

## Article

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

Fiona Wanjiku Moejes<sup>1</sup>, Ovidiu Popa<sup>2,3,†</sup>, Antonella Succurro<sup>3,4,†,\*</sup>, Julie Maguire<sup>1</sup> and Oliver Ebenhöh<sup>2,3,\*</sup>

<sup>1</sup> Bantry Marine Research Station, Gearhies, Bantry, Co. Cork, Ireland; fmoejes@bmrs.ie (F.W.M.); jmaguire@bmrs.ie (J.M.)

<sup>2</sup> Institute of Quantitative and Theoretical Biology, Heinrich-Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany; ovidiu.popa@hhu.de

<sup>3</sup> CEPLAS (Cluster of Excellence on Plant Sciences), Heinrich-Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany

<sup>4</sup> Botanical Institute, University of Cologne, Zùlpicher Straße 47b, 50674 Cologne, Germany

\* Correspondence: a.succurro@uni-koeln.de (A.S.); oliver.ebenhoeh@hhu.de (O.E.)

† These authors contributed equally to this work.

**Abstract:** The pennate diatom *Phaeodactylum tricornutum* is a model organism able to synthesize industrially-relevant molecules. Commercial-scale cultivation currently requires large monocultures, prone to bio-contamination. However, little is known about the identity of the invading organisms. To reduce the complexity of natural systems, we systematically investigated the microbiome of non-axenic *P. tricornutum* cultures from a culture collection in reproducible experiments. The results revealed a dynamic bacterial community that developed differently in “complete” and “minimal” media conditions. In complete media, we observed an accelerated “culture crash”, indicating a more stable culture in minimal media. The identification of only four bacterial families as major players within the microbiome suggests specific roles depending on environmental conditions. From our results we propose a network of putative interactions between *P. tricornutum* and these main bacterial factions. We demonstrate that, even with rather sparse data, a mathematical model can be reconstructed that qualitatively reproduces the observed population dynamics, thus indicating that our hypotheses regarding the molecular interactions are in agreement with experimental data. Whereas the model in its current state is only qualitative, we argue that it serves as a starting point to develop quantitative and predictive mathematical models, which may guide experimental efforts to synthetically construct and monitor stable communities required for robust upscaling strategies.

**Keywords:** microbial communities; host-microbe interactions; mathematical modelling; diatoms; synthetic ecology; algal biotechnology

## 1. Introduction

Microalgae are photosynthesis-driven cells able to store light energy by converting carbon dioxide into carbohydrates, lipids, proteins, and other cellular components with potential biofuel, food, feed, and pharmaceutical and nutraceutical applications [1]. Novel applications also include the use of microalgae as an alternative sustainable development tool [2]. One such microalgae is the pennate diatom *Phaeodactylum tricornutum* that is able to synthesize a number of industrially relevant molecules applicable in: aquaculture as feed in e.g. bivalve, echinoderm, crustacean and fish hatcheries [3,4]; as biomass for biofuels [5,6]; pharmaceuticals and nutraceuticals [5,7–9]; and nanotechnology [10], and bioremediation industries [11]. To fully exploit the industrial potential of microalgal-derived products, substantial quantities of microalgal biomass is required, preferably obtained whilst maintaining low production costs. This is achieved by implementation of large-scale cultivation methods such as open raceway ponds and photobioreactors. The majority of conventional cultivation methods rely on keeping monocultures of the desired species, particularly if the final

product is a bioactive molecule for human consumption [12]. Photobioreactors are closed systems that allow for the production of monoseptic cultures, fully isolated from potential contamination if cultivation protocols are followed correctly [13]. However, high operational costs of photobioreactors might not be sustainable. Another option is open raceway ponds, which are simple open-air cultivation systems that have been in use since the 1950s [1]. They are highly susceptible to contamination, and unless the desired species is a halophile or thermophile [14], 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 [15,16]. Although much research is carried out in the field of microalgal culture upscaling, very little is known about the true identity and characteristics of these invading organisms, responsible for microalgal culture “crashes” which lead to loss of biomass, and therefore, loss of revenue.

Microalgae are not found in monoculture in nature and it is not surprising that imposing such an artificial environment results in unstable large-scale cultures. By understanding rather than attempting to push out these micro-invaders, potential alternatives such as “synthetic ecology” as novel scaling up techniques should be explored [17]. This concept has as a cornerstone the Competitive Exclusion Principle, or Gause’s Law, which states that competition among organisms results in the specialization towards different ecological niches where they can thrive [18,19]. By “synthesizing” a community of organisms that fills every niche in the ecosystem of the microalgal culture and support, rather than harm, the growth of the phototroph, we would automatically optimize the utilization of nutrients and prevent the establishment of other potentially detrimental organisms [17]. In order for synthetic ecology to be a legitimate contender as a novel scaling up technique, a greater understanding of species-specific interactions is required, starting with the bacterial faction, which are present in all of the Earths’ biomes [20], and arguably the key players in maintaining balance within a system. Theoretical ecology employs mathematical models to study the emergent patterns in ecosystems dynamics [21]. Because of the many industrial applications of microbial communities, current research has shown great interest in improving our understanding of such systems [22]. In particular, mathematical models and interdisciplinary approaches are fundamental to unlock the crucial understanding of the mechanisms regulating community dynamics [23,24]. Since the same system can be inspected at different spatio-temporal scales and at a different degree of complexity, it is important to select the most suitable method to describe the biological phenomena under study in mathematical terms [25]. The first ecosystem models at the population-scale date back to the 1920s with the well-know Lotka-Volterra (LV) predator-prey model [26,27]. Since then, LV models have been extensively used to represent cooperation/competition population dynamics with a system of ordinary differential equations (ODEs). In generalized LV models (gLV) the system includes an arbitrary number of co-existing organisms and they directly represent positive/negative pairwise interactions as fixed parameters [28]. Today, a gLV model can be developed by inferring a co-occurrence network from a time series of metagenomics data [29]. This however requires a reasonable number of time resolved metagenomics data and will provide information on direct, one-on-one interactions only.

Diatoms and bacteria have co-evolved for more than 200 million years [30], and their co-existence is most likely based on a “biological barter trade system”, where substances such as trace metals, vitamins, and nutrients (nitrate, phosphate, silicate, carbon) are exchanged. In this work, we built on previous research that investigated algae-bacterial interactions including Provasoli’s work from 1958 where he suggested that bacteria can enhance the growth of algae [31], and subsequent species-specific studies that further corroborated his initial idea [32–35]. We first characterized the relative composition of the bacterial community in non-axenic *P. tricornutum* cultivated in the presence and absence of trace metals, vitamins and sodium metasilicate at different time points. Secondly, using critical peer-reviewed literature we defined the most likely functional roles of the bacterial factions and reconstructed an interaction network. Lastly, from the derived putative

network of interactions, we built an ODE model with modified Verhulst equations for microbial growth that included the direct effect of nutrient availability. Mortality rates were also introduced as dependent on specific bactericidal substances. The mathematical model, even if purely qualitative, with parameters fitted to the available experimental data, served as a proof of concept that data as obtained here is sufficient to reconstruct a theoretical model that a) reproduces the experimental observations, thus demonstrating consistency of our assumptions, and b) allows for testing different hypotheses regarding the nature of the metabolic interactions underlying the ecosystem dynamics. It therefore represents a starting point to gain a deeper understanding of the principles of microbial community dynamics by an iterative experimental and theoretical approach.

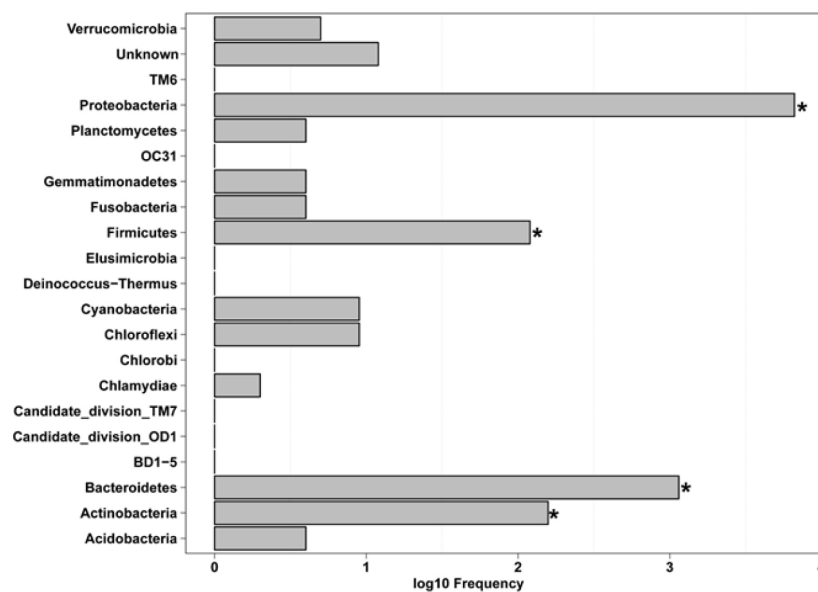
## 2. Results

### 2.1. Characteristics of *Phaeodactylum tricornutum* growth

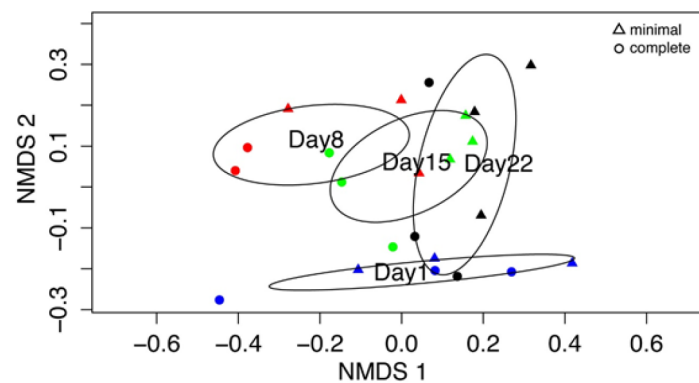
The media composition was shown to have a significant effect on the growth characteristics of *P. tricornutum*. *P. tricornutum* cultivated in minimal media exhibited a statistically significantly ( $p = 0.042$ , unpaired Wilcoxon signed rank) higher cell density ( $11.2 \cdot 10^6$  cells/mL) when compared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

### 2.2. Bacterial community profile of *Phaeodactylum tricornutum* cultures

Bacterial gDNA is analysed as described in the Materials and Methods section. Most of the operational taxonomic units (OTUs) could be assigned to the genera level (Figure A.2). Of the 9727 OTUs identified, 8109 corresponded to known sequences in the SILVA database (v.118) [36]. The OTU abundance at the phylum level showed that 99.97% of all OTUs belonged to Proteobacteria, Bacteroidetes, Actinobacteria and Firmicutes (Figure 1). A comparison of the number of individual reads to the number of unique OTUs showed that the high number of reads per phyla is not the result of a single OTU (Figure A.3). OTUs with hits to known 16S *P. tricornutum* sequences were discarded.



(a)

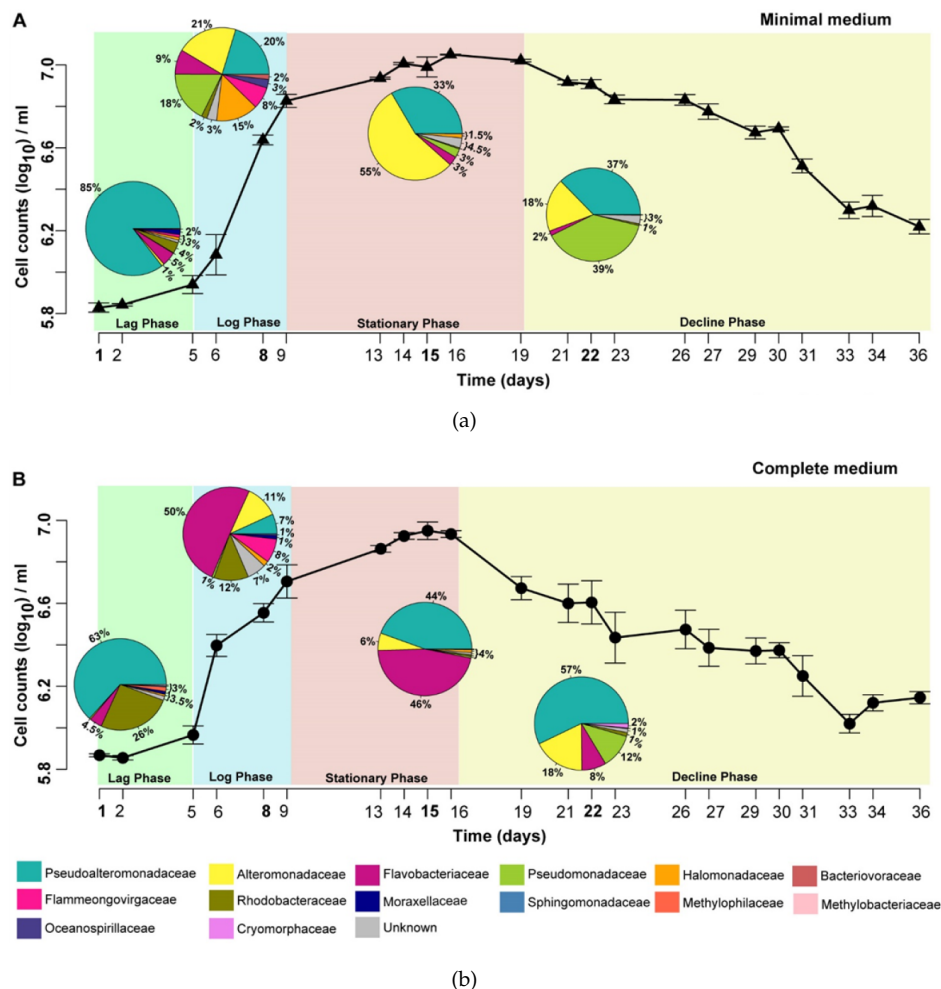


(b)

**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, Bacteroidetes, Actinobacteria and Firmicutes. (b) Ordination plot of bacterial community in the two media conditions for all sampling points. 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 (Figure A.4). 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. All datasets showed a diminished 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 generalized UniFrac distances [37]. 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).



**Figure 2.** Bacterial community profile of *P. tricornutum* (CCAP 1052/1B) over a 36 day period in culture conditions: (a) minimal media; (b) complete media. The growth curves are partitioned into lag (green), log (blue), stationary (red), and decline (yellow) phases. The abundance (%) of the “Top Ten” bacterial families (corresponding colors described in the key) is depicted in pie charts on Days 1, 8, 15 and 22 in both media conditions.

The existence of one dominant family at each investigated time point is a peculiar characteristic. In minimal media (Figure 2(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 (Figure 2(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, Pseudoalteromonadaceae (57%) show clear dominance of the community during the decline phase.



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 generalized UniFrac distance [37]. The significance for the test was assessed by 5000 permutations. The results of the PermanovaG tests support the NMDS ordination, confirming a statistically significant effect in the bacterial community profile at the different sampling points and in the two media conditions whereas no significant effect was found in the experimental replicates (Figure A.5).

### 2.3. 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 to avoid diluting the signal of the less abundant genera. Figures A.6 and A.7 show 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. 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.

In complete media, members of the Pseudoalteromonadaceae family are highly abundant when *P. tricorutum* cell densities are low (63% and 57% on Day 1 and Day 22, respectively). Flavobacteriaceae species dominated (50%) when the *P. tricorutum* culture is growing exponentially (Day 8). Day 15, when *P. tricorutum* cell densities are at their highest, shows co-dominance of both Flavobacteriaceae (46%) and Pseudoalteromonadaceae (44%).

In minimal media, members of the Pseudoalteromonadaceae family are highly abundant when *P. tricorutum* cell densities are low. On Day 22 Pseudomonadaceae (39%) and Pseudoalteromonadaceae (37%) are both overrepresented. When the *P. tricorutum* culture is growing exponentially (Day 8) a cluster of Families dominate; namely Alteromonadaceae (21%), Pseudoalteromonadaceae (20%), Pseudomonadaceae (18%), Halomonadaceae (15%) and Flavobacteriaceae (9%). When the cell density of *P. tricorutum* peaks (Day 15), the Alteromonadaceae species take over (55%).

The bacterial communities within the two media conditions on Day 1 are more closely related than the communities on days 8 and 15 (see Table A.2 for generalized UniFrac distances). As the cultures begin to “crash” (Day 22), the bacterial communities in the two media conditions increase in similarity again. In general, the main families identified show a distinct pattern of disappearance and regeneration within the bacterial community. In the complete media, Pseudoalteromonadaceae species start at 63% (Day 1), drops in abundance to 7% (Day 8) then recovers to 57% (Day 22). Flavobacteriaceae species, in complete media, start at 4.5% (Day 1), increases in abundance to 50% (Day 8), and then falls back to 8% (Day 22). In the minimal media, Alteromonadaceae species have an abundance of only 1% (Day 1), peaks at 55% (Day 15), and decreases down to 18% (Day 22).

### 2.4. Assessment of prospective role of central bacterial families from literature

#### 2.4.1. Pseudoalteromonadaceae

Members of Pseudoalteromonadaceae family have been isolated from coastal, open and deep-sea waters, sediments, marine invertebrates, as well as marine fish and algae [38]. The Pseudoalteromonadaceae family has three genera, namely *Pseudoalteromonas*, *Algicola* and *Psychrosphaera* [39]. Several species of Pseudoalteromonadaceae are reported to possess antibiotic properties with bactericidal effects [40]. For example, concentrated supernatant of a marine bacterium *Pseudoalteromonas* sp. strain A28 contained various enzymes including proteases, DNases, cellulases, and amylases, capable of causing the lysis of the diatom *Skeletonema costatum* [41]. Species of Pseudoalteromonadaceae are also capable of producing cold-adapted enzymes [42–47]. Pseudoalteromonadaceae species can produce extracellular polymeric substances allowing them

to colonise surfaces, enhancing nutrient uptake whilst limiting diffusion of particular substances across the cell membrane [48]. The ability of Pseudoalteromonadaceae species to suppress the growth of competing bacteria could explain the dominance of Pseudoalteromonadaceae in almost all cultures irrespective of media composition, particularly when *P. tricornerutum* abundance is limited (Figure 2, Days 1 and 22). *P. tricornerutum* on the other hand, may protect other bacterial community members from the bacteriolytic ability of Pseudoalteromonadaceae by producing specific antibacterial compounds themselves. Desbois *et al.* showed that *P. tricornerutum* excreted bacteriolytic fatty acids such as eicosapentaenoic acid (EPA; 20:5n-3), nucleotides, peptides, and pigment derivatives that can eliminate unwanted competition for nutrients such as organic phosphates from certain bacteria [49].

#### 2.4.2. Alteromonadaceae

The Alteromonadaceae family consists of 16 (yet annotated) named genera [50] found predominantly in marine environments [39]. 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 [51]. They are able to utilize a vast array of compounds as carbon sources; from glucose to glycerol [39]. Members of this family are known siderophore producers [52–54]. 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 [55]. Most bioactive trace metals, including iron, exist at nanomolar to picomolar concentrations in our oceans, approximately one-millionth of the intracellular concentration in diatoms [56,57]. 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. tricornerutum* (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  $\mu\text{M}$  of iron (III) chloride hexahydrate.

#### 2.4.3. Flavobacteriaceae

Flavobacteriaceae are members of the Bacteroidetes phylum and include over 120 genera [58] found in soil, sediments and seawater (see [59] 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 [60–62]. Members of Flavobacteriaceae can proficiently degrade various biopolymers such as cellulose, chitin and pectin [63,64]. 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 [65,66]. Although the exact mechanisms behind them are not perfectly understood, algal blooms are a consequence of exponential growth of phytoplankton [67]. In this respect, the phase of exponential growth of *P. tricornerutum* in complete media, when our results showed highest abundance of Flavobacteriaceae, is the artificial equivalent of an algal bloom of *P. tricornerutum* (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.

#### 2.4.4. Pseudomonadaceae

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

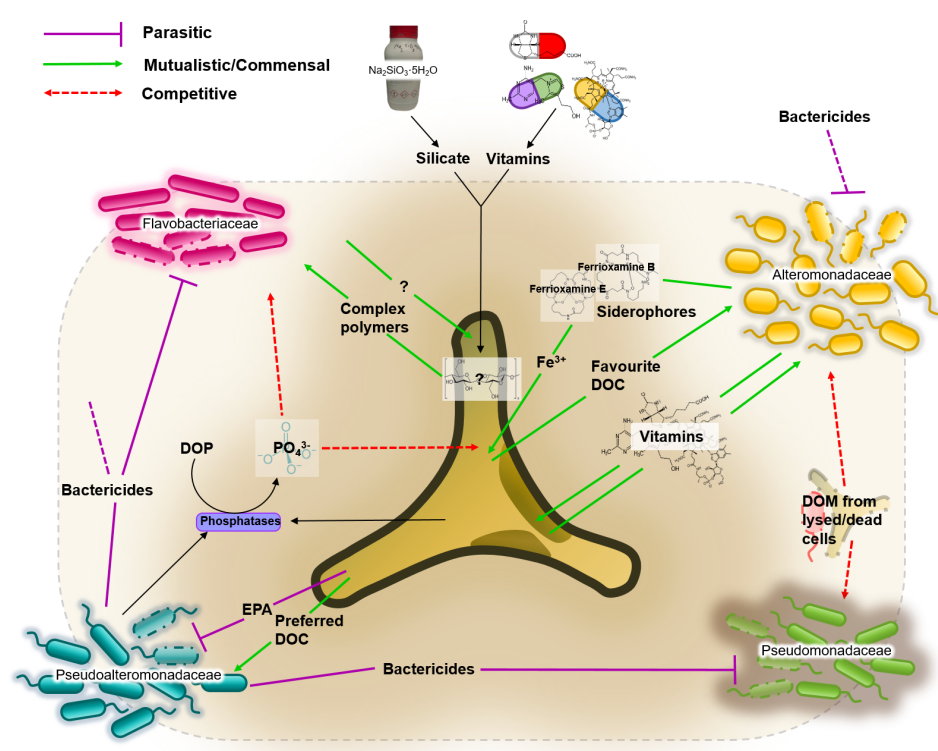
pathogens [68]. Species from the *Pseudomonas* genus are the best studied of the Pseudomonadaceae family, whose sheer genetic diversity explains the ability to thrive in such a wide range of environments [69]. Marine isolates from the *Pseudomonas* genus have been shown to produce a wide range of bioactive compounds, many of which exhibit antibacterial as well as antiviral properties (see [70] 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 Pseudomonadaceae when the *P. tricornutum* culture has “crashed” could be attributed to its ability to produce antibacterial compounds allowing members of this family to begin to thrive in the community through inhibition of its competitors. Given its exceptional genetic diversity, and thus, its metabolic versatility, allows for members of Pseudomonadaceae to be truly saprophytic; providing a hypothetical explanation of its abundance we could measure when the *P. tricornutum* cultures crash (Figure 2, Day 22 in both media conditions).

#### 2.4.5. Network of putative interactions between *Phaeodactylum tricornutum* and identified bacterial families

The putative roles of each of the dominant families are illustrated in Figure 3. We identified from extensive literature review five types of relevant metabolites associated with the identified bacterial families and playing a potential role in diatom-bacterial interactions. These are: bactericidal metabolites; iron; vitamins; dissolved organic carbons; dissolved organic phosphates.

- **Bactericidal Metabolites.** Several species of the Pseudoalteromonadaceae family have been reported to possess bactericidal effects [40]. This ability to suppress the growth of competing bacteria could explain the dominance of Pseudoalteromonadaceae in almost all cultures irrespective of media composition. *P. tricornutum* also demonstrates bactericidal properties by excreting fatty acids (such as eicosapentaenoic acid or EPA), nucleotides, peptides, and pigment derivatives [49].
- **Iron.** Iron acquisition is essential for biological processes such as photosynthesis, respiration and nitrogen fixation. Bacteria produce and excrete siderophores, which scavenge iron. Diatoms are not known to produce siderophores, but genome sequence analyses identified the presence of a gene orthologue of a bacterial ferrichrome binding protein that suggests the possibility of iron (III)-siderophore utilization by *P. tricornutum* [52,71]. Furthermore, it was shown that *P. tricornutum* was able to uptake siderophores ferrioxamines B and E [72].
- **Vitamins.** Prokaryotes are thought to be the main producers of B vitamins [73,74]. Although *P. tricornutum* does not require cobalamin, thiamine and biotin [75], production of organic compounds such as EPA can be considerably enhanced by the bioavailability of co-factors such as cobalamin [76]. This provides the basis for potential mutualistic interactions. For example, Alteromonadales, dominant in our cultures, are thought to be capable of producing B vitamins [77].
- **Dissolved Organic Carbon (DOC).** It is estimated that up to 50% of carbon fixed via phytoplankton-mediated photosynthesis is utilized by marine bacteria [78], mainly as DOC compounds, defined as the organic material < 0.7 µm in size [79]. DOC from diatoms originates either from live cells or recently lysed or grazed cells, which determines the type of DOCs available, and therefore are likely to influence the bacterial consortia associated with the diatom [30].
- **Dissolved Organic Phosphate (DOP).** Both diatoms and bacteria primarily utilize orthophosphate as a source of phosphorus. However, to access phosphate from DOP compounds, both diatoms and bacteria developed mechanisms to release orthophosphate ( $\text{PO}_4^{3-}$ ) from DOP. The mechanism is not species-specific, which consequently means the “free” orthophosphates can be acquired by any organism [80].

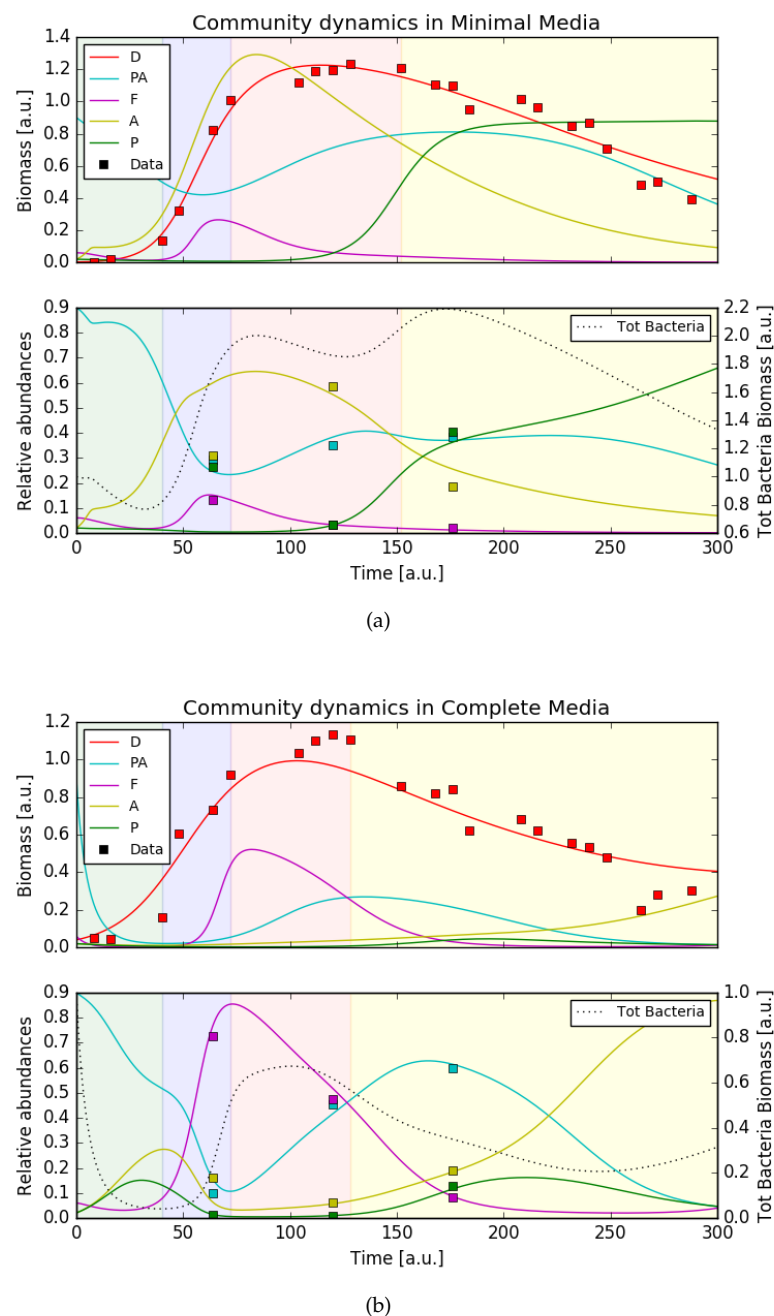




**Figure 3.** Network of putative interactions between *Phaeodactylum tricornutum* and identified bacterial families. The dotted grey line depicts the “phycosphere”; a term coined by Bell and Mitchell in 1972 as an aquatic equivalent of the “rhizosphere”, denoting the region extending outwards from the algal cell in which bacterial growth is stimulated by extracellular products of the alga [81].

### 2.5. Mathematical model simulations

Based on the network of putative interactions between diatoms, bacteria, and the environment, we constructed a dynamic mathematical model, based on generalized growth-laws reflecting the metabolic dependencies (for details on the applied equations see Appendix C.1). Figure 4 presents results of the model simulations after the model parameters were fitted to the data in minimal and complete media conditions, respectively (see Materials and Methods and Appendix C.1). Experimental data are superimposed. The top panel shows biomasses of the five organisms (data available only for the diatom), the bottom panel shows relative bacteria abundance versus time (individual biomass divided by total bacterial biomass). Because of the qualitative nature of the model, units are arbitrary. The figures show that the model is able to reproduce the main features of the bacterial community dynamics, like the disappearance and return of *Pseudoalteromonadaceae* in complete media and the peak of *Alteromonadaceae* at the end of the exponential growth phase of *P. tricornutum* in minimal media.



**Figure 4.** Simulation results (lines) and experimental data (squares) for communities of *P. tricornutum* (D), Pseudoalteromonadaceae (PA), Flavobacteriaceae (F), Alteromonadaceae (A) and Pseudomonadaceae (P) in (a) minimal media and (b) complete 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.

Due to the large number of free parameters, this fit is certainly not unique and therefore, with the data available, it is not possible to make any quantitative statements about the actual interaction parameters. Neither it can be assumed that simulation results are of general validity. Despite

these limitations, the model does represent a possible configuration of diatom-bacteria-environment interactions, which is in agreement with the experimentally observed bacterial dynamics.

In order to test the stability of the bacterial community and how it supports the growth of *P. tricornutum* we run simulations using the same set of parameters (either minimal or complete media conditions) and varying the initial community composition (see Appendix B). In complete media the simulated growth of *P. tricornutum* still fits the experimental data, rather independently from the bacterial community. Also under axenic conditions, diatom growth is predicted to be largely unperturbed. This situation is different in minimal media. Whereas also under these conditions diatom growth is unaffected upon removal of the three bacterial families Pseudoalteromonadaceae, Flavobacteriaceae and Pseudomonadaceae, removing Alteromonadaceae from the community results in no growth at all of *P. tricornutum* (Figure B.2). This behaviour is expected from the hypothesized central role of this bacterial family to supply the diatom with micronutrients. Surprisingly, removing single bacteria from the community in both media conditions still gives in general a good fit of the (recomputed) relative abundances, except when removing Alteromonadaceae in minimal media (as direct consequence of what stated previously) and when removing Pseudoalteromonadaceae in both media conditions (Appendix B). This hints to a relevant role of Pseudoalteromonadaceae in regulating the community composition through its predatory strategy of releasing bactericidal substances. Finally, the community composition at the last time point is overall better captured, possibly because at that time dominance of one family over the others is settled and our network of interactions consistently captures these general trends.

### 3. Discussion

In nature, *P. tricornutum* does not exist as an isolated entity. In fact, it is part of a complex ecosystem, including both environment and other organisms, and whose complete network of interactions is still poorly understood. Microbial ecosystems are of high interest for a wide range of applications in fields like medicine, renewable energy, agriculture. In order to accelerate our knowledge of the fundamental phenomena regulating microbial community dynamics, it is important to implement interdisciplinary approaches [25].

To reduce the complexity typical of a natural, variable system, we investigated non-axenic laboratory strains of *P. tricornutum* from a culture collection, typically used also for industrial applications. In particular, as a first step towards the identification of key mechanisms in ecosystem interactions, we studied in a reproducible laboratory set-up the trends of the bacterial community dynamics during batch growth of *P. tricornutum* in two conditions: minimal and complete media. The cultivation method we developed was designed to compromise between highly controlled small-scale laboratory conditions and a large-scale industrial set-up. Our approach, integrating experiments, bioinformatics and mathematical modeling, illustrates a possible way towards the development of a monitoring pipeline for non-axenic microalgae cultures.

#### 3.1. Experimental observation of the dynamics of the bacterial community associated to *Phaeodactylum tricornutum*

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 indicates a more stable culture in the minimal media (Figure 2). This also suggests that non-axenic cultures of *P. tricornutum* might not require expensive trace metals and vitamins for optimal growth under the conditions provided. In the context of scaling up, this would drastically decrease the production cost of large-scale cultivation of *P. tricornutum*. Simultaneously, the dynamics of the bacterial community reveals that the community in the minimal media increased 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 [82–85]. Our observations are in agreement with more recent hypotheses indicating that diversity generally increases the stability of an ecosystem [86].

### 3.2. Putative network of interactions and validation with a qualitative mathematical model

Our bioinformatics analysis of bacterial gDNA abundance showed clear dominance of four bacterial families: Pseudoalteromonadaceae, Alteromonadaceae, Flavobacteriaceae and Pseudomonadaceae. These bacteria are over-represented in all samples and their relative abundances show different temporal dynamics among the two *P. tricornutum* growth condition. In order to understand the putative functional role of these bacteria we made an extensive literature review. This work revealed interesting insights on the possible metabolic exchanges going on and allowed us to draw interaction links among *P. tricornutum* and its associated bacterial community. We critically considered which metabolites were most relevant for survival (organic carbons for the bacteria, iron, vitamins and phosphates for the diatom) and which ones could play a role in competition and predation among the microbes (bactericidal metabolites). From these considerations we designed a putative network of interactions that was then translated into a mathematical model. Direct metabolic exchanges are known to be central in microbial community interactions [87], but usually population dynamics models like gLV [28] do not include this information. We therefore modified the standard formulation of the Verhulst equation [88] for bacterial growth to include organism-to-organism interactions depending on the production/consumption of metabolites. Metabolites included nutrients, micronutrients and toxins. An ODE model at the population level cannot, of course, capture mechanisms such as metabolic shifts caused by changes in the environment such as the supplementation of minimal or complete media. Therefore, we did not expect to find a unique set of parameters for the model in the two conditions. However the parameters fitted to the data of *P. tricornutum* with four bacterial families still provide good fits in simulations with altered community composition. Considering the limited information that can be extracted from the current experimental data, the model is purely qualitative and provides a proof of concept that a quantitative model can in principle be constructed if dedicated experiments are designed for calibration. The current qualitative model provides therefore a preliminary validation of our putative network of interactions, and serves as motivation for further research bringing the model to a quantitative, predictive level. Indeed, starting with systematic measurements of model parameters in co-cultivation experiments, the simulations can gain predictive power and become a powerful tool towards the goal of synthetic community design and control.

### 3.3. Future research directions

In this study we have demonstrated that the bacterial community associated with non-axenic laboratory strains of *P. tricornutum* is not randomly assembled but follows dynamics that can be reproduced. We postulate that a role within the community can be filled by a number of bacterial species capable of carrying out a certain function (guilds) rather by one specific species of bacteria. 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. Unfilled niches will be seized by bacteria with the ideal metabolic functionality. 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 and to further develop the qualitative mathematical model to understand the specific community roles and the ecological niches. In the context of fundamental research, one approach would be to carry out systematic time-resolved omics studies, which provide a holistic view of the genes (genomics) and metabolites (metabolomics) in a specific biological sample in a non-targeted and non-biased manner [89], and use them to develop an “expanded gLV” mathematical model where the species specific interaction terms depend on the metabolite concentrations. This would allow to derive a network of interactions independent of a priori hypotheses, and thus represent a significant step forward in understanding community dynamics based on metabolic exchanges. In the context of industrial scale up, systematic co-culture experiments with culturable members of the bacterial families of interest, chosen based on desired functional roles, could be used to parametrize a mathematical model like the one we presented and develop it into a powerful predictive tool for

culture monitoring. For example, samplings assessing the community composition can be used to predict the harvesting point and avoid “culture crash”. Synthetic ecology for the scale-up of co-cultivation strategies is extremely relevant for pharma- and nutraceutical, as well as animal feed industries. Further research into co-cultivation approaches and the support of changes policy-makers and legislative bodies (e.g. the European Food Safety Authority) is needed to address urgent and significant societal challenges such as food security.

## 4. Materials and Methods

### 4.1. Strains and culture conditions

All *P. tricornutum* strains were obtained from the Culture Collection of Algae and Protozoa (CCAP) based in Oban, Scotland [90]. All cultures were obtained non-axenic. Based on previous experimental evidence [91], 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 containing sources of nitrogen ( $\text{NaNO}_3$ ) and phosphorus ( $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$ ), as well as trace metals and vitamins with the addition of sodium metasilicate, as per Guillard and Ryther 1962 [92] and Guillard 1975 [93], and (2) minimal media which contained just sources of nitrogen ( $\text{NaNO}_3$ ) and phosphorus ( $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$ ) at the same concentration as in the F/2 medium recipe. All cultures were grown in hanging 5L polyethylene bags with a “V” shaped bottom prepared using a heat sealer (Figure A.1). All cultures had a modified aeration system provided by a 10ml pipette attached to the main pressurised air supply via  $0.2 \mu\text{m}$  sterile air filters. Cultures were kept at  $18\text{--}20^\circ\text{C}$  and 24hr light at an average of  $132.3 \mu\text{mol m}^{-2} \text{s}^{-1}$ . All cultures, irrespective of media condition, were inoculated with 250ml from the same 5L stock culture of actively growing non-axenic *P. tricornutum*.

### 4.2. Growth measurements

Growth was monitored every 24 to 48 h using a light microscope and carrying out cell counts of each culture in quadruplicate. 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 a  $-20^\circ\text{C}$  freezer.

### 4.3. Genomic DNA extraction

All samples from Day 1, 8, 15, and 22 were thawed in a water bath set at  $25^\circ\text{C}$ . As per de Gouvion Saint Cyr *et al.* [94], samples were centrifuged for 5 mins 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 5 mins at 2000g. This was repeated twice to ensure that majority of the bacteria attached to the pellet were released and were included in the community analysis. Genomic DNA extraction was carried out in the Aquaculture and Fisheries Development Centre, University College Cork. Mo Bio’s PowerWater® DNA Isolation Kit (catalogue no. 14900-100-NF) was utilized to carry out the genomic DNA extraction. 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.

### 4.4. Barcoded 16S-V6-Next Generation Sequencing

Ion Torrent™ barcoded Next Generation Sequencing protocol was used to sequence the bacterial gDNA [95,96]. Amplification of the V6 hyper variable region of 16S rRNA with forward and reverse



primers (Supplementary Table S2) was carried out. Ion Reporter™ software assembled all the raw sequencing data and sorted all the reads using the unique sample-specific barcode sequences and removed them from the reads. The outcome was raw FASTQ files which were ready for analysis using bioinformatics tools.

#### 4.5. Bioinformatics analysis

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* [97]. Primers were trimmed with fastq-mcf (version 1.04.807) [98], the resulting sequences were quality filtered and clustered into OTUs with usearch (version 8.0.1517; 32Bit – opensource) [99,100]. Taxonomy assignment was done by QIIME (version 1.9.0) [101] with the implemented uclust classifier based on 97% sequence identity to the reference 16S sequences from SILVA 111 database [36]. Statistical analyses were performed in R [102]. The complete protocol containing all processing steps is available on GitHub (see Supplementary Materials).

#### 4.6. Mathematical model

Starting from our understanding of the organism-to-organism interactions, we developed a dynamic model consisting of 13 ordinary differential equations (ODEs) and including 56 (55 free) parameters (see Appendix C.1). The model was built from the following working hypotheses:

1. the growth rate  $\gamma$  of each population followed a standard Verhulst equation parametrized with a carrying capacity and scaled by Monod-type terms that encode the dependence on necessary nutrients as a positive scaling factor  $< 1$ ;
2. the mortality rate of each population was inversely proportional to  $(1+\gamma)$ , to account for the fact that cells during replication (high growth rate) were healthier;
3. additional contributions to population mortality was given by the presence in the environment of noxious elements like bactericidal substances;
4. changes in metabolite concentrations are in general directly proportional to the growth  $\gamma$  of the consumers and producers;
5. in the event of micronutrient scarcity (Iron and Vitamins in our model), *P. tricornutum* will secrete more organic carbons favored by those bacteria able to provide the needed micronutrients.

The parameters were fitted using a genetic algorithm [103] which was 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 Appendix C.2). The model was written in Python (Python Software Foundation, <https://www.python.org/>) and is available on GitHub with instructions and scripts for running (see Supplementary Materials).

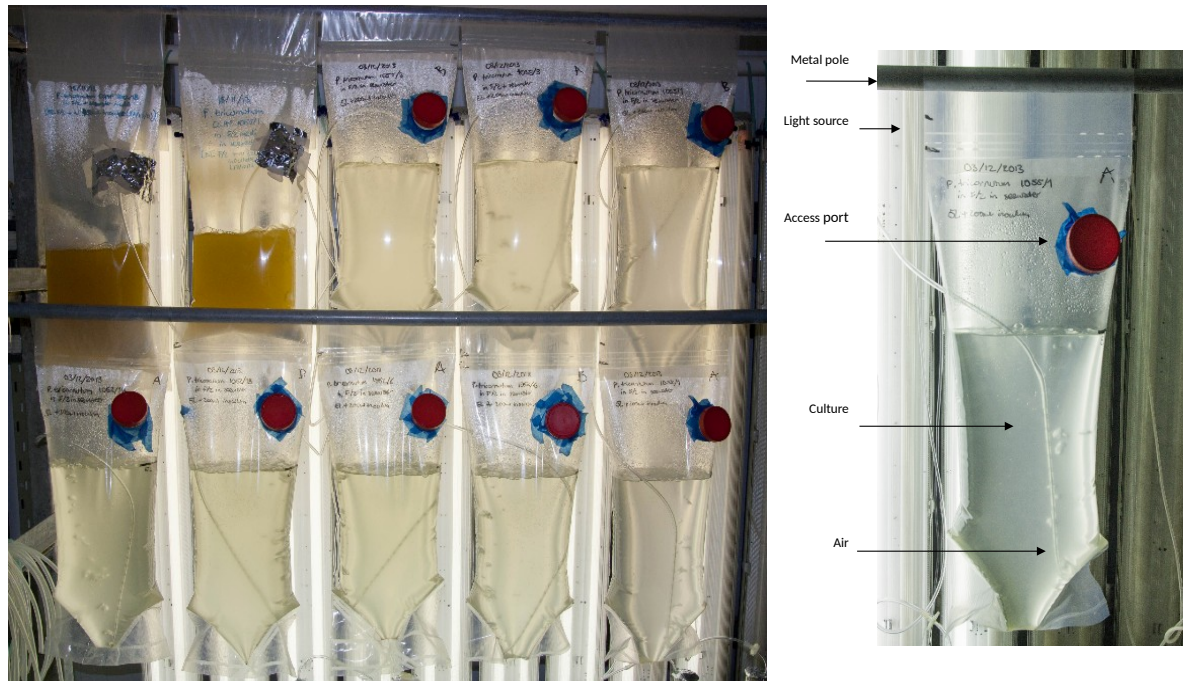
**Supplementary Materials:** The bioinformatics analysis steps are available online at <https://github.com/QTB-HHU/16SV6-Sequence-Analysis.git>. The mathematical model in python with instructions to run simulations is available online at <https://github.com/QTB-HHU/communityODE>.

**Acknowledgments:** This work was supported by the European Commission Seventh Framework Marie Curie Initial Training Network project 'AccliPhot' (grant agreement number PITN-GA-2012-316427) to F.W.M. and A.S.; and the Deutsche Forschungsgemeinschaft, Cluster of Excellence on Plant Sciences CEPLAS (EXC 1028) to O.P., A.S. and O.E. 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.

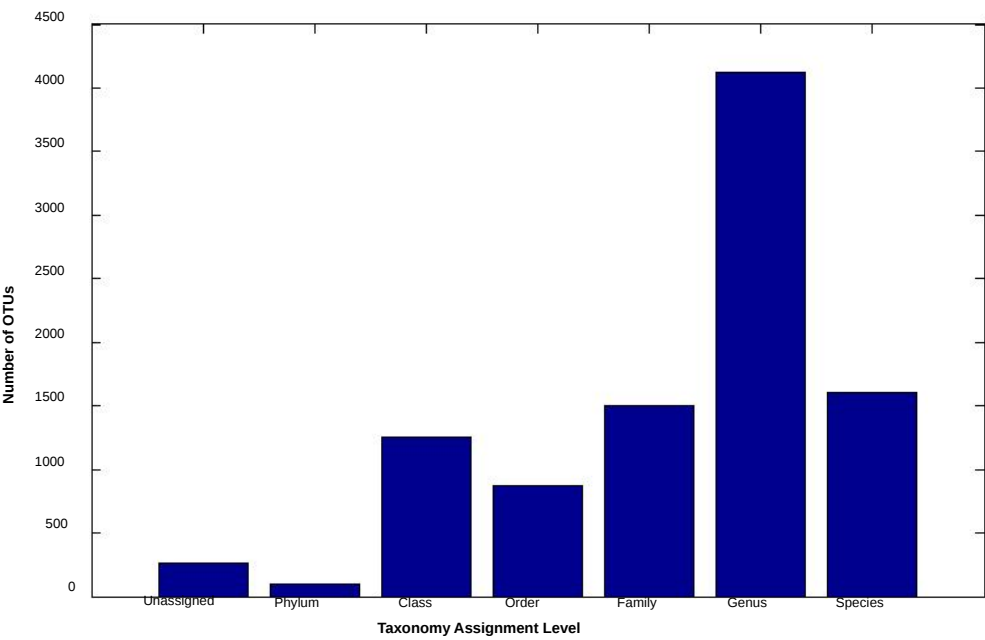
**Author Contributions:** F.W.M., O.E. and J.M. conceived and designed the experiments; F.W.M. performed the experiments; O.P. performed the bioinformatics and statistical data analysis; A.S. developed the mathematical model and performed simulations; all authors contributed to the interpretation of the results and wrote the paper.

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

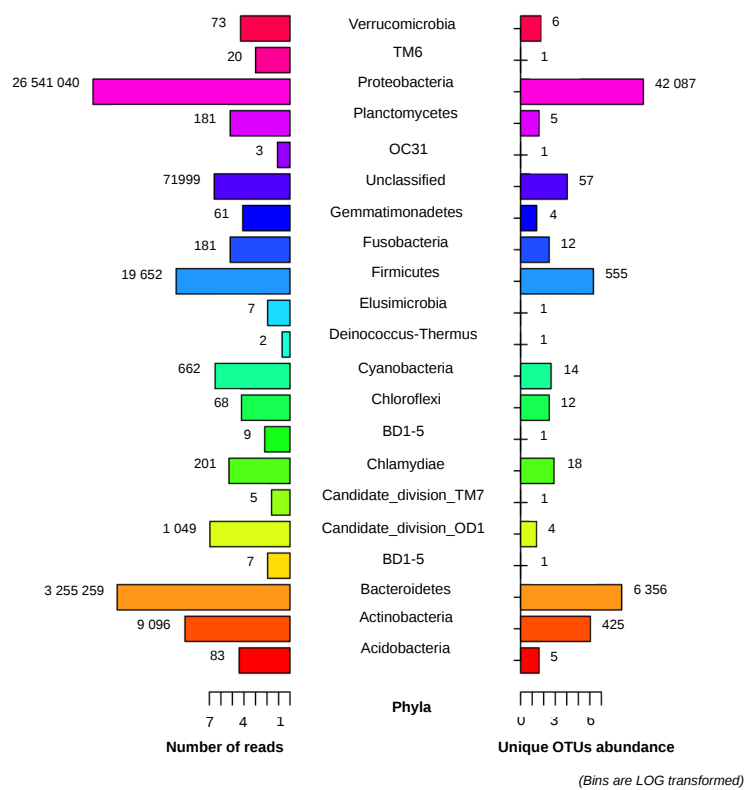
## Appendix Experimental setup and data analysis



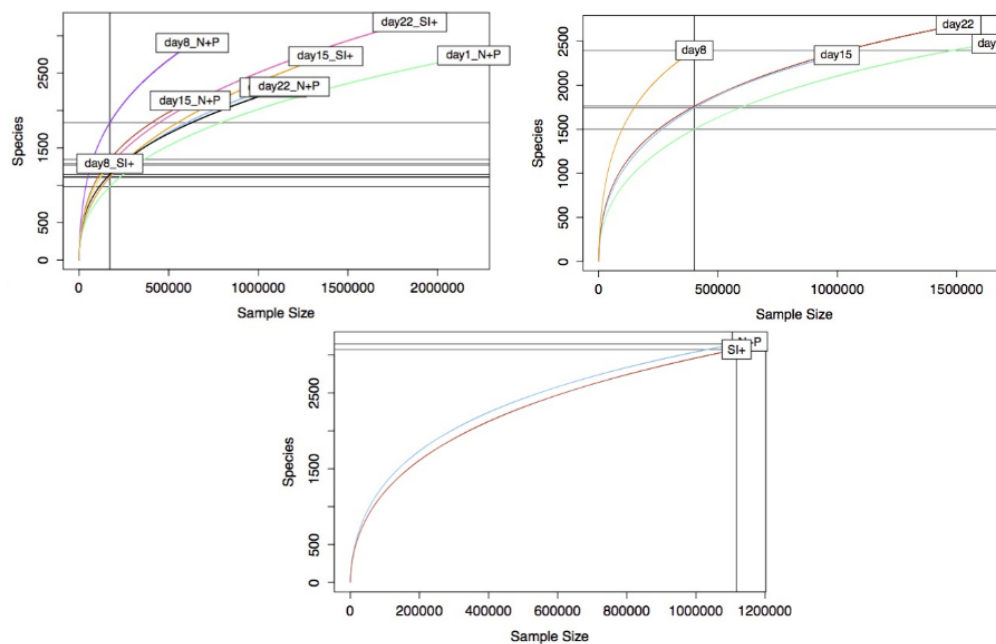
**Figure A.1. Non-axenic *Phaeodactylum tricornutum* culture set up.** 5L polyethylene bags with a “V” shaped bottom were created using the heat sealer machine. The bags were then rinsed and filled with 5L of filtered seawater. Afterwards each bag was sealed and hung approximately 30 cm from the light 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 an 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)



**Figure A.2. Operational Taxonomic Unit (OTU) Taxonomy Assignment Level.** The 16S rRNA gene sequences were clustered to defined Operational Taxonomic Units (OTUs) at  $\geq 97\%$  sequence identity. Most OTUs could be assigned to the genera level, using the SILVA database (v.118) [36]

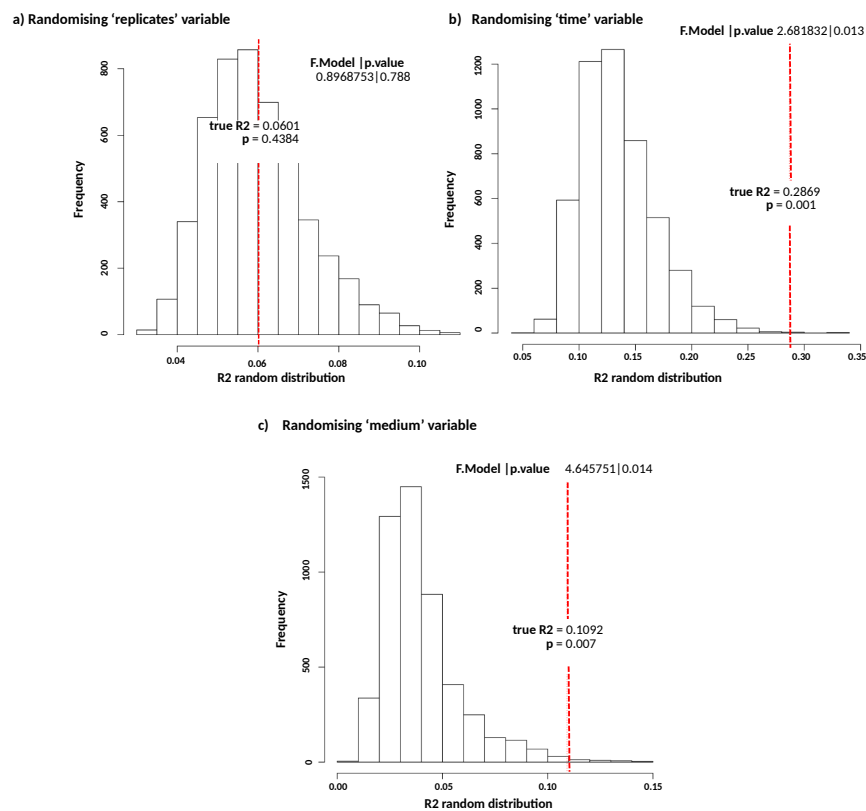


**Figure A.3. Number of reads per unique OTU abundance (at the phylum level).** A comparison of the number of individual reads to the number of unique OTUs shows that phyla with high number of reads do not result in single OTUs

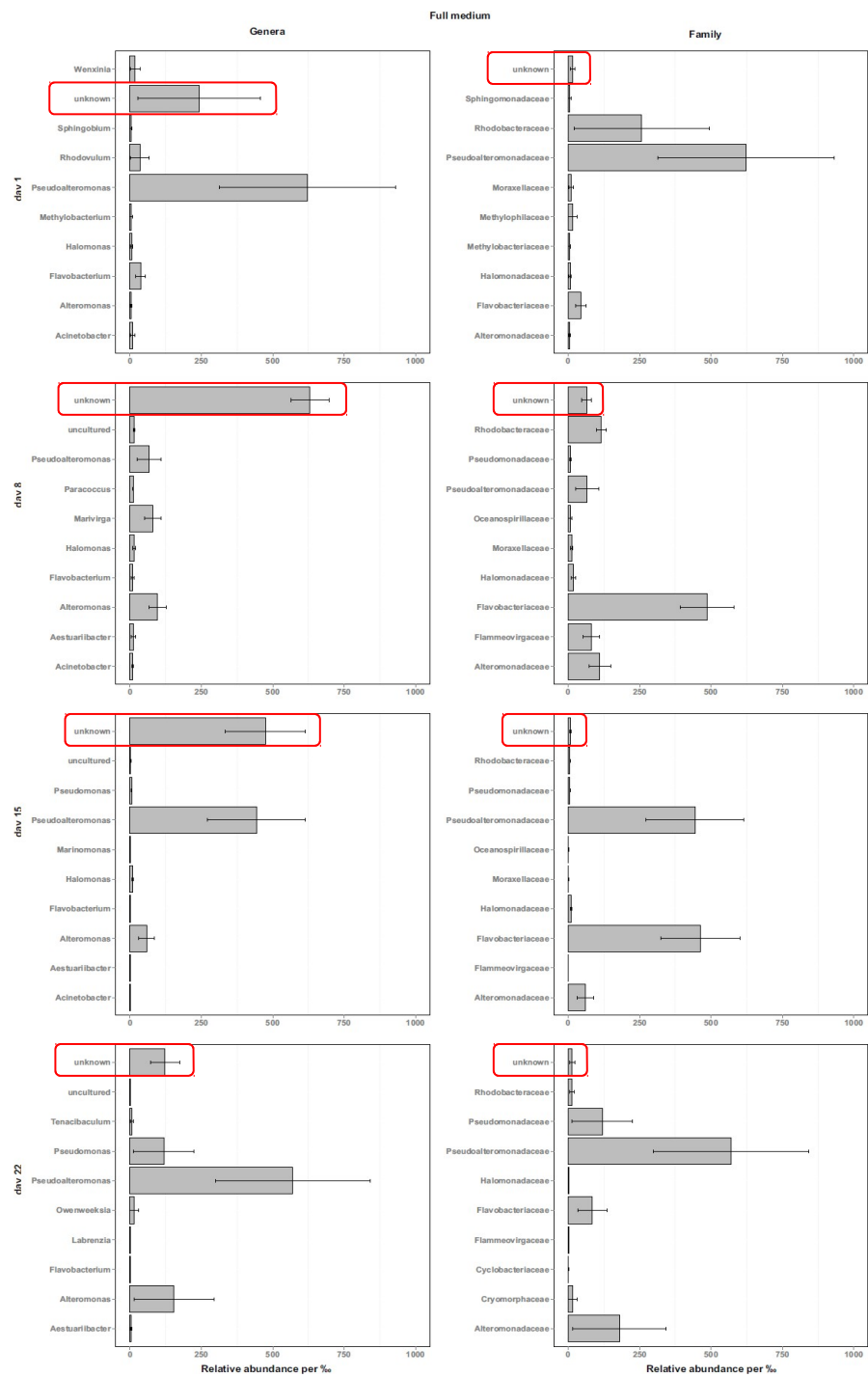


**Figure A.4. Alpha diversity.** Rarefaction curves were used to evaluate the Alpha diversity in the different media conditions as well as at the different time points. Species richness in both minimal and complete media was  $\sim 3\,000$ . Species richness over time remained between  $\sim 2\,400$  and  $2\,600$ , with reduced species richness ( $\sim 1\,300$ ) 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 ( $\sim 3\,000$ ) was shown on Day 22

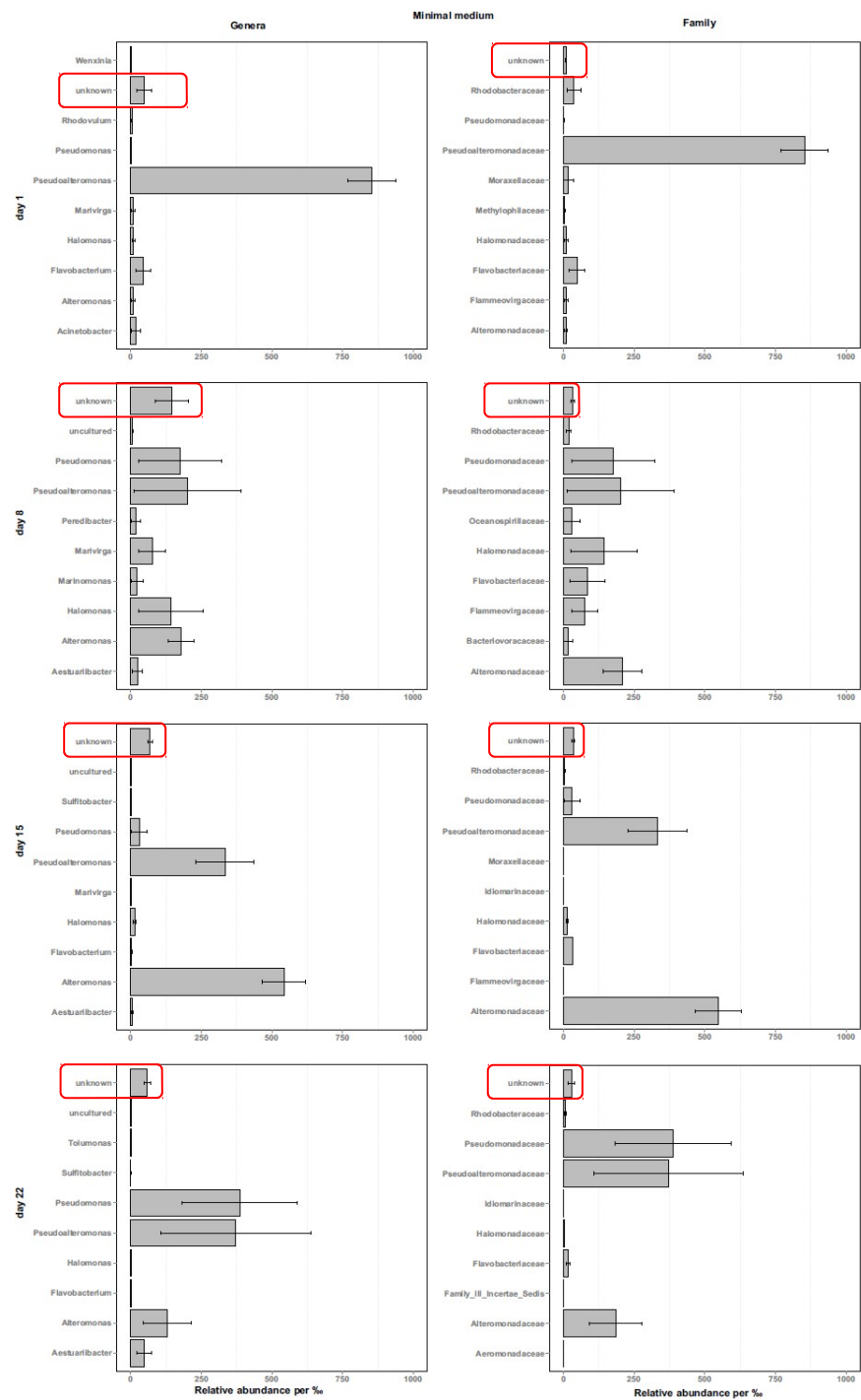




**Figure A.5. Beta diversity.** A modified version of PermanovaG was used to carry out permutational multivariate analysis of variance using multiple distance matrices. The distance matrices [24x24] were previously calculated based on the generalized UniFrac distance [37], weighted UniFrac and unweighted UniFrac [104] distance. The significance for the test was assessed by 5000 permutations. a) shows no significant effect between the replicates (p-value of 0.4384). b) shows a significant effect for the time variable (p-value of 0.001). c) shows also shows a significant effect for the medium variable (p-value of 0.007)



**Figure A.6. Comparison between bacterial community at genera level and family level (complete media).** We show no dynamical difference within the genera that cannot be observed at the family level. Encircled in red, there are a greater number of OTUs that could not be assigned a taxonomy ('unknowns') at the genera level than 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



**Figure A.7. Comparison between bacterial community at genera level and family level (minimal media).** We show no dynamical difference within the genera that cannot be observed at the family level. Encircled in red, there are a greater number of OTUs that could not be assigned a taxonomy ('unknowns') at the genera level than 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

**Table A.1. 16S V6 rRNA primer sequences.** ‘Max’ is the complete media. ‘Min’ is the minimal media. ‘A’, ‘B’, and ‘C’ are the three replicates

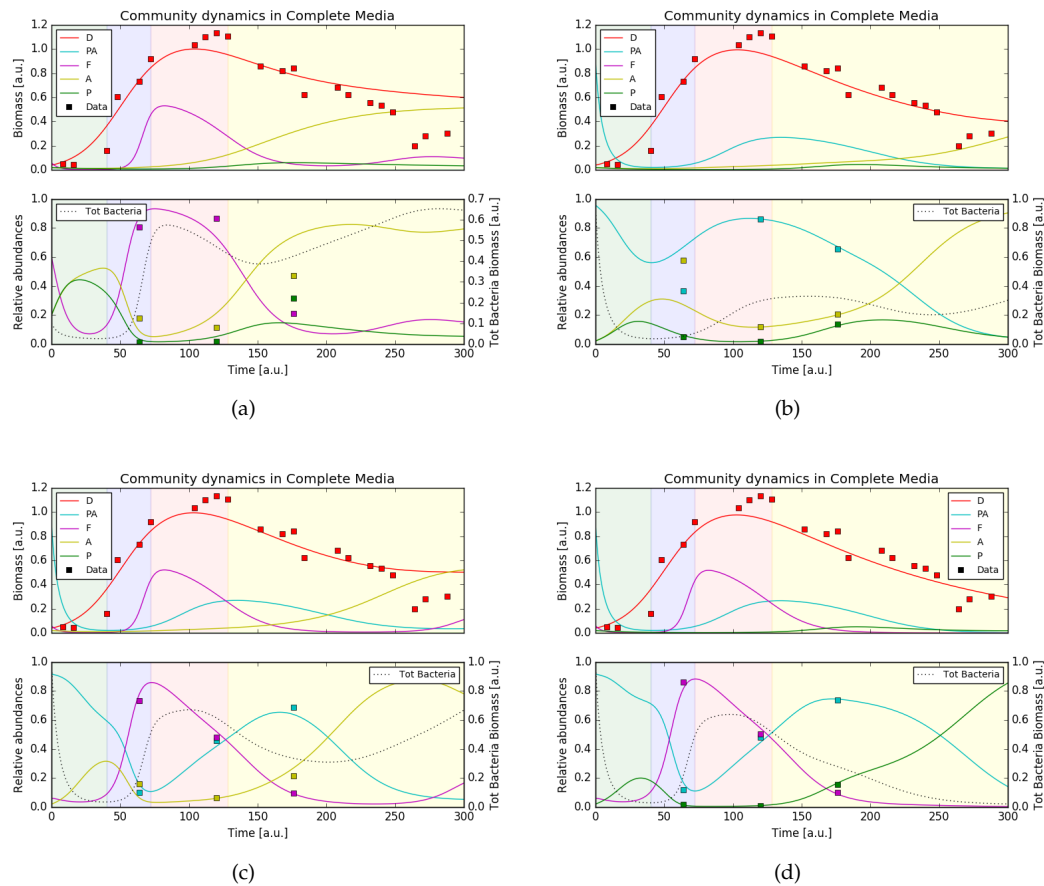
Description	Barcode Name	Adaptor A Sequence	Barcode Sequence	Forward V6 Primer Sequence	Reverse V6 Primer Sequence
Day 1_Max_A	IonXpress_071	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TGAGGCTCCGAC	MWACGCGARGAACCTTACC	CGACARCCATGCANCACCT
Day 1_Max_B	IonXpress_072	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGAAGGCCACAC	MWACGCGARGAACCTTACC	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	CGAAGCGATTG	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	CCTGCCATTTCG	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	CCGAACAATTG	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

**Table A.2. Generalized UniFrac distances of bacterial communities in complete and minimal media over time.** Generalized UniFrac distance contains an extra parameter  $\alpha$  controlling the weight on abundant lineages so the distance is not dominated by highly abundant lineages.  $\alpha = 0.5$  has overall 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

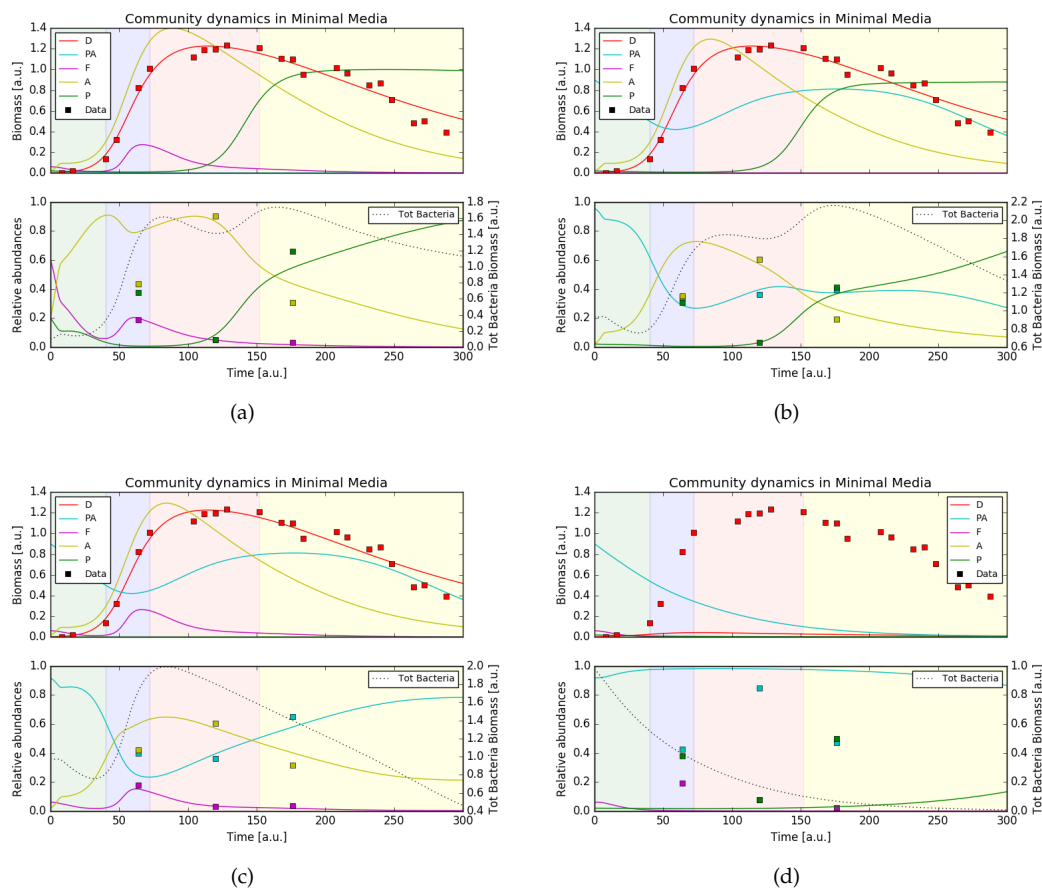
**Appendix Mathematical model simulation results with reduced community**

The simulations are performed in complete and minimal media conditions removing one bacteria from the community.



**Figure B.1.** Simulations of diatom and bacterial community in complete media removing from the community: (a) Pseudoalteromonas; (b) Flavobacterium; (c) Pseudomonas; (d) Alteromonas.





**Figure B.2.** Simulations of diatom and bacterial community in minimal media removing from the community: (a) Pseudoalteromonas; (b) Flavobacterium; (c) Pseudomonas; (d) Alteromonas.

## Appendix Mathematical model additional material

The mathematical model is a system of 13 ODEs describing the variation in time of the populations (cell counts) of

- *P. tricornutum* (*D*);
- *Pseudoalteromonas* (*PA*);
- *Flavobacterium* (*F*);
- *Alteromonas* (*A*);
- *Pseudomonas* (*P*);

and 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* ( $\text{DOC}_{PA}$  and  $\text{DOC}_A$ , respectively);
- the complex polymers (COP) consumed by *F*;
- generic vitamins (Vit) and iron (Fe) needed by *D* and produced by *A*;
- bactericidal molecules (EPA and Bac, produced by *D* and by *PA* respectively);
- the dissolved organic matter (DOM).

The model has 55 unknown free parameters:

- 5 carrying capacities  $CC$ ;
- 34 maximal rates  $v$ ;
- 15 Monod-type coefficients  $K$ ;
- the fraction of  $\text{DOC}_A$ -dependent growth of *A*,  $\epsilon_{\text{DOC}_A}$ .

### Appendix C.1 ODEs system

Five ODEs describe the variation in time of the populations of organism  $O$ , with  $\gamma^O$  and  $\delta^O$  being its growth and death rate:

$$\frac{dD}{dt} = \gamma^D D - \delta^D D \quad (C.1)$$

$$\frac{dPA}{dt} = \gamma^{PA} PA - \delta^{PA} PA \quad (C.2)$$

$$\frac{dF}{dt} = \gamma^F F - \delta^F F \quad (C.3)$$

$$\frac{dA}{dt} = \gamma^A A - \delta^A A \quad (C.4)$$

$$\frac{dP}{dt} = \gamma^P P - \delta^P P \quad (C.5)$$

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

$$\frac{d\text{Vit}}{dt} = v_{\text{Vit}}^{\text{prod}(A)} \gamma^A A - v_{\text{Vit}}^{\text{cons}(D)} \gamma^D D \quad (C.6)$$

$$\frac{d\text{Fe}}{dt} = v_{\text{Fe}}^{\text{prod}(A)} \gamma^A A - v_{\text{Fe}}^{\text{cons}(D)} \gamma^D D \quad (C.7)$$

$$\frac{d\text{DOC}_{PA}}{dt} = v_{\text{DOC}_{PA}}^{\text{prod}(D)} \gamma^D D - v_{\text{DOC}_{PA}}^{\text{cons}(PA)} \gamma^{PA} PA \quad (C.8)$$

$$\frac{d\text{DOC}_A}{dt} = (v_{\text{DOC}_A}^{\text{prod}(D)} + \phi) \gamma^D D - v_{\text{DOC}_A}^{\text{cons}(A)} \gamma^A A \quad (C.9)$$

$$\frac{d\text{COP}}{dt} = (v_{\text{COP}}^{\text{prod}(D)} + \psi) \gamma^D D - v_{\text{COP}}^{\text{cons}(F)} \gamma^F F \quad (C.10)$$

$$\frac{d\text{EPA}}{dt} = v_{\text{EPA}}^{\text{prod}(D)} \gamma^D D - v_{\text{EPA}}^{\text{deg}} \text{EPA} \quad (C.11)$$

$$\frac{d\text{Bac}}{dt} = v_{\text{Bac}}^{\text{prod}(PA)} \gamma^{PA} PA - v_{\text{Bac}}^{\text{deg}} \text{Bac} \quad (C.12)$$

$$\frac{d\text{DOM}}{dt} = v_{\text{DOM}}^{\text{prod}(D)} \gamma^D D - v_{\text{DOM}}^{\text{cons}(A)} \gamma^A A - v_{\text{DOM}}^{\text{cons}(P)} \gamma^P P \quad (C.13)$$

$\phi$  and  $\psi$  are additional terms for  $\text{DOC}_A$  and  $\text{COP}$  production respectively (see Section C.1.1).  $v_J^{\text{deg}}$  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 Monod-type coefficient  $K_J^O$  and eventually maximal rates  $v_J^O$ :

$$\gamma^D = v_{\gamma}^D \cdot \frac{\text{Vit}}{\text{Vit} + K_{\text{Vit}}^D} \frac{\text{Fe}}{\text{Fe} + K_{\text{Fe}}^D} \left(1 - \frac{D}{CC^D}\right) \quad (\text{C.14})$$

$$\delta^D = v_{\delta}^D \frac{1}{1 + \gamma^D} \quad (\text{C.15})$$

$$\gamma^{PA} = v_{\gamma}^{PA} \frac{\text{DOC}_{PA}}{\text{DOC}_{PA} + K_{\text{DOC}_{PA}}^D} \left(1 - \frac{PA}{CC^{PA}}\right) \quad (\text{C.16})$$

$$\delta^{PA} = v_{\delta}^{PA} \left(1 + \frac{v_{\text{EPA}}^{PA} \cdot \text{EPA}}{\text{EPA} + K_{\text{EPA}}^D}\right) \frac{1}{1 + \gamma^{PA}} \quad (\text{C.17})$$

$$\gamma^F = v_{\gamma}^F \frac{\text{COP}}{\text{COP} + K_{\text{COP}}^D} \left(1 - \frac{F}{CC^F}\right) \quad (\text{C.18})$$

$$\delta^F = v_{\delta}^F \left(1 + \frac{v_{\text{Bac}}^F \cdot \text{Bac}}{\text{Bac} + K_{\text{Bac}}^D}\right) \frac{1}{1 + \gamma^F} \quad (\text{C.19})$$

$$\gamma^A = \gamma_{\text{DOC}_A}^A + \gamma_{\text{DOM}}^A \quad (\text{C.20})$$

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

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

$$\delta^A = v_{\delta}^A \left(1 + \frac{v_{\text{Bac}}^A \cdot \text{Bac}}{\text{Bac} + K_{\text{Bac}}^A}\right) \frac{1}{1 + \gamma^A} \quad (\text{C.23})$$

$$\gamma^P = v_{\gamma}^P \frac{\text{DOM}}{\text{DOM} + K_{\text{DOM}}^P} \left(1 - \frac{P}{CC^P}\right) \quad (\text{C.24})$$

$$\delta^P = v_{\delta}^P \left(1 + \frac{v_{\text{Bac}}^P \cdot \text{Bac}}{\text{Bac} + K_{\text{Bac}}^P}\right) \frac{1}{1 + \gamma^P} \quad (\text{C.25})$$

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

#### Appendix C.1.1 $\text{DOC}_A$ and COP production

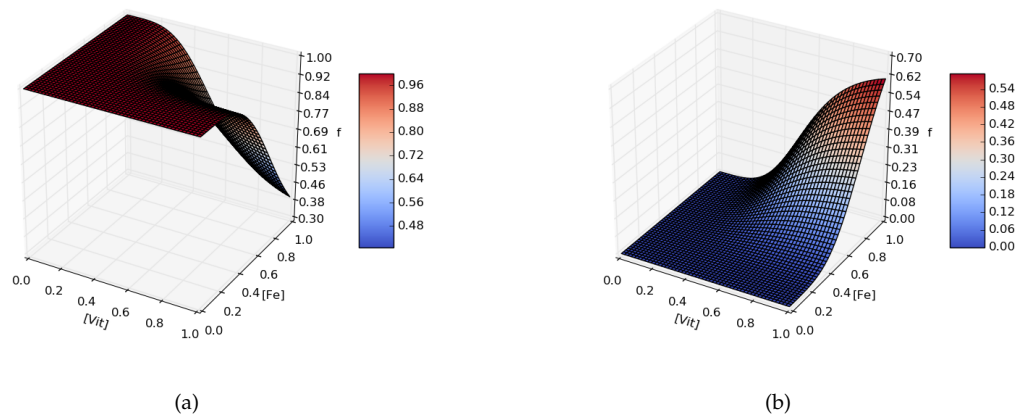
When  $D$  is grown in minimal media conditions, the emergence of  $A$  is observed over  $F$ . From this observation we hypothesise that  $D$  can produce extra organic carbons for either  $A$  or  $F$  depending on the scarcity of micronutrients to favor the growth of  $A$  if more Vit or Fe is needed.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

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

$$\psi = v_{\text{DOC}_A\text{COP}}^D \cdot \xi \quad (\text{C.27})$$

$$\xi = \frac{\text{Vit}^4}{\text{Vit}^4 + K_{\text{Vit}}'^D} \frac{\text{Fe}^4}{\text{Fe}^4 + K_{\text{Fe}}'^D} \quad (\text{C.28})$$

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 C.1).



**Figure C.1.** Example for  $\text{DOC}_A$  ((a),  $1 - \xi$ ) and  $\text{COP}$  ((b),  $\xi$ ) additional production rates dependent on Vit and Fe availability in the media. Here  $K_{\text{Vit}}^D = 0.1$ ,  $K_{\text{Fe}}^D = 0.5$ .

### Appendix C.2 Parameter fitting

The model has 56 parameters, of which 55 are free parameters (see Table C.1). Being a qualitative model, we do not aim at interpreting the absolute parameter values in a biological sense.

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

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

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 C.2.1) 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  $\{\mathcal{P}_i\}^{G_{\max}}$ .

#### Appendix C.2.1 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:

$$s = \sqrt{\sum_t (x_t - X_t)^2} \quad (\text{C.29})$$

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:

$$s = \sum_t \sqrt{\sum_o \left(1 - e^{\frac{r_{o_t} - R_{o_t}}{r_{o_t}}}\right)^2} \quad (\text{C.30})$$

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.

**Table C.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 losing information on differences among MM and CM conditions. Only average values, and not experimental errors, are taken into account.

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 C.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.

t	Complete Media				Minimal Media			
	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

## Appendix C.2.2 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. C.29 is used. For B-type fits, the fitness score of Eq. C.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, 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  $\mathcal{P}(D)_{MM2} \pm 5\%$ ,  $\mathcal{P}(A, F, P, PA, deg)_{MM2} \pm 80\%$

The last rounds of fitting were run on different sets of parameters considered equally good. The final parameter sets  $\mathcal{P}$  are presented in Table C.4.

**Table C.4.** Final parameter sets used for simulation in CM ( $\mathcal{P}_{CM}$ ) and MM ( $\mathcal{P}_{MM}$ ). Also reported are the overall average and standard deviation values from all the last rounds of fitting.

	$\mathcal{P}_{CM}$	$\mu(\mathcal{P}_{CM})$	$\sigma(\mathcal{P}_{CM})$	$\mathcal{P}_{MM}$	$\mu(\mathcal{P}_{MM})$	$\sigma(\mathcal{P}_{MM})$
$K_{Bac}^A$	0.562780	0.476829	0.235751	0.23821	0.281329	0.257813
$K_{DOC_A}^A$	0.463690	0.249183	0.152283	0.02253	0.332969	0.365116
$K_{DOM}^A$	1.043490	0.526842	0.339699	0.82552	0.671145	0.475618
$v_{Bac}^A$	1.884690	1.433304	0.495578	0.94702	1.085873	0.889537
$CC^A$	1.230920	1.112280	0.351905	2.74047	1.564099	0.668637
$v_{\delta}^A$	0.036310	0.067159	0.085177	0.01697	0.058289	0.113161
$v_{DOC_A}^{cons(A)}$	0.504220	0.504324	0.149340	1.30073	0.806015	0.315439
$\epsilon_{DOC_A}^{cons(A)}$	0.204470	0.464587	0.226651	0.99257	0.670187	0.250496
$v_{DOM}^{cons(A)}$	0.186340	0.493623	0.298676	0.74598	0.349684	0.246408
$\epsilon_{DOM}^{cons(A)}$	0.795530	0.508826	0.229097	0.00743	0.323377	0.248408
$v_{DOM}^{prod(A)}$	0.030470	0.052871	0.094576	0.06702	0.130514	0.176652
$v_{Fe}^{prod(A)}$	0.134290	0.118939	0.028298	0.19948	0.206115	0.091565
$v_{\gamma}^A$	0.329520	0.434683	0.340005	0.34841	1.037434	0.510106
$v_{Vit}^{prod(A)}$	0.954240	0.644251	0.317607	1.09226	1.241136	0.618984
$v_{Bac}^{deg}$	0.108110	0.409152	0.213390	0.07769	0.263422	0.330743
$v_{EPA}^{deg}$	0.350050	0.373353	0.117858	0.57995	0.784058	0.412260
$K_{Fe}^D$	0.488680	0.583597	0.358354	0.02979	0.124157	0.134217
$K_{Fe}^D$	1.199730	1.048777	0.343732	0.33321	0.486298	0.256807
$K_{Vit}^D$	0.844900	0.645544	0.226955	0.46274	0.346839	0.088571
$K_{Vit}^D$	0.469600	0.903463	0.501364	0.09782	0.339058	0.356208
$v_{DOC_A}^D$	0.314330	0.853140	0.557108	0.36480	0.546552	0.330353
$CC^D$	1.875200	1.584920	0.515701	1.57897	1.427444	0.420945
$v_{COP}^{prod(D)}$	1.005110	1.268507	0.463020	0.70666	0.736754	0.370734
$v_{\delta}^D$	0.007180	0.016960	0.051891	0.00681	0.013765	0.049375
$v_{DOC_A}^{prod(D)}$	1.770740	0.987437	0.490546	1.65657	1.350378	0.421454
$v_{DOC_{PA}}^{prod(D)}$	1.055270	0.990547	0.415673	0.83897	1.081959	0.591094
$v_{DOM}^{prod(D)}$	0.135980	0.653181	0.351953	0.54133	0.565364	0.231813
$v_{EPA}^{prod(D)}$	1.214350	0.899207	0.360058	1.28659	1.070999	0.478554
$v_{Fe}^{cons(D)}$	0.665740	0.755699	0.241111	0.31684	0.363974	0.099207
$v_{\gamma}^D$	0.194310	0.200395	0.069030	0.52737	0.562459	0.149546

Continued on next page

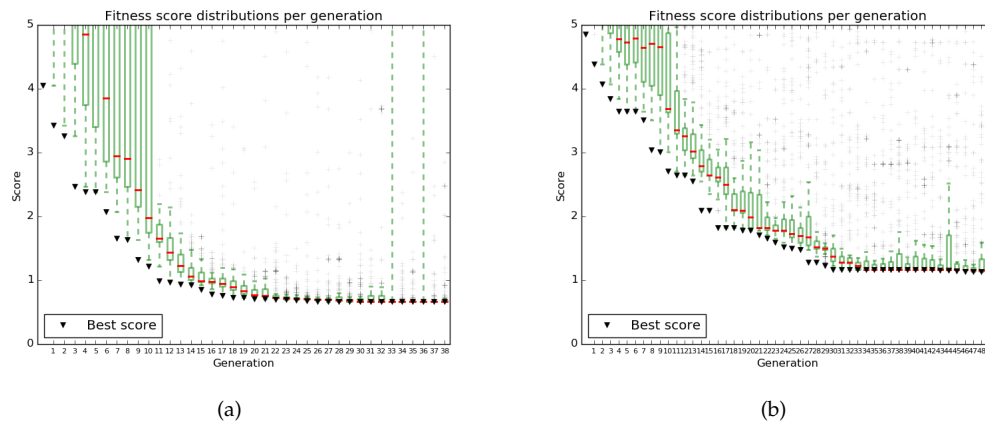
Table C.4 – Continued from previous page

	$\mathcal{P}_{CM}$	$\mu(\mathcal{P}_{CM})$	$\sigma(\mathcal{P}_{CM})$	$\mathcal{P}_{MM}$	$\mu(\mathcal{P}_{MM})$	$\sigma(\mathcal{P}_{MM})$
$v_{Vit}^{cons(D)}$	0.367880	0.566566	0.416514	1.78450	0.909564	0.404000
$K_{Bac}^F$	0.936420	0.583317	0.263447	0.16731	0.299761	0.198921
$K_{COP}$	0.477700	0.588674	0.234155	0.74525	0.451922	0.315203
$v_{Bac}^F$	0.184360	0.311845	0.105780	0.23234	1.237169	1.012458
$CC^F$	1.351050	1.206888	0.384417	0.54187	1.074951	0.758107
$v_{COP}^{cons(F)}$	0.139320	0.175972	0.045416	0.57005	0.330531	0.127086
$v_{\delta}^F$	0.382820	0.318181	0.144775	0.18005	0.200895	0.150824
$v_{DOM}^{prod(F)}$	0.092860	0.080066	0.074871	0.00984	0.135875	0.283818
$v_{\gamma}^F$	0.765450	0.726578	0.223690	1.50156	0.888556	0.545041
$K_{Bac}^P$	0.020100	0.148399	0.132143	0.16823	0.326145	0.294570
$K_{DOM}^P$	0.609800	0.560853	0.171693	1.12080	0.688621	0.413129
$v_{Bac}^P$	1.009740	1.238831	0.430709	2.11081	1.419892	0.958683
$CC^P$	1.301320	1.277678	0.407513	1.17750	2.585117	0.869802
$v_{\delta}^P$	0.020440	0.069033	0.167148	0.01591	0.036821	0.150349
$v_{DOM}^{cons(P)}$	0.698330	0.523151	0.203572	0.17625	0.124345	0.136063
$v_{DOM}^{prod(P)}$	0.195450	0.189091	0.103414	0.03107	0.116528	0.203261
$v_{\gamma}^P$	0.820720	0.440066	0.249502	0.57938	0.527859	0.284980
$K_{DOC_{PA}}$	0.245720	0.351941	0.221873	0.42128	0.564689	0.489764
$K_{EPA}$	0.755570	0.541606	0.297743	0.05329	0.404319	0.414359
$v_{EPA}^{PA}$	1.577050	1.484135	0.474372	2.65508	1.368551	0.892632
$v_{Bac}^{prod(PA)}$	0.819580	0.848959	0.264185	0.28618	0.568550	0.438164
$CC^{PA}$	0.995130	1.029216	0.323852	1.28138	1.477045	0.533872
$v_{\delta}^{PA}$	0.221040	0.284309	0.181709	0.01861	0.052638	0.102986
$v_{DOC_{PA}}^{cons(PA)}$	0.236820	0.254966	0.144860	0.41130	0.249994	0.182493
$v_{DOM}^{prod(PA)}$	0.130620	0.110548	0.125387	0.01816	0.108930	0.154334
$v_{\gamma}^{PA}$	0.327430	0.468045	0.287832	0.12769	0.350329	0.210662

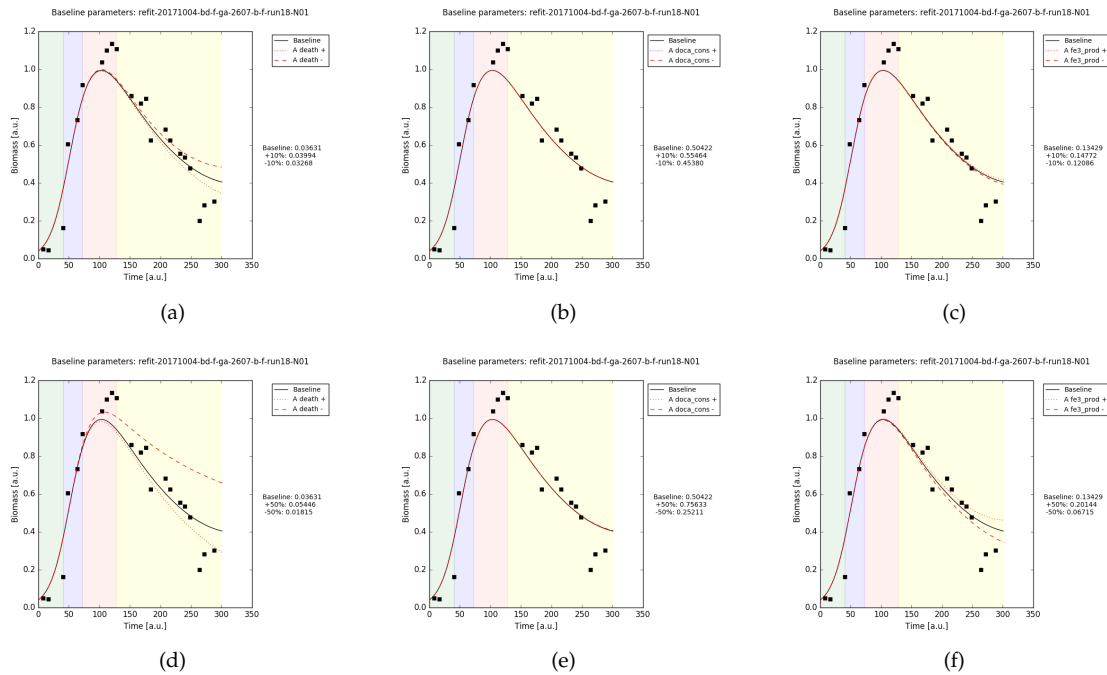
Appendix C.2.3 Sanity checks of the parameter fits

The parameters of the algorithm were chosen to obtain a satisfactory convergence of the fit (Figure C.2).

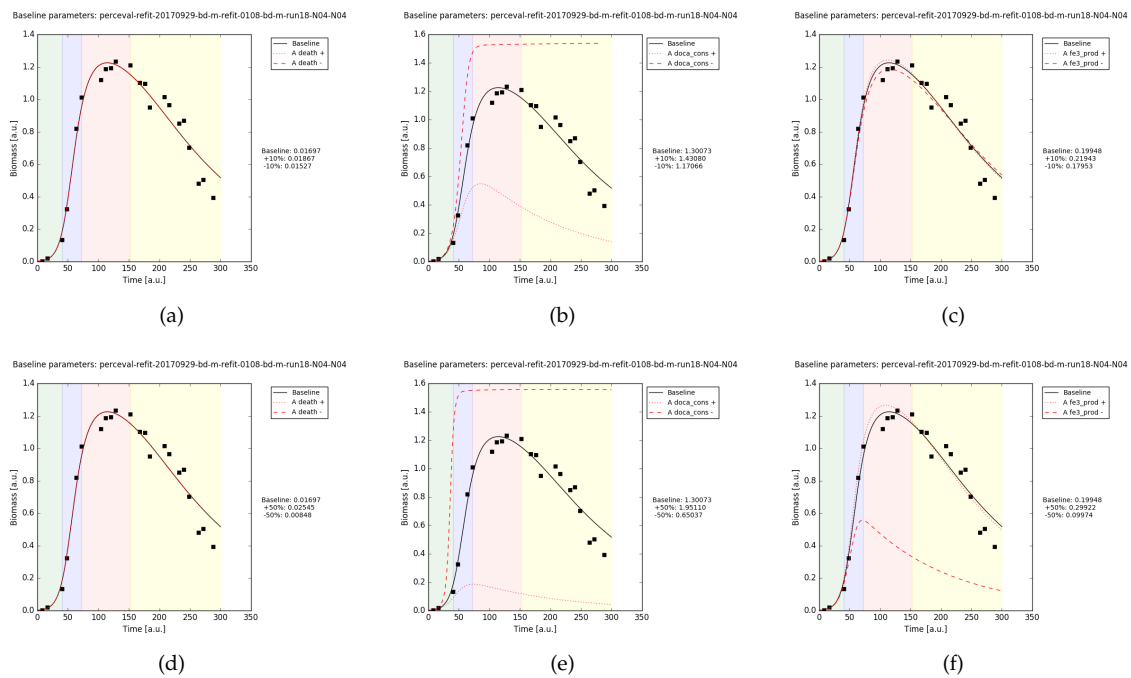
We checked the effect of varying the parameters  $\delta^A$ ,  $v_{DOC_A}^{cons(A)}$  and  $v_{Fe}^{prod(A)}$  (the only bacterial parameters observed to influence the biomass growth curve in CM) by  $\pm 10\%$  and  $\pm 50\%$ . The diatom growth is almost insensitive to these variations in CM (Figure C.3), while it shows greater effects in MM (Figure C.4).



**Figure C.2.** Distribution of fitness scores in populations over generations for the genetic algorithm runs chosen to perform the last fitting iteration in minimal (a) and complete (b) media conditions.

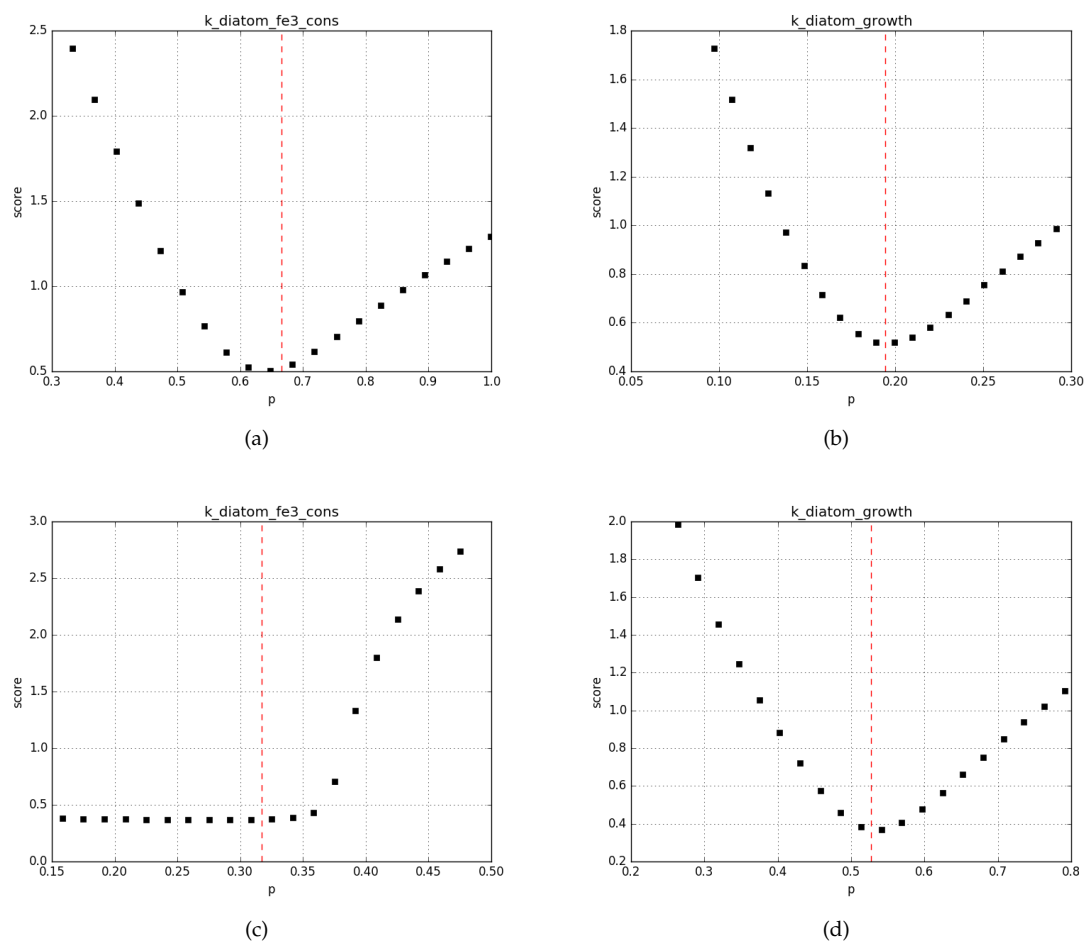


**Figure C.3.** Diatom growth in CM simulation results. The parameters  $\delta^A$ ,  $v_{\text{DOC}_A}^{\text{cons}(A)}$  and  $v_{\text{Fe}}^{\text{prod}(A)}$ , are varied by  $\pm 10\%$  (a, b, c respectively) and by  $\pm 50\%$  (d, e, f respectively).



**Figure C.4.** Diatom growth in MM simulation results. The parameters  $\delta^A$ ,  $v_{\text{DOC}_A}^{\text{cons}(A)}$  and  $v_{\text{Fe}}^{\text{prod}(A)}$ , are varied by  $\pm 10\%$  (a, b, c respectively) and by  $\pm 50\%$  (d, e, f respectively).

Parameter profiling shows that the algorithm correctly converges towards local minima and that in general those minima are rather stable to perturbation  $p \pm 50\%$ . Figure C.5 shows examples of the most unstable profiles from this first set of fits.



**Figure C.5.** Profiling of the parameters  $v_{\text{Fe}}^{\text{cons}(D)}$  in CM (a) and MM (c) and  $\gamma^D$  in CM (b) and MM (e). The red line shows the value chosen by the fit.

## Bibliography

1. Chisti, Y. Biodiesel from microalgae. *Biotechnology Advances* **2007**, *25*, 294–306.
2. Moejes, F.W.; Moejes, K.B. Algae for Africa: Microalgae as a source of food, feed and fuel in Kenya. *African Journal of Biotechnology* **2017**, *16*, 288–301.
3. Ryther, J.; Goldman, J. Microbes as food in mariculture. *Annual Reviews in Microbiology* **1975**, *29*, 429–443.
4. Tredici, M.R.; Biondi, N.; Ponis, E.; Rodolfi, L.; Zittelli, G.C.; Burnell, G.; Allan, G. Advances in microalgal culture for aquaculture feed and other uses. In *New Technologies in Aquaculture: Improving Production Efficiency, Quality and Environmental Management*; Burnell, G.; Allan, G., Eds.; Woodhead Publishing Ltd: Cambridge, 2009; pp. 610–676.
5. Reboloso-Fuentes, M.; Navarro-Pérez, A.; Ramos-Miras, J.; Guil-Guerrero, J. Biomass nutrient profiles of the microalga *Phaeodactylum tricornutum*. *Journal of Food Biochemistry* **2001**, *25*, 57–76.
6. Kates, M.; Volcani, B.E. Lipid components of diatoms. *Biochimica et biophysica acta* **1966**, *116*, 264–278.
7. Fajardo, A.R.; Cerdán, L.E.; Medina, A.R.; Fernández, F.G.A.; Moreno, P.A.G.; Grima, E.M. Lipid extraction from the microalga *Phaeodactylum tricornutum*. *European Journal of Lipid Science and Technology* **2007**, *109*, 120–126.
8. Siron, R.; Giusti, G.; Berland, B. Changes in the fatty acid composition of *Phaeodactylum tricornutum* and *Dunaliella tertiolecta* during growth and under phosphorus deficiency. *Marine Ecology Progress Series* **1989**, *55*, 95–100.

9. Yashodhara, B.M.; Umakanth, S.; Pappachan, J.M.; Bhat, S.K.; Kamath, R.; Choo, B.H. Omega-3 fatty acids: a comprehensive review of their role in health and disease. *Postgraduate Medical Journal* **2009**, *85*, 84–90.
10. Francius, G.; Tesson, B.; Dague, E.; Martin-jézéquel, V.; Dufrêne, Y.F. Nanostructure and nanomechanics of live *Phaeodactylum tricornutum* morphotypes. *Environmental microbiology* **2008**, *10*, 1344–56.
11. Santaefemia, S.; Torres, E.; Mera, R.; Abalde, J. Bioremediation of oxytetracycline in seawater by living and dead biomass of the microalga *Phaeodactylum tricornutum*. *Journal of Hazardous Materials* **2016**, *320*, 315–325.
12. Mata, T.; Martins, A.; Caetano, N. Microalgae for biodiesel production and other applications: a review. *Renewable and Sustainable Energy Reviews* **2010**, *14*, 217–232.
13. Grima, E.; Fernández, F. Photobioreactors: light regime, mass transfer, and scaleup. *Journal of Biotechnology* **1999**, *70*, 231–247.
14. Parmar, A.; Singh, N.; Pandey, A. Cyanobacteria and microalgae: a positive prospect for biofuels. *Bioresource Technology* **2011**, *102*, 10163–10172.
15. Day, J.G.; Thomas, N.J.; Achilles-Day, U.E.M.; Leakey, R.J.G. Early detection of protozoan grazers in algal biofuel cultures. *Bioresource Technology* **2012**, *114*, 715–719.
16. Wang, H.; Zhang, W.; Chen, L.; Wang, J.; Liu, T. The contamination and control of biological pollutants in mass cultivation of microalgae. *Bioresource Technology* **2013**, *128*, 745–750.
17. Kazamia, E.; Aldridge, D.C.; Smith, A.G. Synthetic ecology – A way forward for sustainable algal biofuel production? *Journal of Biotechnology* **2012**, *162*, 163–169.
18. Gause, G. *The Struggle for Existence*; 1934; p. 163.
19. Hardin, G. The competitive exclusion principle. *Science* **1960**, *131*, 1292–7.
20. Dykhuizen, D. Santa Rosalia revisited: why are there so many species of bacteria? *Antonie van Leeuwenhoek* **1998**, *73*, 25–33.
21. Hagstrom, G.I.; Levin, S.A. Marine Ecosystems as Complex Adaptive Systems: Emergent Patterns, Critical Transitions, and Public Goods. *Ecosystems* **2017**, *20*, 458–476.
22. Widder, S.; Allen, R.J.; Pfeiffer, T.; Curtis, T.P.; Wiuf, C.; Sloan, W.T.; Cordero, O.X.; Brown, S.P.; Momeni, B.; Shou, W.; Kettle, H.; Flint, H.J.; Haas, A.F.; Laroche, B.; Kreft, J.U.; Rainey, P.B.; Freilich, S.; Schuster, S.; Milferstedt, K.; van der Meer, J.R.; Großkopf, T.; Huisman, J.; Free, A.; Picioreanu, C.; Quince, C.; Klapper, I.; Labarthe, S.; Smets, B.F.; Wang, H.; Soyer, O.S. Challenges in microbial ecology: building predictive understanding of community function and dynamics. *The ISME Journal* **2016**, *10*, 2557–2568.
23. Song, H.S.; Cannon, W.; Beliaev, A.; Konopka, A. Mathematical Modeling of Microbial Community Dynamics: A Methodological Review. *Processes* **2014**, *2*, 711–752.
24. Zomorodi, A.R.; Segrè, D. Synthetic Ecology of Microbes: Mathematical Models and Applications. *Journal of molecular biology* **2015**, *428*, 837–861.
25. Succurro, A.; Moejes, F.W.; Ebenhöf, O. A Diverse Community To Study Communities: Integration of Experiments and Mathematical Models To Study Microbial Consortia. *J Bacteriol* **2017**, *199*, e00865–16.
26. Lotka, A. *Elements of Physical Biology*; 1925.
27. Volterra, V. Fluctuations in the abundance of a species considered mathematically. *Nature* **1926**, *118*, 558–560.
28. Hofbauer, J.; Hutson, V.; Jansen, W. Coexistence for systems governed by difference equations of Lotka-Volterra type. *Journal of mathematical biology* **1987**, *25*, 553–70.
29. Berry, D.; Widder, S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Frontiers in Microbiology* **2014**, *5*, 219.
30. Amin, S.A.; Parker, M.S.; Armbrust, E.V. Interactions between diatoms and bacteria. *Microbiology and molecular biology reviews : MMBR* **2012**, *76*, 667–84.
31. Provasoli, L. Nutrition and ecology of Protozoa and Algae. *Annual review of microbiology* **1958**, *12*, 279–308.
32. Delucca, R.; McCracken, M.D. Observations on interactions between naturally-collected bacteria and several species of algae. *Hydrobiologia* **1977**, *55*, 71–75.
33. Suminto.; Hirayama, K. Application of a growth-promoting bacteria for stable mass culture of three marine microalgae. *Hydrobiologia* **1997**, *358*, 223–230.



34. Bruckner, C.G.; Rehm, C.; Grossart, H.P.; Kroth, P.G. Growth and release of extracellular organic compounds by benthic diatoms depend on interactions with bacteria. *Environmental Microbiology* **2011**, *13*, 1052–1063.
35. Amin, S.A.; Hmelo, L.R.; van Tol, H.M.; Durham, B.P.; Carlson, L.T.; Heal, K.R.; Morales, R.L.; Berthiaume, C.T.; Parker, M.S.; Djunaedi, B.; Ingalls, A.E.; Parsek, M.R.; Moran, M.a.; Armbrust, E.V. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **2015**, *522*, 98–101.
36. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glockner, F.O. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **2013**, *41*, D590–D596.
37. Chen, J.; Bittinger, K.; Charlson, E.S.; Hoffmann, C.; Lewis, J.; Wu, G.D.; Collman, R.G.; Bushman, F.D.; Li, H. Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* **2012**, *28*, 2106–2113.
38. Ivanova, E.P.; Flavier, S.; Christen, R. 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. *International Journal of Systematic and Evolutionary Microbiology* **2004**, *54*, 1773–1788.
39. Rosenberg, E.; DeLong, E.F.; Lory, S.; Stackebrandt, E.; Thompson, F. *The Prokaryotes*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2014.
40. Bowman, J. Bioactive compound synthetic capacity and ecological significance of marine bacterial genus Pseudoalteromonas. *Marine drugs* **2007**, *5*, 220–241.
41. Lee, S.o.; Kato, J.; Takiguchi, N.; Kuroda, A.; Ikeda, T. Involvement of an Extracellular Protease in Algicidal Activity of the Marine Bacterium Pseudoalteromonas sp. Strain A28. *Applied and Environmental Microbiology* **2000**, *66*, 4334–4339.
42. Venkateswaran, K.; Dohmoto, N. Pseudoalteromonas peptidolytica sp. nov., a novel marine mussel-thread-degrading bacterium isolated from the Sea of Japan. *International Journal of Systematic and Evolutionary Microbiology* **2000**, *50*, 565–574.
43. Chen, X.; Xie, B.; Lu, J.; He, H.; Zhang, Y. A novel type of subtilase from the psychrotolerant bacterium Pseudoalteromonas sp. SM9913: catalytic and structural properties of deseasin MCP-01. *Microbiology* **2007**, *153*, 2116–2125.
44. Khudary, R.A.; Venkatachalam, R.; Katzer, M. A cold-adapted esterase of a novel marine isolate, Pseudoalteromonas arctica: gene cloning, enzyme purification and characterization. *Extremophiles* **2010**, *14*, 273–285.
45. Lu, M.; Wang, S.; Fang, Y.; Li, H.; Liu, S.; Liu, H. Cloning, expression, purification, and characterization of cold-adapted  $\alpha$ -amylase from Pseudoalteromonas arctica GS230. *The protein journal* **2010**, *29*, 591–597.
46. Albino, A.; Marco, S.; Maro, A.D. Characterization of a cold-adapted glutathione synthetase from the psychrophile Pseudoalteromonas haloplanktis. *Molecular BioSystems* **2012**, *8*, 2405–2414.
47. He, H.; Guo, J.; Chen, X.; Xie, B.; Zhang, X. Structural and functional characterization of mature forms of metalloprotease E495 from Arctic sea-ice bacterium Pseudoalteromonas sp. SM495. *PloS one* **2012**, *7*, e35442.
48. Holmström, C.; Kjelleberg, S. Marine Pseudoalteromonas species are associated with higher organisms and produce biologically active extracellular agents. *FEMS Microbiology Ecology* **1999**, *30*, 285–293.
49. Desbois, A.P.; Mearns-Spragg, A.; Smith, V.J. A Fatty Acid from the Diatom Phaeodactylum tricornutum is Antibacterial Against Diverse Bacteria Including Multi-resistant Staphylococcus aureus (MRSA). *Marine Biotechnology* **2009**, *11*, 45–52.
50. LPSN. List of Prokaryotic names with Standing in Nomenclature - Alteromonadaceae, 2016.
51. Ivanova, E.P.; Mikhailov, V.V. A new family of Alteromonadaceae fam. nov., including the marine proteobacteria species Alteromonas, Pseudoalteromonas, Idiomarina i Colwellia. *Mikrobiologiya* **2001**, *70*, 15–23.
52. Amin, S.A.; Green, D.H.; Hart, M.C.; Küpper, F.C.; Sunda, W.G.; Carrano, C.J. Photolysis of iron, siderophore chelates promotes bacterial, algal mutualism. *Proceedings of the National Academy of Sciences* **2009**, *106*, 17071–17076.

53. Reid, R.; Butler, A. Investigation of the mechanism of iron acquisition by the marine bacterium *Alteromonas luteoviolaceus*: Characterization of siderophore production. *Limnology and oceanography* **1991**, *36*, 1783–1792.
54. Holt, P.D.; Reid, R.R.; Lewis, B.L.; Luther, G.W.; Butler, A. Iron(III) coordination chemistry of alterobactin A: A siderophore from the marine bacterium *Alteromonas luteoviolacea*. *Inorganic Chemistry* **2005**, *44*, 7671–7677.
55. Vraspir, J.M.; Butler, A. Chemistry of marine ligands and siderophores. *Annual review of marine science* **2009**, *1*, 43–63.
56. Bruland, K.W.; Donat, J.R.; Hutchins, D.a. Interactive influences of bioactive trace metals on biological production in oceanic waters. *Limnology and Oceanography* **1991**, *36*, 1555–1577.
57. Morel, F.M.M.; Price, N.M. The biogeochemical cycles of trace metals in the oceans. *Science* **2003**, *300*, 944–7.
58. LPSN. List of Prokaryotic names with Standing in Nomenclature - Flavobacteriaceae, 2016.
59. Yoon, J.; Jo, Y.; Kim, G.J.; Choi, H. *Gramella lutea* sp. nov., a Novel Species of the Family Flavobacteriaceae Isolated from Marine Sediment. *Current microbiology* **2015**, *71*, 252–8.
60. Glöckner, F.O.; Fuchs, B.M.; Amann, R. Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. *Applied and environmental microbiology* **1999**, *65*, 3721–6.
61. Abell, G.; Bowman, J. Ecological and biogeographic relationships of class Flavobacteria in the Southern Ocean. *FEMS Microbiology Ecology* **2005**, *51*, 265–277.
62. DeLong, E.F.; Preston, C.M.; Mincer, T.; Rich, V.; Hallam, S.J.; Frigaard, N.U.; Martinez, A.; Sullivan, M.B.; Edwards, R.; Brito, B.R.; Chisholm, S.W.; Karl, D.M. Community genomics among stratified microbial assemblages in the ocean's interior. *Science (New York, N.Y.)* **2006**, *311*, 496–503.
63. Manz, W.; Amann, R.; Ludwig, W.; Vancanneyt, M.; Schleifer, K.H. 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* **1996**, *142*, 1097–1106.
64. Kirchman, D.L. The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS microbiology ecology* **2002**, *39*, 91–100.
65. Cottrell, M.T.; Kirchman, D.L. Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacter cluster consuming low- and high-molecular-weight dissolved organic matter. *Applied and environmental microbiology* **2000**, *66*, 1692–7.
66. Pinhassi, J.; Sala, M.M.; Havskum, H.; Peters, F.; Guadayol, Ò.; Malits, A.; Marrasé, C. Changes in bacterioplankton composition under different phytoplankton regimens. *Applied and Environmental Microbiology* **2004**, *70*, 6753–6766.
67. Smayda, T. Harmful algal blooms: their ecophysiology and general relevance to phytoplankton blooms in the sea. *Limnology and oceanography* **1997**, *42*, 1137–1153.
68. Starr, M.P.; Stolp, H.; Trüper, H.G.; Balows, A.; Schlegel, H.G. *The Prokaryotes*; Springer Berlin Heidelberg: Berlin, Heidelberg, 1981.
69. Anzai, Y.; Kim, H.; Park, J.Y.; Wakabayashi, H.; Oyaizu, H. Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. *International Journal of Systematic and Evolutionary Microbiology* **2000**, *50*, 1563–89.
70. Isnansetyo, A.; Kamei, Y. Bioactive substances produced by marine isolates of *Pseudomonas*. *Journal of industrial microbiology & biotechnology* **2009**, *36*, 1239–48.
71. Soria-Dengg, S.; Reissbrodt, R.; Horstmann, U. Siderophores in marine coastal waters and their relevance for iron uptake by phytoplankton: experiments with the diatom *Phaeodactylum tricornutum*. *Mar. Ecol. Prog. Ser.* **2001**, *220*, 73–82.
72. Soria-Dengg, S.; Horstmann, U. Ferrioxamines B and E as iron sources for the marine diatom *Phaeodactylum tricornutum*. *Marine Ecology Progress Series* **1995**, *127*, 269–277.
73. Provasoli, L. Organic regulation of phytoplankton fertility. In *The sea: ideas and observations on progress in the study of the seas*; Hill, M., Ed.; Wiley-Interscience: New York, 1963; pp. 165–219.
74. Provasoli, L.; Carlucci, A. Vitamins and growth regulators. In *Algal Physiology and Biochemistry, Botanical monographs*, 10; Stewart, W., Ed.; Blackwell Scientific Publications: California, 1974; pp. 741–787.
75. Croft, M.T.; Warren, M.J.; Smith, A.G. Algae need their vitamins. *Eukaryotic Cell* **2006**, *5*, 1175–1183.

76. Yongmanitchai, W.; Ward, O.P. Growth of and omega-3 fatty acid production by *Phaeodactylum tricornutum* under different culture conditions. *Applied and Environmental Microbiology* **1991**, *57*, 419–425.
77. Sañudo-Wilhelmy, S.A.; Gómez-Consarnau, L.; Suffridge, C.; Webb, E.A. The Role of B Vitamins in Marine Biogeochemistry. *Annual Review of Marine Science* **2014**, *6*, 339–367.
78. Azam, F.; Fenchel, T.; Field, J.; Gray, J.; Meyer-Reil, L.; Thingstad, F. The Ecological Role of Water-Column Microbes in the Sea. *Marine Ecology Progress Series* **1983**, *10*, 257–263, [3294].
79. Stocker, R. Marine Microbes See a Sea of Gradients. *Science* **2012**, *338*, 628–633.
80. Persson, G.; Jansson, M.; Kluwer, C. Phosphate uptake and utilization by bacteria and algae. *Hydrobiologia* **1988**, *170*, 177–189.
81. Bell, W.; Mitchell, R. Chemotactic and growth responses of marine bacteria to algal extracellular products. *The Biological Bulletin* **1972**, *143*, 265–277.
82. MacArthur, R. Fluctuations of animal populations and a measure of community stability. *Ecology* **1955**, *36*, 533–536.
83. Gardner, M.; Ashby, W. Connectance of large dynamic (cybernetic) systems: critical values for stability. *Nature* **1970**, *228*, 784.
84. Pimm, S. The complexity and stability of ecosystems. *Nature* **1984**, *307*, 321–326.
85. Elton, C. *The Ecology of Invasions by Animals and Plants*; 1958.
86. McCann, K.S. The diversity-stability debate. *Nature* **2000**, *405*, 228–33.
87. Zelezniak, A.; Andrejev, S.; Ponomarova, O.; Mende, D.R.; Bork, P.; Patil, K.R. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc Natl Acad Sci U S A* **2015**, *112*, 6449–54.
88. Verhulst, P. Notice sur la loi que la population suit dans son accroissement. *Correspondance mathématique et physique de l'Observatoire de Bruxelles* **1838**, *10*, 113–121.
89. Horgan, R.P.; Kenny, L.C. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist* **2011**, *13*, 189–195.
90. Gachon, C.M.; Day, J.G.; Campbell, C.N.; Pröschold, T.; Saxon, R.J.; Küpper, F.C. The Culture Collection of Algae and Protozoa (CCAP): A biological resource for protistan genomics. *Gene* **2007**, *406*, 51 – 57. Marine Genomics.
91. Moejes, F.W. Dynamics of the bacterial community associated with *Phaeodactylum tricornutum* cultures: a novel approach to scaling up microalgal cultures. PhD thesis, Heinrich Heine University Dusseldorf, 2016.
92. Guillard, R.; Ryther, J. Studies of marine planktonic diatoms: I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (Cleve) Gran. *Canadian journal of microbiology* **1962**, *8*, 229–239.
93. Guillard, R. Culture of phytoplankton for feeding marine invertebrates. In *Culture of marine invertebrate animals*; W. L. Smith and M. H. Chanley., Ed.; Plenum Press: New York, 1975; pp. 29–60.
94. de Gouvion Saint Cyr, D.; Wisniewski, C.; Schrive, L.; Farhi, E.; Rivasseau, C. Feasibility study of microfiltration for algae separation in an innovative nuclear effluents decontamination process. *Separation and Purification Technology* **2014**, *125*, 126–135.
95. Quail, M.; Smith, M.E.; Coupland, P.; Otto, T.D.; Harris, S.R.; Connor, T.R.; Bertoni, A.; Swerdlow, H.P.; Gu, Y. A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics* **2012**, *13*, 341.
96. Grada, A.; Weinbrecht, K. Next-Generation Sequencing: Methodology and Application. *Journal of Investigative Dermatology* **2013**, *133*, e11–4.
97. Pylro, V.S.; Roesch, L.F.W.; Morais, D.K.; Clark, I.M.; Hirsch, P.R.; Tótola, M.R. Data analysis for 16S microbial profiling from different benchtop sequencing platforms. *Journal of Microbiological Methods* **2014**, *107*, 30–37.
98. Aronesty, E. Command-line tools for processing biological sequencing data, 2011.
99. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461.
100. Edgar, R.C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods* **2013**, *10*, 996–8.
101. Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Peña, A.G.; Goodrich, J.K.; Gordon, J.I.; Huttley, G.A.; Kelley, S.T.; Knights, D.; Koenig, J.E.; Ley, R.E.; Lozupone, C.A.; McDonald, D.; Muegge, B.D.; Pirrung, M.; Reeder, J.; Sevinsky, J.R.; Turnbaugh, P.J.; Walters,

- W.a.; Widmann, J.; Yatsunenko, T.; Zaneveld, J.; Knight, R. QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **2010**, *7*, 335–336.
102. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. **2015**, pp. <http://www.R-project.org/>.
103. Mitchell, M. *An introduction to genetic algorithms*; MIT Press: Cambridge, MA, 1996.
104. Lozupone, C.; Knight, R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology* **2005**, *71*, 8228–8235, [<http://aem.asm.org/content/71/12/8228.full.pdf+html>].