The Microbiome of Bioreactors Containing Mass-Cultivated Marine Diatoms for Industrial Carbon Capture and Utilization

Nerea Johanna Aalto[1,2], Ingeborg Hulda Giæver[1], Gunilla Kristina Eriksen[1], Linn Israelsen[1], Stina Krsmanovic[1], Sebastian Petters[1], Hans C. Bernstein[1,2]*

Affiliations:

1. Norwegian College of Fisheries Sciences, UiT - The Arctic University of Norway, Tromsø, Norway

2. The Arctic Centre for Sustainable Energy, UiT - The Arctic University of Norway, Tromsø, Norway

***Correspondence**: Hans C. Bernstein, UiT-The Arctic University of Norway, BFE-fak., Postboks 6050 Langnes, 9037 Tromsø, NORWAY*; Phone: +47 776 46 114; Email:* Hans.C.Bernstein@uit.no

Running Title: Industrial microbial communities for decarbonization

Keywords: *Algae, Mass cultivation, Phycosphere, Community Succession, Porosira glacialis*

Abstract

Marine microalgae are a promising innovation platform for carbon capture and utilization (CCU) biotechnologies to mitigate industrial greenhouse gas emissions. However, industrial-scale cultivation of algal mono-cultures is challenging and often unscalable. Non-axenic microalgae in large semi-open photobioreactors lead to the co-cultivation of diverse microbial communities. There is limited knowledge about the "bioreactor ecology" involving microalgae interacting with the microbiome and its subsequent impact on process stability and productivity. In this study, we describe the semi-continuous industrial mass cultivation of the cold-adapted marine diatom, *Porosira glacialis* UiT201, by investigating the prokaryotic and microeukaryotic (phytoplankton and heterotrophic protist) communities. Data were collected in two consecutive time series experiments, representing the initiation and operation of an preindustrial scale CCU photobioreactor (300,000 liters). The first experiment experienced a culture "crash" of the focal strain after 39 days, while the second culture remained stable and "healthy" for 60 days. The results highlight that this mass cultivation system represents a unique industrial marine microbial ecosystem. The succession of the prokaryotic community was primarily driven by species replacement, indicating turnover due to selective bioreactor conditions and/or biological interactions. Nonetheless, the bioreactor consistently harbors a recurring and abundant core microbiome, suggesting that the closely associated bacterial community is influenced by microalgae-specific properties and can endure a dynamic and variable environment. The observed culture collapse of *P. glacialis* coincided with changes in the core microbiome structure and different environmental growth conditions compared to the stable and "healthy" experiment. These findings imply that cohabiting microbial taxa within industrial microalgae cultivation likely play a critical role in stabilizing the conversion of industrial CO_2 into marine biomass, and changes in community structure serve as an indicator of process stability.

1 Introduction

Growing levels of atmospheric carbon have an impact on biodiversity, human health, the economy, and the global geopolitical landscape. As the world's climate nears new tipping points [1], it is imperative that we not only reduce emissions but also take new steps toward carbon capture and utilization (CCU). Innovative biotechnologies that capitalize on photosynthetic carbon fixation, which naturally controls global carbon cycles, will contribute significantly to the solution [2]. Marine microalgae are specifically promising because they have a much higher carbon fixation capacity than plants. Living microalgae make up just 1-2% of global biomass, but they account for almost 50% of the world's primary production [3, 4]. The utility of microalgae for CCU is being demonstrated at an industrial installation in (Arctic) northern Norway, that harnesses cold-adapted marine diatoms to capture CO_2 at the source from an operational ferrosilicon smelter plant [5]. This is a prominent example of how innovative algal biotechnology can be deployed for CCU by taking advantage of locally isolated diatom strains adapted to the seasonality and seawater resources available to coastal industry locations. A recent study reported a difference in the photophysiology of the Arctic diatom *Porosira glacialis* UiT201 compared to temperate microalgae that enables strong photosynthetic flexibility in the dynamic light and temperature environment, as is uniquely experienced in coastal, Arctic regions [6].

Scalingup mass cultivation of microalgae for CCU must, however, overcome several barriers that include managing the microbial populations within large industrial scale photobioreactors. As axenic mass cultivation of microalgae is technically challenging, large photobioreactors are often populated by multi-domain microbial communities and function as their own unique ecosystems [7, 8]. A better understanding of how these communities assemble and behave under stable and unstable – i.e., reactor crash – conditions is needed to scale bioreactors up to volumes capable of effective industrial CO_2 capture. Open airlift column bioreactors are

72 effective for scaled-up industrial operation, land space utilization, and improved CO_2 gas-to-
73 liquid transfer [5]. However, the challenge linked to open-systems is a constant exposure to the
74 environment and thereby capacity of recruiting new microbial taxa to the system [8]. The extent
75 to which a bioreactor recruits microbial taxa from the environment depends on how the inputs
76 (e.g., water, nutrients, innocula) and control volume are exposed to the environment. This
77 presents a unique intersection between bioprocess engineering for CCU and microbial ecology.
78
79 Coinhabiting microbial taxa within industrial photobioreactors are not necessarly harmful.
80 Infact, previous studies have demonstrated that the majority of the interspecific interactions
81 between microalgae and bacteria are mutualistic [9-11]. The beneficial associations are based
82 on exchange of synthesized compounds enhancing growth of both partners [12, 13]. The
83 invasion of other microbial eukaryotes i.e., phytoplankton and heterotrophic protist (hereafter
84 referred as "microeukaryotes") instead can cause a rapid culture collapse of the focal strain [8].
85 Some mass cultivation attempts used euryhaline and high pH tolerate focal strain to prevent the
86 contamination of competitive or predative species [14-16]. However, extreme conditions
87 reduce the number of potential strains and might complicate the operational processes. To
88 uncover the potential and risks associated with the coexisting microbiome, a shift in focus from
89 focal strain-centered investigations to ones that encompass the entire biological component of
90 the production system is necessary [17].
91
92 Major knowledge gaps exist in our understanding of how to control or even measure the
93 complex interactions between microalgae and bacteria. Only few previous studies have
94 measured the microbial diversity of photobioreactors and none (to our knolwege) have
95 investigated the effects on the stability of CCU-specific operations. In order to address some of
96 these knowledge gaps, mass cultivation of the marine centric diatom *Porosira glacialis* UiT201

97 in a preindustrial scale CCU operation was combined with an investigation of the community
98 composition and temporal dynamics of the bioreactor microbiome. This study systematically
99 explores three research questions: (i) Which prokaryotic and microeukaryotic taxa are
100 recurrently observed in mass cultivation alongside the focal diatom strain? (ii) What are the
101 discernible succession patterns within the prokaryotic and microeukaryotic consortia of the
102 bioreactor microbiome, and how are these patterns influenced by the bioreactor's fluctuating
103 environmental conditions? and (iii) In what ways do the outcomes of two sequential time-series
104 experiments, one concluding with a premature culture crash at 39 days and the other achieving
105 sustained culture stability for 60 days, converge or diverge in terms of microbiome structure
106 and bioreactor growth conditions?
107

108 **2 Materials and Methods**

109 *2.1 Algae inoculum*

110 The focal strain used to inoculate bioreactors was the centric diatom *Porosira glacialis* (UiT-
111 201). It was originally single-cell isolated from a sediment sample collected in the Barents Sea
112 (N 76° 27.54′, E 033° 03.54′) in 2014 [18]. The monoculture was maintained in a temperature
113 regulated incubator (Termaks KBP 6151) at 8 °C with a scalar irradiance of 65 µmol m^{-2} s^{-1}
114 and diluted weekly with filtered (0.45 µm) and autoclaved seawater containing inorganic
115 marco- and micronutrients from a plant nutrient mix YaraTera Kristalon Purple (Yara, Norway)
116 and sodium metasilicate pentahydrate ($Na_2SiO_3 \cdot 5H_2O$; Permakem, Norway).
117

118 *2.2 Photobioreactor operation*

119 Mass cultivation of *P. glacialis* took place in northern Norway (69.22°N; 18.08°E) in two
120 consecutive time series experiments: 19.08-27.09 and 19.10-17.12 in 2021. Hereafter the first
121 time series recording is referred to as "T1-crash" due to unplanned discontinuation of

122 cultivation caused by contamination and collapse of *P. glacialis* culture and the second as "T2-
123 healthy" in which the focal strain was maintained at targeted abundance and growth rates until
124 the end of time series. The cultivation platform is integrated into an industrial ferrosilicon
125 smelting plant, Finnfjord AS, in collaboration with UiT – The Arctic University of Norway.
126 Detailed description of the mass cultivation design and operation is given in Eilertsen et al. [5].
127 The platform contains four vertical column airlift photobioreactors with volumes of 6 m^3, 6 m^3,
128 14 m^3, and 300 m^3 [5]. This study deployed the largest 300 m^3 reactor operated with culture
129 volume of 150 m^3 under semicontinuous conditions. Seawater for cultivation medium was
130 obtained from an adjacent fjord at the depth of 25 m and filtered through a multichannel
131 filtration unit with 0.45 µm as a smallest cartridge pore size and through a UV-sterilization unit
132 (Model: MR3-350PP; Ultraaqua, Denmark). The sea water inflow to the reactor is however not
133 sterile, despite the robust filtration system. Compressed air and flue gas from the factory with
134 CO_2 content of 3-7% were added into the bottom of the reactor using a rotating gas dispersing
135 device [5]. Flue gas was sparged continuously during the cultivation periods except for three
136 dates: 28.10., 08.12. and 14.12.2021. Illumination was provided for the reactors by ambient
137 sunlight plus submerged LED lights. Additional artificial illumination was provided by a
138 collection of 200 x 50W LED light units, placed above the culture surface of the bioreactor.
139 The average light intensity in the 300 m^3 reactor, provided by the collection of sources, was 36
140 µmol m^{-2} s^{-1}. Inorganic nutrients (nitrite, nitrate and phosphate) a plant nutrient mix YaraTera
141 Kristalon Purple (Yara, Norway), and $Na_2SiO_3 \cdot 5H_2O$ (Permakem, Norway) were added
142 manually to the culture from the start of T1-crash and following 9-days (19.10.-29.10.2021).
143 Thereafter, the nutrients were automatically added using a dosing pump to ensure continuous
144 addition of nutrients. On average, 1231 g 24 h^{-1} of Kristalon Purple and 1321 g 24 h^{-1} of
145 $Na_2SiO_3 \cdot 5H_2O$ were dosed as liquid solutions into the bioreactor.
146

147 *2.3 Inoculation of the 300 m^3 photobioreactor*

148 Prior to each time series experiment, the inoculation volume of *P. glacialis* was increased in
149 two smaller (6 m^3 and 14 m^3) bioreactors. The 300 m^3 bioreactor was pre-filled with ~45 000 L
150 of filtered seawater before the inoculum was introduced. Thereafter, the volume was increased
151 for the next two days until a volume of 150 m^3 was reached as this was the maximum operation
152 capacity for the light delivery system. Semicontinuous operation was maintained by
153 continuously adding filtered seawater for five days a week (Monday – Friday morning) with a
154 mean inflow of 2100 L h^{-1}. The culture surplus was extracted from overflow (150 m^3) and algae
155 was harvested.

156

157 *2.4 Sample collection and growth condition monitoring*

158 Daily measurements were obtained during scheduled work shifts (Monday-Friday) from the
159 300 m^3 bioreactor for: temperature (C°; Endress-Hauser sensor iTHERMA ModuLine TM131
160 w) and dissolved oxygen (DO mg L^{-1}; WTWMulti 360 meter with CellOx 325 sensor, Xylem
161 Anlaytics, Welheim, Germany) from subsurface at ~1 m depth. Liquid samples for cell counts
162 were collected daily and for analysis of inorganic nutrients every second day. Samples were
163 collected twice weekly (Monday and Friday) for genomic DNA isolation and dissolved organic
164 carbon (DOC mg L^{-1}). These samples were also collected from the 300 m^3 bioreactor pre-filled
165 with filtered intake water (hereafter referred as "SW bioreactor"), from inoculum (sample from
166 6 m^3 and 14 m^3 were combined), and once or twice from intake water used to semicontinuously
167 dilute the reactor.

168

169 *2.5 Sample processing and analyses*

170 Concentration of inorganic nutrients of nitrite-nitrate (NO_2^-+NO_3^-), phosphate (PO_4^{3-}), and
171 silicate ($Si(OH)_4$) were analyzed immediately after filtering sample through a GF/C microfiber

172 filter (Whatman). Nutrient specific Spectroquant kit (Merck Millipore, US) was used and the
173 concentration was determined spectrophotometrically using a 96-well plate with a microplate
174 reader (FilterMax F5, Molecular Devices, CA) in which the absorbance was measured at 540
175 nm. DOC samples were filtered through 0.2 µm cellulose syringe filters (Corning®) into acid
176 washed and pre-combusted 60 mL glass vials and stored in dark at +4 °C for 12 months before
177 acidification with 250 µl of 2N HCl to determine the concentration with TOC-VPH analyzer
178 (Shimadzu, Japan). Modified Utermöhl method was used to obtain an estimation of *P. glacialis*
179 cell density. Cells from two separate transects were counted from a 4-well dish with inverted
180 light microscope (Zeiss PrimoVert) after 2 h of settling (Nunc, Thermo Fisher Scientific, US).
181 Depending on the culture density up to 1 L of sample was filtered through each (n = 5 per
182 sampling day) 0.22 µm Sterivex polyethersulfone membrane filter unit (Millipore) to obtain
183 biomass for DNA extraction and following amplicon sequencing of 16S and 18S small subunit
184 ribosomal RNA (SSU rRNA) gene. The filters were immediately stored at -80 °C for 3 months
185 before DNA extraction took place at the UiT – The Arctic University of Norway.
186
187 *2.6 Microbiome – DNA extraction and Illumina amplicon sequencing*
188 DNeasy Power Water Kit (Qiagen, Carlsbad, CA) was used for the purification of genomic
189 DNA. DNA was extracted from the collected biomass following the manufacturer's protocol
190 and stored at -80 °C until PCR. DNA quality and quantity were assessed by the NanoDrop™
191 Spectrophotometer and Qubit fluorometer (Invitrogen TM). Bacterial communities were
192 identified following the Earth Microbiome Project (EMP) [19] 16S Standard Illumina library
193 preparation protocol, using the forward-barcoded 515f (5′-3′ GTGYCAGCMGCCGCGGTAA)
194 [20] and 806r (5′-3′ GGACTACNVGGGTWTCTAAT) [21] primers targeting the V4 region
195 of 16S SSU rRNA. Targeted gene sequencing of microeukaryotes was determined following
196 the EMP 18S Illumina Amplicon Protocol using the universal Euk_1391f (5′-3′

197 AATGATACGGCGACCACCGAGATCTACAC) and the barcoded EukBr (5′-3′
198 TGATCCTTCTGCAGGTTCACCTAC) primers targeting the V9 region of the 18S SSU rRNA
199 [22, 23]. The PCR products were confirmed by gel electrophoresis in 1% agarose gels and
200 visualized with GelRed® staining. Sequencing was performed on an Illumina MiSeq
201 instrument (Illumina, San Diego, CA) at the Environmental Sample Preparation and
202 Sequencing Facility at Argonne National Laboratory (Lemont, IL, USA).
203
204 *2.7 Microbiome – Bioinformatic pipeline*
205 Forward and reverse Illumina reads were imported into QIIME2 and thereafter demultiplexed
206 using the complementary barcode file, for each sequencing run, respectively. These steps were
207 executed using the EMP paired-end flag. DADA2 V2021.0 algorithm within QIIME2 was used
208 for denoising and merging the reads [24]. During the denoising steps of the DADA2 package,
209 the reads were pooled, filtered and chimera-checked. Additionally, the replicated reads were
210 discarded. After removing the chimeric sequences and singletons, the reads of the 16S runs and
211 18S runs were merged, respectively. The DADA2 statistic on sequence reads is provided in
212 Supplementary Data S1. Subsequently, the DADA2 package was used for inferring and filtering
213 amplicon sequence variants (ASVs). Taxonomy was assigned by aligning these ASV sequences
214 against the SILVA V138.1 database, using a 16S and 18S rRNA self-trained classifier [25, 26].
215 Classifiers were trained using RESCRIPt [27]. Although the 16S and 18S based taxonomic
216 classification within SILVA V138.1 database is a standardized approach, it has limitations
217 regarding classification and taxonomic resolution. It is noted that manual correction of
218 taxonomic annotation was beyond the scope of this study. Nevertheless, order *Enterobacterales*
219 was manually changed to order *Alteromonadales* as the main classified families obtained within
220 *Enterobacterales* in our 16S dataset were *Colwelliaceae*, *Alteromonadaceae* and
221 *Pseudoalteromonadaea* that are known to be members of *Alteromonadales* [28]. The

222 phylogenetic rooted tree, generated with QIIME2 using MAFFT algorithm, was used for
223 downstream diversity analysis.
224
225 *2.8 Microbiome – Diversity analysis*
226 All downstream analyses were completed using R [29]. The microbiome time-series data was
227 analyzed using the "microeco" package (v16.0) [30] unless another package is specifically
228 mentioned. Prior to community analysis, all reads that belonged to mitochondria, chloroplast
229 and kingdom *Eukaryota* or were unassigned (no assigned kingdom) were removed from the
230 16S dataset. Filtering from the 18S dataset removed ASVs assigned to the kingdoms of *Bacteria*
231 and *Archae*a and phyla of *Vertebrata*, *Arthropoda*, *Nematozoa*, *Annelida*, *Cnidaria*, *Mollusca*,
232 *Tunicate*, *Phragmoplastophyt*a and *Basidiomycota* as well as unassigned reads. Both datasets,
233 16S and 18S, contained a high number of ASVs with low number of reads. For example, 2074
234 ASVs found just in 300 m^3 bioreactor comprised only 0.6% of the total reads within 16S dataset.
235 As we are only considering taxa with higher abundance, an abundance-filter was used to remove
236 rare and possibly ambiguous ASVs from the datasets. The filter-threshold included only ASVs
237 which mean relative abundance within the complete 16S and 18S dataset was > 0.01%,
238 respectively. As a result of this the number of ASVs decreased from 7736 to 430 and from 1577
239 to 58 within the 16S and 18S datasets, respectively. After these filtration steps the final library
240 size comprised 5908-18328 and 2924-417990 reads in 16S and 18S datasets, respectively.
241 Counts of unique ASVs per sample were used to measure richness and Shannon´s diversity
242 index was used to combine both richness and evenness – i.e., alpha-diversity [31]. The
243 `multcomp´ R package was used to perform *post hoc* multiple comparison test with Tukey's
244 HSD [32] on alpha-diversity of different sample types between and within T1-crash and T2-
245 healthy using generalized linear models with Poisson distribution. Potential core prokaryotic
246 taxa associated with the focal strain from inoculum to the end of time series were identified

247 applying an occurrence filter with 75% presence threshold separately on T1-crash and T2-
248 healthy 16S dataset.
249
250 The compositional differences in prokaryotic and microeukaryotic communities – i.e., beta
251 diversity was measured with unweighted Jaccard distance metric to emphasize
252 presence/absence of community members. Three sample days, both from the beginning and the
253 end of 300 m^3 bioreactor time series, were included to infer how the bioreactor microbiome
254 composition in the start and end differs and how they relate to the communities in inoculum,
255 intake water and SW bioreactor (pre-filled 300 m^3 reactor with filtrated seawater). The
256 community differences were visualized in Principal Coordinate Analysis (PCoA) and the
257 pairwise permutational multivariate analysis of variance (PERMANOVA) was performed to
258 test the (dis)similarity associated to each sample type and chosen time series fraction within
259 and between T1-crash and T2-healthy. Beta diversity was also measured via Jaccard distances
260 between T1-crash and T2-healthy bioreactor time series samples, respectively. The Jaccard
261 distances were partitioned to account for the contributions of "nestedness" (species loss) and
262 "turnover" (species gain) towards the measure of total dissimilarity – i.e., $\beta_{Jac} = \beta_{Jne} + \beta_{Jtu}$ [33,
263 34]. This was obtained using the R package `betapart´ [35]. The associated changes within total
264 beta diversity, turnover, and nestedness, respectively, by sampling day were tested using
265 "adonis" function (PERMANOVA) with 999 permutations from the R package `vegan´ [36].
266 Taxonomic recruitment of prokaryotes across time series experiments was estimated by
267 calculating the fraction of ASVs that were unobserved at any previous cultivation days [34].
268 This metric was adapted from the data analysis of a previous study investigating community
269 succession and is available at https://github.com/pnnl/brislawn-2018-founders-species.
270

271 *2.9 Microbiome – Clustering around temporal patterns*

272 The K-medoids clustering method was depolyed following the previously described approach
273 applied to microbiome time series data [37, 38]. It was used to investigate shifts in microbiome
274 structure by examining community members that share similar succession patterns across the
275 crash and healthy time series. Prior to clustering the ASV counts were preprocessed through
276 normalization steps including variance stabilizing transformation (R package `DESeq2´) [39],
277 detrending (R package `pracma´) [40], and scaling via z-transformation to convert values on z-
278 score scale to ensure intercomparability between ASVs. As the K-medoids is a distance-based
279 clustering method pairwise Euclidean distances were calculated before each ASV was
280 partitioned into K clusters according to similarity criterion using the R package `cluster [41].
281 The clustering efficacy – a balance between number of clusters and data overfitting – was
282 assessed via the Calinski-Harabasz index (Supplementary Figure S1) [42, 43].

283

284 *2.10 Growth conditions of 300 m^3 bioreactor – Statistics*

285 The Mann-Whitney U test (also known as Wilcoxon rank sum test) was performed to reveal
286 differences in bioreactor physical and biogeochemical conditions between T1-crash and T2-
287 healthy. The potential association between the measured bioreactor conditions and their
288 relationship with identified temporal patterns (cluster medoids) was obtained using Spearman's
289 rank correlation analysis. Non-parametric analyses were conducted as half of the included
290 factors indicated non-normal distribution based on Shapiro-Wilk normality test (Supplementary
291 Table S1).

292

293 *2.11 Data repository and reproducible analyses*

294 The unprocessed genetic sequencing data for both SSU rRNA 16S and 18S gene amplicons is
295 available in ENA repository under project PRJEB66331. All data including ASV, taxonomy

296 and relative abundance tables, and bioreactor measurements used for analysis and graphing
297 along with R markdown script are available on the Open Science Framework (osf.io) as part of
298 this project: https://osf.io/k2egy/.
299

300 **3 Results**

301 *3.1 The bioreactor´s microbiome*

302 Open mass cultivation of marine diatoms within the industrial CCU setting resulted in a
303 complex microbiome that accompanied growth of the target strain *P. glacialis*. The prokaryotic
304 and microeukaryotic community members were diverse and relatively dynamic during both
305 crashed and healthy cultivation periods. This was observed across two time series experiments
306 with lengths of 39 (crashed) and 60 days (healthy), respectively. The first time series – T1-crash
307 – ended with the loss of dominating abundance of the focal strain, *P. glacialis*, as determined
308 by decrease in growth and microscopy observations revealed severe increases of dead and lysed
309 cells. During the second time series from a mass cultivation run – T2-healthy – *P. glacialis* kept
310 its growth and abundance until the operation was ended and biomass was harvested after 60
311 days. Despite the major differences in conditions and outcomes, the observed richness and
312 diversity, especially for the prokaryotic component of the microbiome, across both time series
313 experiments shared major similarities (Figure 1 and Supplementary Figure S2). The total
314 number of prokaryotic ASVs in T1-crash was 339 and in T2-healthy 368, and the corresponding
315 microeukaryotic richness was 47 and 56 ASVs, respectively. Prokaryotic richness and diversity
316 (Shannon) measured in the 300 m^3 bioreactor did not differ between the T1-crash and T2-
317 healthy (Tukey's HSD results given in Supplementary Table S2 and S3), as in total 291 and
318 256 unique ASVs were observed across time (Figure 1a,b), and the Shannon diversity index
319 ranged from 1.7 to 3.2 in T1-crash and 1.6 to 3.2 in T2-healthy (Supplementary Figure S2).
320 However, only 54-128 ASVs in T1-crash and 49-113 in T2-healthy were present in each sample
321 day, which corresponds to the richness measured within each inoculum, but not within the SW

322 bioreactor (pre-filled 300 m^3 reactor with filtrated seawater) or intake water samples, which
323 comprised a higher number of ASVs (Figure 1a, b). A clear decrease in Shannon diversity was
324 observed in the middle of T1-crash and T2-healthy. As a somewhat similar pattern was
325 observed in richness, especially in T2-healthy, the change was only partially driven by
326 evenness. The microeukaryotic portion of the bioreactor microbiome (measured via 18S rRNA
327 gene amplicons) showed mainly a contrasting result as compared to the prokaryotic community
328 (Figure 1c, d). The SW reactor and intake water samples comprised a smaller or corresponding
329 number of unique ASVs than the inoculum and per day 300 m^3 bioreactor samples. Also, the
330 mean number of observed ASVs but not Shannon diversity in the bioreactor across time was
331 significantly different between T1-crash and T2-healthy (Tukey's HSD, $p < 0.001$;
332 Supplementary Table S4 and S5). In both time series, the richness and changes that occurred
333 over time were somewhat similar until day 32. Thereafter, the number of observed ASVs stayed
334 constant in T2-healthy, whereas in T1-crash a strong decrease was observed towards the end
335 (Figure 1c, d). This pattern was not inferred from the Shannon diversity index, which instead
336 revealed similar diversity, especially in the end of both time series experiments (Supplementary
337 Figure S2).
338

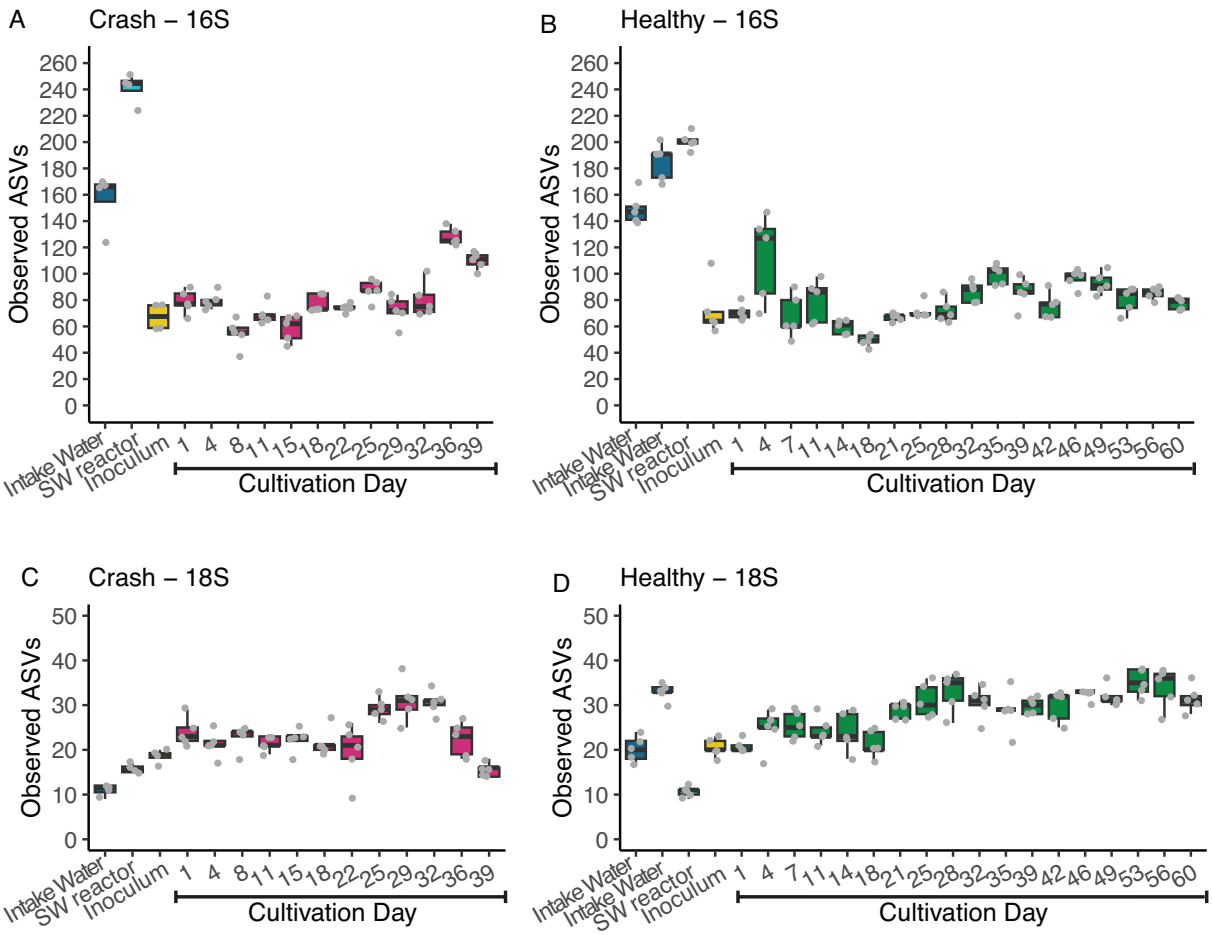


Figure 1. Alpha diversity measured as observed number of unique ASVs in each sample type and across bioreactor time series cultivations. Prokaryotic richness (16S) in time series experiment (A) crash and (B) healthy. Microeukaryotic richnes (18S) in time series experiment (C) crash and (D) healthy. The values of biological replicates (n = 4 or n = 5) are shown over boxplots. Note the different y-axis range between 16S and 18S datasets.

3.2 Mass cultivation has an abundant core microbiome

Nearly the same number of ASVs and taxonomic assignments within prokaryotes were found when ASVs were filtered on occurrence (i.e., presence) at a threshold of 75% among the inoculum and 300 m^3 bioreactor samples. These were 27 and 21 in T1-crash and 25 and 22 in T2-healthy, respectively, that were present throughout the mass cultivation runs. The core ASVs accounted for 20-50% of the unique ASVs observed each sample day in bioreactor of both time series runs. Fourteen of the core ASVs, categorized into nine different bacterial orders, were discovered in the bioreactor and inoculum of the two time series periods. Between 44% and 91% of the prokaryotic community's daily relative abundance in T1-crash and,

355 correspondingly, between 33% and 91% in T2-healthy, were the taxa that the core ASVs were
356 classified into (core and other ASV(s), henceforth referred to as core taxa) (Figure 2a, b).
357 Supplementary Table S6 contains a list of all the core ASVs together with the corresponding
358 taxonomic assignments.
359

360 *3.3 The reactor's prokaryotic community structure was distinct between crash versus healthy*
361 *production*

362 When the 16S ASVs were allocated to the taxonomic level order, the prokaryotic community
363 in both time series studies exhibited a dominance of three groups (Figure 2a, b). Specifically,
364 in the T1-crash, the community was dominated by taxa classified within the order
365 *Rhodobacterales* (22%-75% of the relative abundance) until sample day-11, comprising mainly
366 ASVs assigned to genus *Planktotalea* (Figure 2a). In contrast, T2-healthy, *Rhodobacterales*
367 became the most abundant class in the end of the time series, after sample day 46 (28%-54%
368 of relative abundance) and was dominated by ASVs not assigned to a lower taxonomic level
369 (Figure 2b). Taxa indicative to the order *Flavobacteriales* were predominant, notably ASVs
370 assigned to genera NS9 marine group and *Winogradskyella*, for the first half of the T2-healthy
371 time series with exception of sample day-14, while in T1-crash the predominance of the same
372 genera was shorter and limited in the middle of the time series (Figure 2a, b). The third
373 predominant prokaryotic order was different between the time series experiments as taxa
374 classified within order *Chitinophagales*, mainly ASVs assigned to the family *Saprospiraceae*,
375 became the most abundant in the end of T1-crash (25%-64% of relative abundance) (Figure
376 2a). In contrast to that in T2-healthy taxa classified as *Opitutales*, comprised almost solely by
377 genus indicative to *Lentimonas*, had a high relative abundance in most of the sample days and
378 was the most abundant on sample day-14 (49% of the relative abundance) and in the middle of
379 the second half of the time series (Figure 2b).

380

381 *3.4 The bioreactor's microeukaryotic community structure was similar until the culture*
382 *collapse in T1-crash*

383 The focal strain, *P. glacialis*, is currently classified within class *Mediophyceae* [44], although
384 this classification was not given by the SILVA v138 database as 19 different ASVs indicative
385 to class *Mediophyceae* were assigned under the same annotation (*Mediophyceae*) down to
386 genus level. Among these ASVs, ASV *032943128192fa7f0a17b2d2991a72f4*, comprised 2-3
387 magnitude higher number of reads compared to the other ASVs. It is therefore considered that
388 most of the relative abundance within class *Mediophyceae* is composed by the focal strain, *P.*
389 *glacialis* in the 18S microeukaryotic communities.

390

391 The community structure of microeukaryotes at the taxonomic level class was similar between
392 T1-crash and T2-healthy samples until a rapid turn from *P. glacialis* dominance to
393 predominance of ASVs classified within the class *Bacillariophyceae* after sample day 29 (59%-
394 89% of relative abundance) in T1-crash (Figure 2c, d). It is notable that members of
395 *Bacillariophyceae* were present from the start in both time series with similar abundances until
396 strong occupation took place in T1-crash whereas in T2-healthy the relative abundance stayed
397 constant until the end of experiment. This contamination was caused by small pennate diatoms,
398 as confirmed by preliminary microscopic detection (Supplementary Figure S3). The most
399 abundant ASV within *Bacillariophyceae* was *f030aa085cd7c8b4bb6538425a95161c*. ASVs
400 assigned to the class *Bacillariophyceae* were also present in the community in the SW
401 bioreactor and intake water samples in the T1-crash, but to a lower degree in the T2-healthy
402 run, where taxa representative of the class *Chrysophyceae* were prevalent, particularly in the
403 intake water. However, *Chrysophyceae* did not show increasing occupation across time in the
404 T2-healthy microbiome (Figure 2c, d).

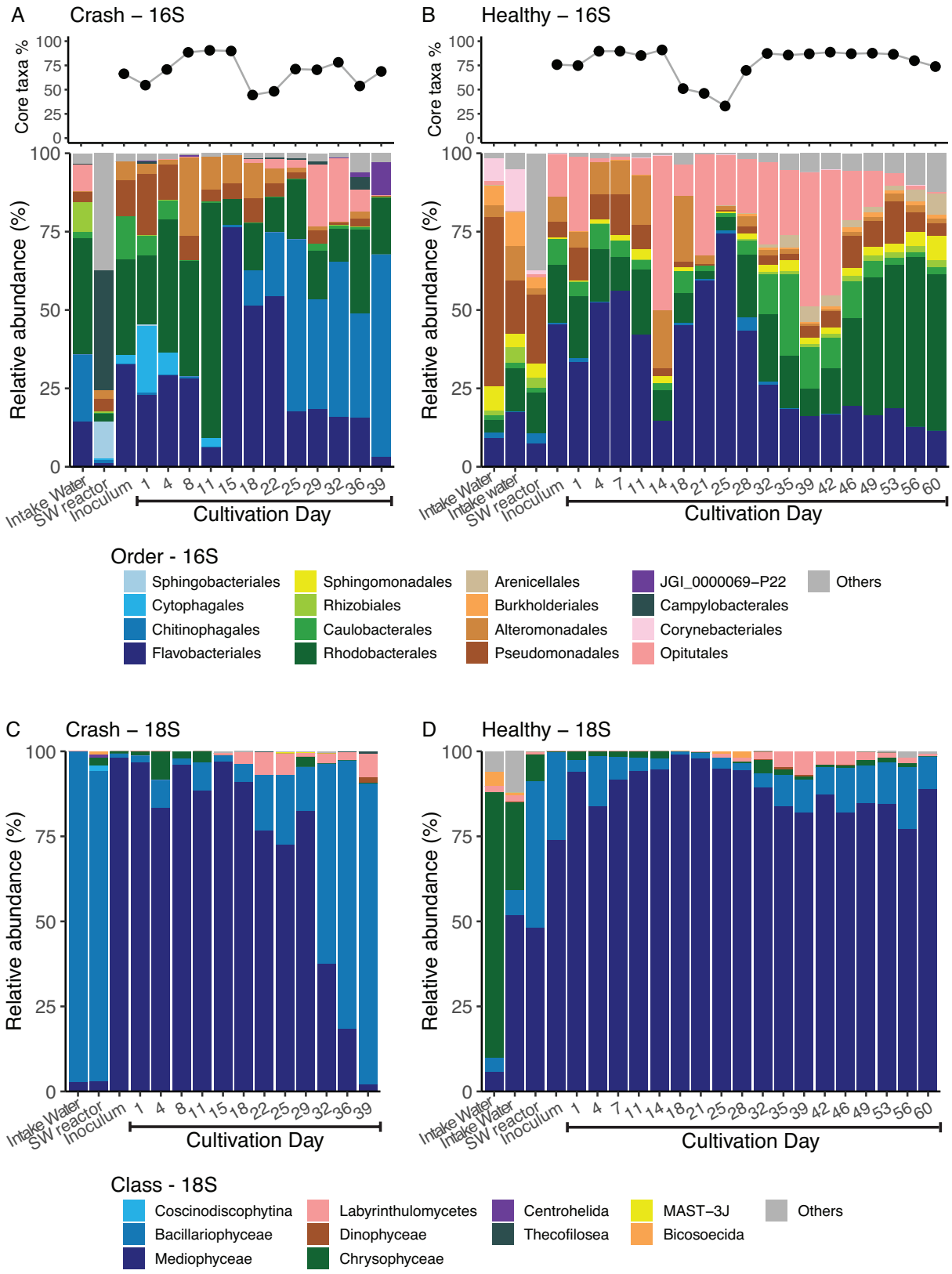


Figure 2. Comparison of community structure in each sample type and across bioreactor time series cultivations. The change in relative abundance of core taxa (core ASVs and other ASVs classified within the same taxonomic assignment) from inoculum to cultivation end and the taxonomic composition of the most abundant prokaryotes classified within taxonomic level order in (A) crash and (B) healthy time series. Taxonomic composition of the most abundant

411 microeukaryotes classified within taxonomic level class in (C) crash and (D) healthy time
412 series.
413

414 *3.5 Beta diversity – microbiome community composition between sample types and across*
415 *time*

416 The community composition of prokaryotes and microeukaryotes changed from inoculation to
417 the end-point in both time series experiments as inferred from pairwise PERMANOVA
418 comparison of the three first (start) and last (end) sample days (T1-crash: 16S, $R^2 = 0.50$ and p
419 < 0.01 & 18S, $R^2 = 0.61$ and $p < 0.01$; T2-healthy: 16S, $R^2 = 0.61$, $p < 0.01$ & 18S, $R^2 = 0.63$,
420 $p < 0.01$). The communities within start- and end-time fractions were significantly dissimilar
421 to those obtained in corresponding inoculum, intake water, and SW bioreactor samples and
422 these sample types were also dissimilar in relation to each other (Supplementary Table S7 and
423 S8). The Jaccard dissimilarity distances within 16S and 18S datasets, respectively, were
424 visualized by ordination with PCoA which showed that 39.7% and 44.0% of the total variation
425 among all included prokaryotic and microeukaryotic samples, respectively, were explained by
426 the two first principal coordinate axis (Figure 3). Closer inspection of the ordination results
427 indicated that in both time series experiments, the community composition in the early time
428 series were more similar with the parallel inoculum community, except for microeukaryotes in
429 T1-crash, than in the end as the communities exhibited strong differentiation across time
430 (Figure 3). The communities within SW bioreactor and intake water were clearly distinct from
431 those in inoculum and 300 m^3 bioreactor samples but also from each other, although the 300
432 m^3 bioreactor was pre-filled with intake water (Figure 3).

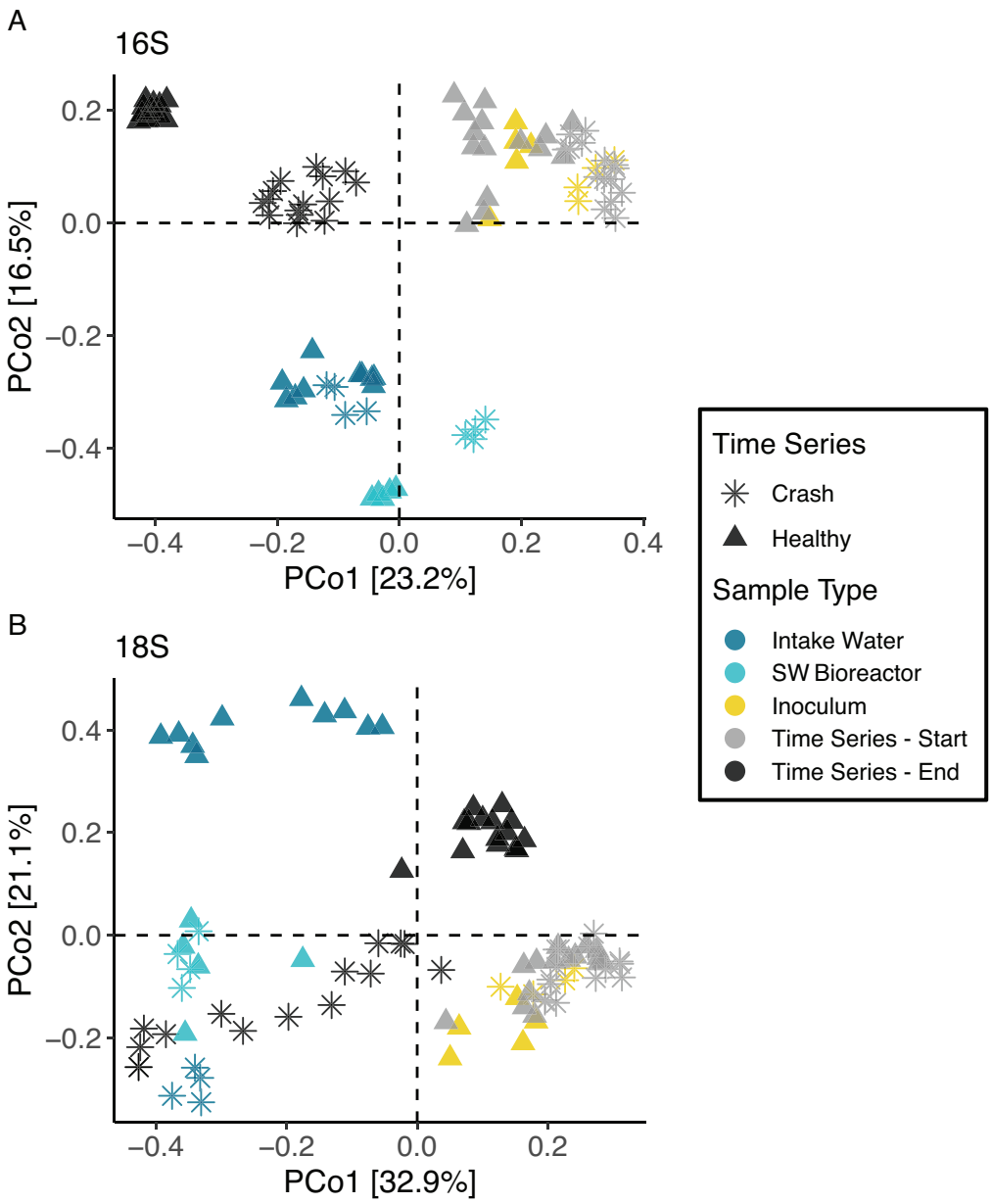
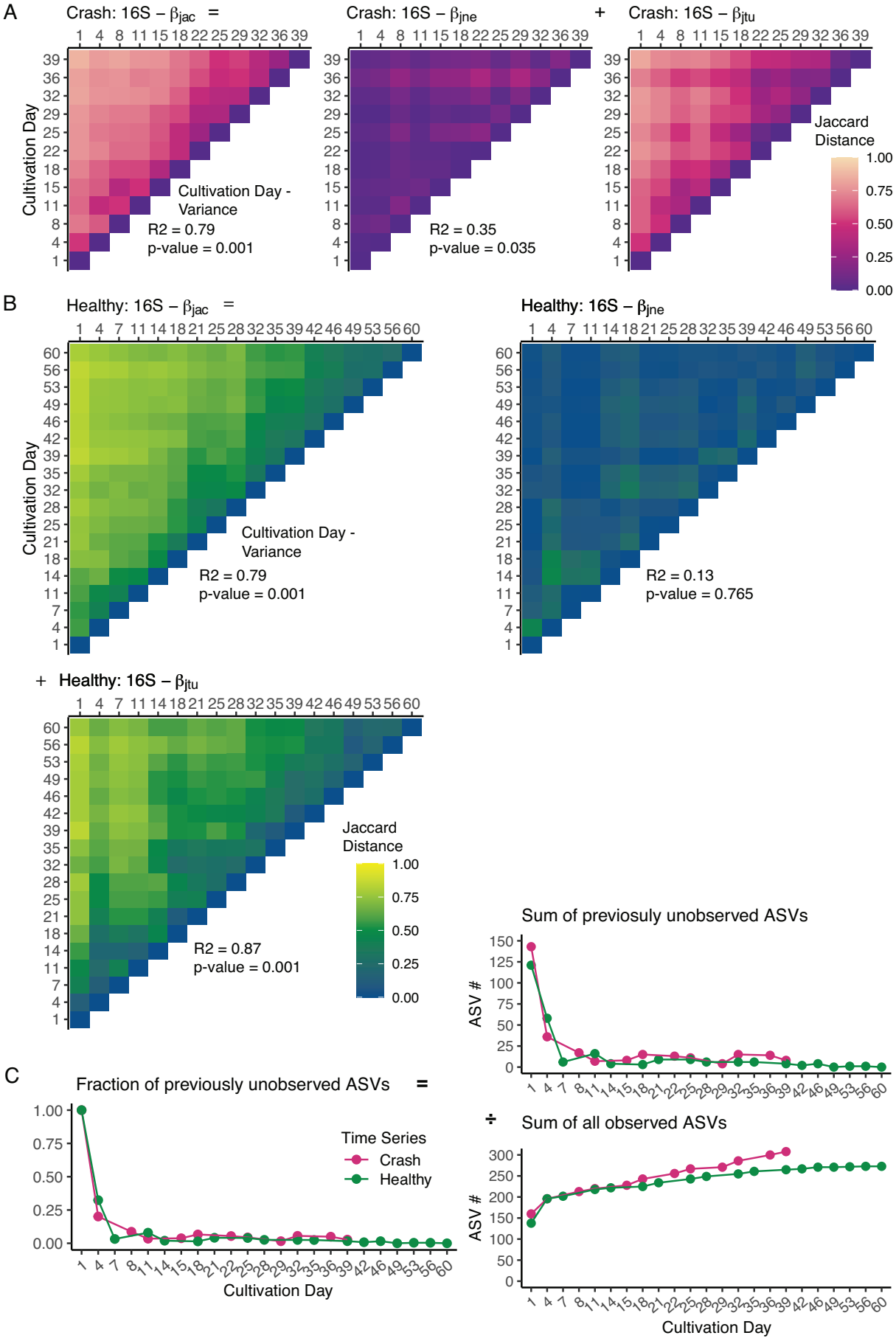


Figure 3. Beta diversity in prokaryotic and microeukaryotic commmunity composition between sample types and bioreactor time series start- and end-time fractions measured with unweighted Jaccard distances and visualized via Principal coordinate analysis. Dissimilarity within (A) prokaryotic and (B) microeukaryotic communities. Time series start-time fraction comprises cultivation days 1, 4 and 8 in crash and 1, 4 and 7 in healthy experiment. The corresponding end-time fractions are 32, 36 and 39 for crash and 53, 56 and 60 for healthy time series.

3.6 Beta diversity – contribution of species-loss and replacement to the bioreactor´s community succession patterns

Temporal variation in the community composition of the bacterial fraction of the microbiome was driven to a greater degree by turnover (species replacement) than nestedness (species loss)

445 in both time series (Figure 4a, b). This inference was obtained by analysis that disentangles the
446 additive contributions of these two antithetic processes to the total beta diversity (as measured
447 by the Jaccard distance) between all-time points across T1-crash and T2-healthy, respectively
448 [33]. The variance attributed to the sampling day was much lower from nestedness than from
449 turnover and total Jaccard in both time series as these explained over 84% and 78% of the
450 observed temporal variation associated with the time points, respectively, in both time series
451 runs (Figure 4 and Supplementary Table S9). Strong turnover is also indicated by the observed
452 prokaryotic richness which stayed rather constant across T1-crash and T2-healthy as under
453 nestedness some level of decrease would be expected through temproal species loss (Figure 1).
454 The bioreactor is exposed to several sources of new taxa that were not continuously monitored
455 in this study, including the input sea water and factory particulate matter from the air. Therefore,
456 the recruitment of prokaryotes was inferred from a metric that indicates the fraction of new taxa
457 introduced to the system at each time point in the time series experiments (Figure 4c). The
458 results revealed that the system is exposed to the detectable seed bank of available community
459 members very quickly. The introduction of previously unobserved ASVs clearly decreased after
460 the 11–14 days of cultivation, and less than 8% and 5% of the ASVs were new in the rest of the
461 cultivation days in T1-crash and T2-healthy, respectively (Figure 4c). Although both of the time
462 series experiments indicated similar pattern, the T1-crash recruited new taxa faster than T2-
463 healthy towards the end of time series.
464


A
Crash: 16S – βjac =
Crash: 16S – βjne
+
Crash: 16S – βjtu
Cultivation Day
Cultivation Day - Variance
R2 = 0.79
p-value = 0.001
R2 = 0.35
p-value = 0.035
Jaccard Distance
1.00
0.75
0.50
0.25
0.00
B
Healthy: 16S – βjac =
Healthy: 16S – βjne
Cultivation Day
Cultivation Day - Variance
R2 = 0.79
p-value = 0.001
R2 = 0.13
p-value = 0.765
+ Healthy: 16S – βjtu
Jaccard Distance
R2 = 0.87
p-value = 0.001
Sum of previosuly unobserved ASVs
ASV #
C
Fraction of previously unobserved ASVs =
Time Series
Crash
Healthy
Cultivation Day
÷ Sum of all observed ASVs
ASV #
Cultivation Day

466 Figure 4. Changes in prokaryotic communities across time. Unweighted Jaccard distance
467 measured between cultivation days in (A) crash and (B) healthy time series experiments. The
468 total Jaccard distance (β_{jac}) is partiotioned by contributions from nestedness (β_{jne}; species loss)
469 and turnover (β_{jtu}; species replacement) i.e., $\beta_{jac} = \beta_{jne} + \beta_{jtu}$. Lighter color represents cultivation
470 days with high dissimilarity. Variance and correspondig significance associated with
471 cultivation day are shown for each process. Introduction of new prokaryotic ASVs over time in
472 crash and healthy time series experiments (C). The fraction of previously unobserved ASVs
473 equals the number of ASVs that were not observed in any of the previous cultivation days and
474 the sum of the total number of ASVs observed at the corresponding time point.
475

476 *3.7 Temporal succession patterns and cooccurring taxa in distinct bioreactor growth*
477 *environments*

478 The physical and chemical environment in the large 300 m^3 photobioreactor, as well as cell
479 density of *P. glacialis*, were different between mass cultivation runs. This was established using
480 the Mann-Whitney U test that showed statistically significant differences between *P. glacialis*
481 cell number and raw fluorescence, concentration of inorganic nutrients, DOC, and bioreactor
482 temperature – i.e., most of the monitored factors except pH level (Supplementary Table S10).
483 The measured values of these factors were higher in T2-healthy compared to the T1-crash apart
484 from DOC concentration and bioreactor temperature which were lower (Figure 5 and
485 Supplementary Figure S4).

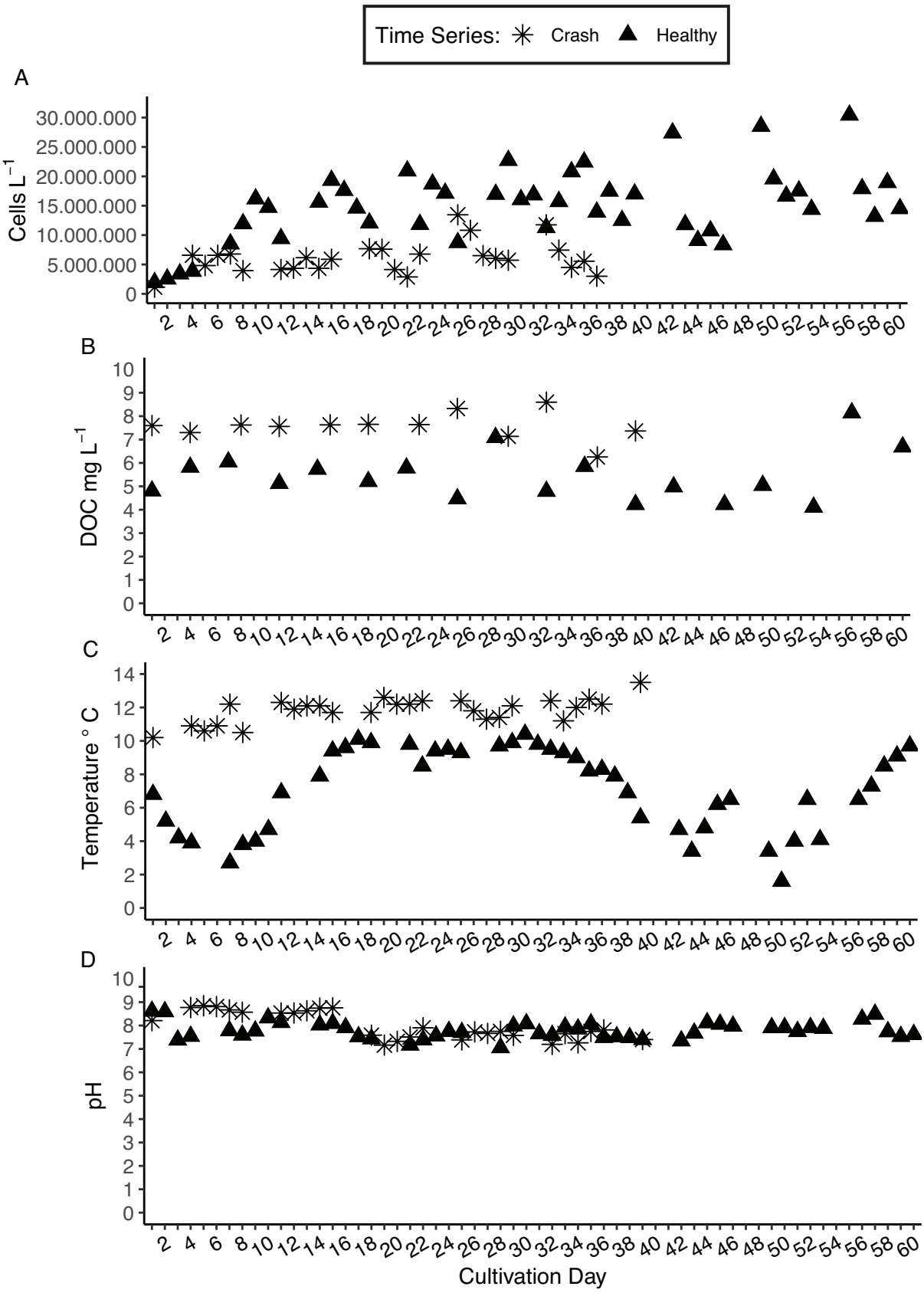


Figure 5. Variation in bioreactor environment across time in crash and healthy cultivation. Daily or twice a week measurements of (A) *P. glacialis* cell density, (B) DOC concentration, (C) bioreactor temperature and (D) pH level. Note that the time series crash was ended after cultivation day-39.

491

492 The potential association of community members to the variable bioreactor environment that
493 prevailed across the time series experiments was evidenced through correlation with temporal
494 patterns. A K-medoid clustering algorithm was used to reduce the temporal complexity within
495 the bioreactor microbiome by identifying the major temporal patterns and revealing the related
496 community members within each time series (T1-crash and T2-healthy) and dataset (16S and
497 18S). The ASVs within both time series and datasets were divided into three clusters based on
498 assessment of the Calinski-Harabasz index which was used as a quality metric (Supplementary
499 Figure S1) [43]. The succession patterns among prokaryotes, as represented by a cluster specific
500 medoid taxon, revealed that part of the community members from both time series datasets
501 (cluster 2 in T1-crash and T2-healthy) exhibited similar temporal dynamics with increase in
502 abundance in the middle of the time series (Figure 6a, b). The medoid of these clusters indicated
503 a significant positive relationship to one of the two factors measuring *P. glacialis* cell density.
504 Yet, the association to the other growth condition factors was not consistent (Figure 6c, d).
505 Also, most of the prokaryotic taxa within cluster 2 (T1-crash and T2-healthy) belonged to the
506 same highly abundant orders *Flavobacteriales*, *Rhodobacterales*, *Opitutales* and
507 *Alteromonadales* (Figure 6e, f). The K-medoid clustering also identified a high number of
508 prokaryotes in T1-crash (cluster 1 and 3) that first decreased and thereafter increased in their
509 abundance towards the culture crash of *P. glacialis* (Figure 6a). Especially in cluster 1 the
510 temporal pattern of medoid showed a strong inverse relationship with *P. glacialis* cell density
511 and DOC concentration and indicated an intensive proliferation of ASVs classified within
512 *Rhodobacterales* and *Chitinophagales* in the last two sample days as these orders comprised
513 the major taxonomic groups within this cluster (Figure 6c, d).

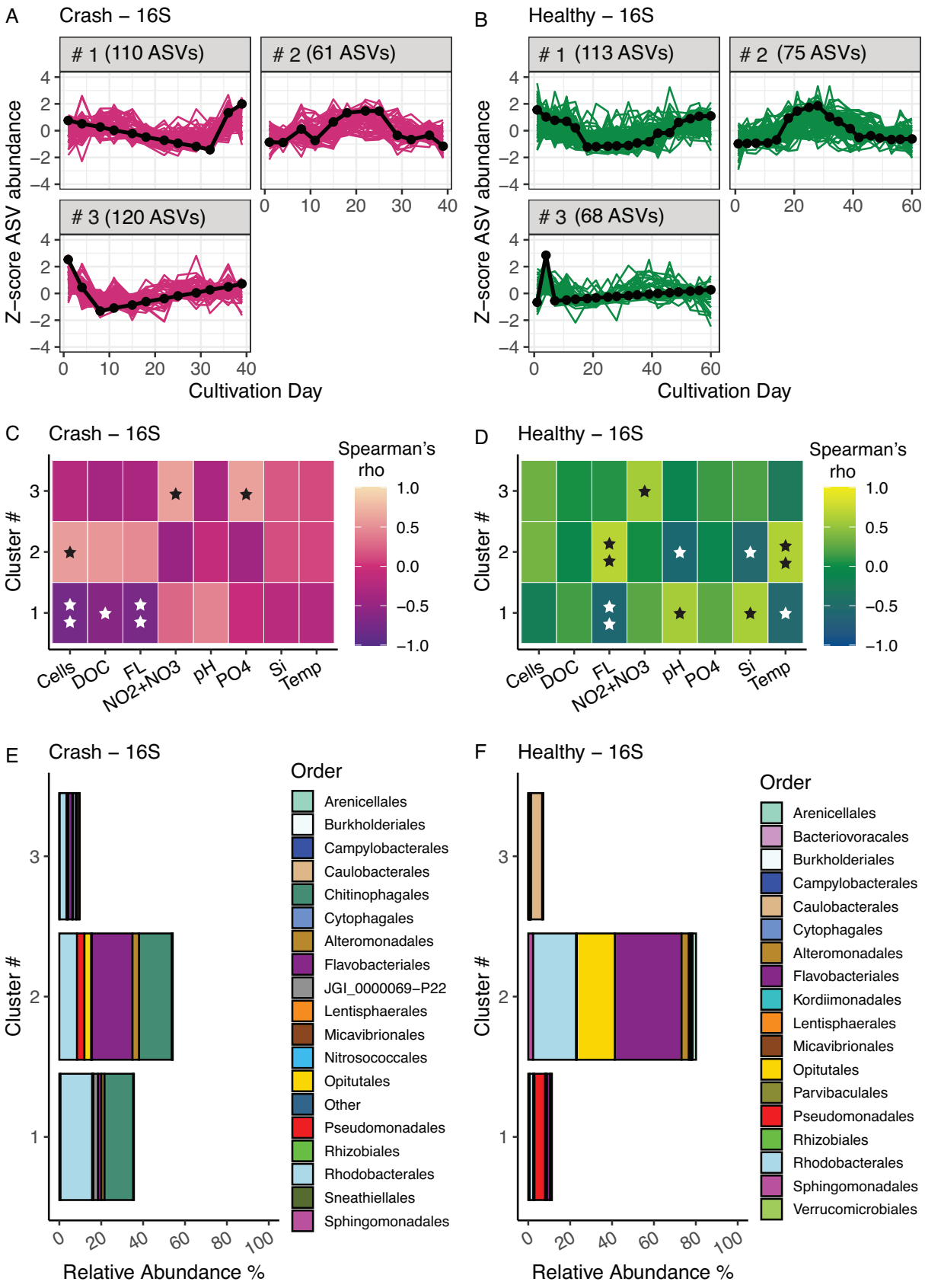


Figure 6. The major temporal succession patterns within prokaryotic communities as assessed via K-medoid clustering analysis with z-transformation and detrending of 16S ASVs. Temproal dynamics of prokaryotic taxa and identified number of ASVs in bioreactor (A) crash and (B)

healthy time series experiemnts. The black line indicates cluster specific medoid taxon, a representative for cluster's temporal pattern. ASV specific z-score is a y-axis and a value 0 denotes the mean abundance. Heatmap of Spearman's correaltion between temporal dynamics of cluster specific medoid taxon (z-scores) and measured bioreactor factors in time series (C) crash and (D) healthy. The color indicates Spearmna's rank correlation coefficient (rho) and the significance is denoted with ★ (★, $p \leq 0.05$; ★ ★, $p \leq 0.01$; ★ ★ ★, $p \leq 0.001$). Cells, *P. glacialis* cell density L^{-1}; DOC, DOC mg L^{-1}; FL, raw fluorescence as a proxy for increase/decrease of *P. glacialis* cell density; NO2-NO3, NO_2^- + NO_3^- µM; pH, pH level; PO4, PO_4^{3-} µM; Si, $Si(OH)_4$ µM, Temp, bioreactor temperature ºC. The most common prokaryotic taxa at the taxonomic level order within each cluster and their mean relative abundance in a whole community in time series (E) crash and (F) healthy.

The cluster medoids identified within 18S microeukaryotic community revealed a strong fluctuation in relative abundances and weak correlation to bioreactor conditions across time in both time series (Figure 7a-d). ASVs classified within class *Mediophyceae* and *Bacillariophyceae* were the main groups driving distinct temporal patterns in T1-crash and T2-healthy. In both time series experiments the class *Bacillariophyceae* belonged to a cluster in which the medoid exhibited strong decrease in relative abundance in the middle of time series (Figure 7e, f).

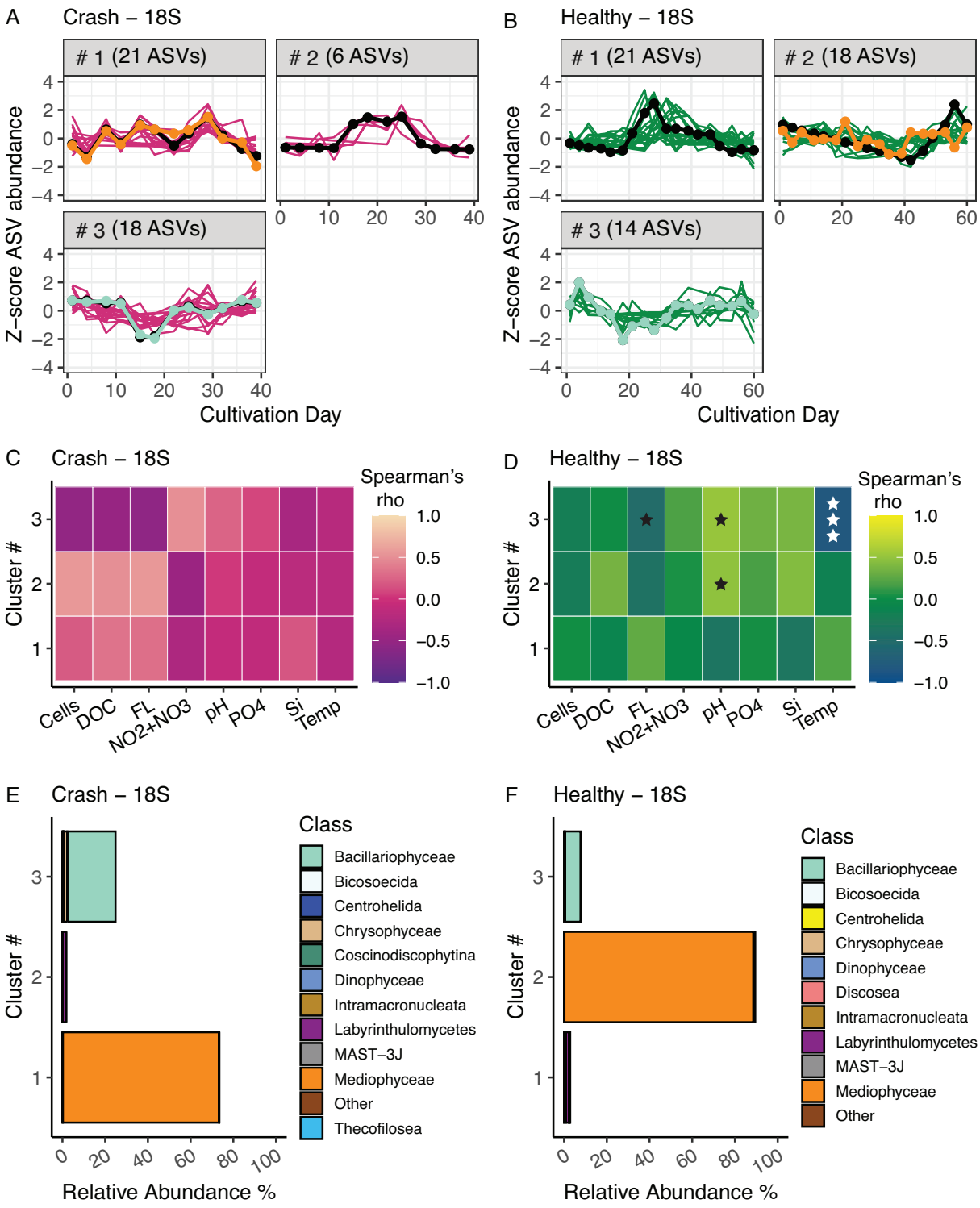


Figure 7. The major temporal succession patterns within microeukaryotic communities as assessed via K-medoid clustering analysis with z-transformation and detrending of 16S ASVs. Temproal dynamics of prokaryotic taxa and identified number of ASVs in bioreactor (A) crash and (B) healthy cultivation experiemnts. The black line indicates cluster specific medoid taxon, a representative for cluster's temporal pattern. ASV specific z-score is a y-axis and a value 0 denotes the mean abundance. The orange line specifies temporal dynamics of the most abundant ASV classified as *Mediophyceae* representing an expected focal strain, and the light green line specifies the most abundant ASVs classified as *Bacillariophyceae*, representing a potential pennate contaminant. Heatmap of Spearman's correaltion between temporal dynamics of

547 cluster specific medoid taxon (z-scores) and measured bioreactor factors in time series (C) crash
548 and (D) healthy. The color indicates Spearmna's rank correlation coefficient (rho) and the
549 significance is denoted with ★ (★, $p \leq 0.05$; ★★, $p \leq 0.01$; ★★★, $p \leq 0.001$). Cells, *P. glacialis*
550 cell density L^{-1}; DOC, DOC mg L^{-1}; FL, raw fluorescence as a proxy for increase/decrease of
551 *P. glacialis* cell density; NO2-NO3, NO_2^- + NO_3^- µM; pH, pH level; PO4, PO_4^{3-} µM; Si,
552 $Si(OH)_4$ µM, Temp, bioreactor temperature ºC. The most common microeukaryotic taxa at the
553 taxonomic level class within each cluster and their mean relative abundance in a whole
554 community in time series (E) crash and (F) healthy.
555

556 **4 Discussion**

557 This study was designed to measure and compare the microbial community succession patterns
558 of an industrial scale CCU photobioreactor during both successful and crashed operations.
559 While contamination control is important, it is not usually feasible to operate large scale, open
560 airlift photobioreactors as axenic cultures of microalgae. Hence, industrial scale production of
561 microalgal biomass in open systems is also mass cultivation of the associated microbial
562 community members. This is not necessarily a hindrance to the goals of algal CCU where pure
563 cultures are not required for success, if the bioreactor environment can effectively select for a
564 stable microbial community that promotes growth and abundant proliferation of a focal strain
565 capable of converting CO_2-enriched flue gas into biomass and bioproducts**.** This study
566 highlights an example of how a unique core microbiome, associated with cold-water marine
567 environments, stabilizes within a photobioreactor used for mass cultivation of diatoms and how
568 these types of microbiomes represent an important signature and possibly a controlling variable
569 associated with successful operation.
570

571 *4.1 CCU bioreactors are unique industrial marine microbial ecosystems*

572 This study characterized a diverse microbial ecosystem within an industrial scale
573 photobioreactor that was distinct from sea water inflow and inoculum (Figure 2 and 3). This is
574 despite the fact that the operation has been designed to promote growth of large marine diatoms
575 – i.e., *P. glacialis.* The microbial community structure is also sensitive to unique temporal
576 succession patterns that serve as signatures of success or crash. New taxa are recruited to the

577 system via inputs from the semicontinuous inflow of sea water from the adjacent costal bay
578 [45] and likely the air-water exchanges within the industrial environment. The majority of
579 taxonomic recruitment was observed within the prokaryotic portion of the microbiome but
580 observations were also made with respect to microeukaryotic taxa. However, the reactor
581 microbiome´s recruitment of new prokaryotic taxa stabilizes very quickly (Figure 4c). Although
582 the bioreactor might be exposed to a larger number of taxa e.g., ASVs removed by the 0.01%
583 abundance threshold or ASVs not detected, these taxa are not taking hold in the community.
584 Changes in community structure were observed on the time scale of days, in a similar fashion
585 to previous observations from temperate and coastal cold-water microbial ecosystems [38, 46].
586 It is expected that the diverse community comprises members with different ecological
587 properties, which potentially have a strong impact on the function of the production system in
588 relation to CCU [17]. In addition, most of the major bacteria groups *Flavobacteriales*,
589 *Rhodobacterales* and orders within *Gammaproteobacteria* found in the 300 m^3 bioreactor have
590 frequently been reported to associate with diatom driven phytoplankton blooms in natural
591 ecosystems or in diatom monoculture experiments (Figure 2a, b) [11, 31, 47, 48].
592
593 Bacteria that are commonly associated with marine diatoms are known to utilize different alga-
594 derived organic substrates – *i.e.,* DOC that is released across different bloom or growth stages
595 or generated through temporal substrate degradation [46, 49]. These dynamic processes at least
596 partially underpin the successional blooming patterns of the prokaryotic groups that were
597 identified in this study and other investigations [38, 47]. The orders *Opitutales*
598 (*Verrumicrobiota*) and *Chitinophagales* which also belonged to the most abundant community
599 ASVs from the current study are not considered either uncommon or specific to our production
600 system as they have also been documented in environmental and experimental studies.
601 Members within *Opitutales* are likely part of free-living community thriving in eutrophic

602 environments such as wastewater related microalgae cultivations, with a capacity to degrade
603 complex polysaccharides [50]. Members within the *Saprospiraceae* family (*Chitinophagales*)
604 have been reported to interact with diatoms [51].
605
606 *4.2 Importance of a strong core microbiome in mass cultivation*
607 The relatively long and consecutively repeated temporal coverage on microbial community
608 composition provided us an excellent opportunity to infer reoccurring community succession
609 processes. The bioreactor harbors a prominent core microbiome although the taxonomic
610 composition underwent effective replacement of prokaryotes (strong turnover) between time
611 points (Figure 4). We found that the core ASVs comprised a high portion of the total number
612 of unique ASVs observed on each sample day in both time series. Many prokaryotic taxa that
613 are introduced by intake water and the pre-filled bioreactor at the beginning are replaced after
614 a while as they do not integrate into the core community, as evidenced by the strong
615 contribution of turnover towards total community dissimilarity. Consequently, the bioreactor
616 functions as a selective environment because of the very different physical and biogeochemical
617 conditions that these taxa suddenly encounter in comparison to the coastal environment, such
618 as low pH and elevated levels of light, nutrients, and DOC, or because of the core taxa's
619 competition and antagonistic relationships, which inhibit the growth of the other taxa. It is also
620 expected that *P. glacialis* imposes a selective force that shapes core microbiome as partially
621 evinced by the observed changes in the prokaryotic community when small pennate diatom
622 contaminates supplanted the *P. glacialis* focal strain. The relative strength of diatom-driven
623 selection compared to the physical bioreactor environment remains unclear, although it is
624 understood that the closely associated community is structured by microalgae-specific manner
625 and can persist through environmental variation [12, 52, 53]. The high similarity in the core
626 microbiome between the two cultivation experiments indicates for a certain level of specificity

627 with the production system as it is designed for mass cultivation of *P. glacialis*. It is noted that
628 at least part of the taxa introduced to the bioreactor by intake water influx may reside in the
629 system below detection or were excluded from the analysis along with the applied abundance
630 filter (0.01%). The core taxa (including core and other ASVs classified within the same
631 taxonomic assignment) accounted for an average of 69% and 77% of the relative abundance
632 within bioreactor samples taken from T1-crash and T2-healthy, respectively (Figure 2a, b).
633 Nevertheless, on the majority of sample days, the high abundance was accounted for by a small
634 number of core taxa, whose predominance fluctuated and was only permanent for a short while.
635 Consequently, the temporal change in the prokaryotic community structure was significantly
636 influenced by a small subset of core taxa. The effects of strong temporal variations in
637 community structure and the interactions of the (core) microbiome on the productivity, stability,
638 and/or controllability of the production system are not well understood. Increasing complexity,
639 or the number of interactions within cooccurring relationships, might positively affect stability
640 and, in turn, have an impact on the growth and biomass concentration of the focal strain. [54].
641 On the other hand, a previous study on the microbiome of marine sponges inferred that a dense
642 core microbiome with few abundant taxa had a stronger positive impact on community level
643 stability than more diverse assemblies of sparsely abundant taxa. [55].
644
645 *4.3 P. glacialis is a promising candidate for mass cultivation*
646 Despite the relatively diversified coinhabiting microbiome, *P. glacialis* grows robustly in the
647 mass cultivation system from inoculation through semicontinuous operation of the 300 m^3
648 photobioreactor. While this work is the first to look at the whole community of the bioreactor,
649 mass cultivation of *P. glacialis* for CCU has been successfully demonstrated in industrial scale
650 trials for several years before [5]. Reactor crash events are rare and conversion of industrial
651 CO_2 to *P. glacialis* biomass is typically successful at scale, regardless of cooccurring

652 microeukaryotes identified in this study. This relatively novel marine diatom is a viable agent
653 for large scale algal CCU and production of high-value biomass [18].
654
655 *4.4 Biological factors associated with reactor crash*
656 Even with a well controlled inoculation and a robust start during mass cultivation, the reactor
657 can occasionally crash. One of the main obstacles when expanding the size of a bioreactor or
658 switching from a closed to an open cultivation system is the high input water volume delivered
659 into the reactor in production systems with (semi)continuous operation, which makes them
660 susceptible to contamination. This is a well-recognized challenge for open cultivation systems
661 even though bioreactor crash events are rarely documented in peer reviewed scientific literature.
662 The research reports on biomass production collapse or culture crash often attribute it to
663 contaminants exhibiting predatory behavior or resource competition [8, 15, 56]. In the present
664 study, the culture collapse in T1-crash was presumably not caused by predation but rather
665 competition or antagonist interactions from another marine diatom or a member from the
666 prokaryotic community. Our study design does not allow for a conclusive mechanistic
667 understanding of the factors that lead to the culture crash; however, our findings suggest two
668 taxa that showed temporal blooming patterns in relation to the observed crash. The
669 *Saprospiraceae* family (order *Chitinophagales*) was the only highly prevalent member of the
670 core community in T1-crash but not in T2-healthy. As algicidal activity has been connected on
671 filamentous members of *Saprospiraceae* [51, 57], the presence of an antagonistic community
672 member against *P. glacalis* should be confirmed in further investigations. We also observed a
673 strongly increasing abundance of pennate diatoms within class Bacillariophyceae towards the
674 end of T1-crash, supported by intake water inflow (Figure 2c), that might have caused
675 competition over resources. Although the cause behind the culture crash remains highly

676 tentative, our findings provide a premise for targeted experiments to uncover taxa with
677 functions that inhibit algal production in this specific industrial system.
678
679 *4.5 Variable environments and blooming patterns within the bioreactor*
680 The large scale mass cultivation system is also unique in relation to the prevailing environment
681 created in the bioreactor and therefore difficult to replicate. This is primarily a result of seasonal
682 variations in the coastal environment and constant production optimization. Although the
683 measured growth conditions varied and were distinct between the time series experiments, the
684 K-medoids clustering did not reveal a clear correlation between the temporal abundance
685 patterns of microeukaryotes and bioreactor factors. Instead, the impact on the prokaryotic
686 community structure might have been stronger (Figure 6). For example, the direct impact of
687 lower and more variable temperature in T2-healthy than in T1-crash on the growth of focal
688 strain was probably minor as *P. glacialis* is shown to obtain a strong growth across the
689 temperature range of 2-12 ºC [18]. The growth rate of marine bacteria is known to have a
690 pronounced positive relationship within the similar temperature range [58]. Also, an elevated
691 temperature was reported to increase the abundance of microalgae-associated bacteria and
692 simultaneous carbon assimilation when natural seawater was studied in experimental conditions
693 [59]. Moreover, major differences were observed on the cell density of *P. glacialis* and
694 biogeochemical environment. While the cell density of *P. glacialis* was clearly lower across
695 T1-crash than T2-healthy, the measured DOC concentration revealed the opposite (Figure 5).
696 The dataset does not provide a clear insight on why the DOC level was almost constantly 2.1
697 mg L^{-1} higher in T1-crash than in T2-healthy or how it potentially governed the cooccurring
698 microbial community. DOC excretion can reflect physiological state of microalgae as a
699 response to stress or it can be triggered by biological interactions [60-62]. The lower cell density
700 in this case does not directly relate to decreased growth (before culture crash) which in turn

701 could indicate stress, because the dilution rate was higher in T1-crash than T2-healthy.
702 However, the concentration of inorganic nutrients, especially silicate in the first half of the time
703 series, was clearly lower in T1-crash than in T2-healthy and therefore might have impacted the
704 cellular response of *P. glacialis* (Supplementary Figure S4).
705
706 *4.6 Core microbiome – an important milestone for further research*
707 This study uncovered the core microbiome associated with mass cultivation of microalgae for
708 CCU and elucidates the role of recurring prokaryotic taxa associated with the open, large scale
709 industrial photobioreactor designed to biomass production of the cold-water diatom *P.*
710 *glacialis*. The findings of this study show that a stable and recurring core microbiome can
711 significantly influence the productivity and stability of industrial-scale open photobioreactors.
712 The core microbiome, comprising key prokaryotic taxa within orders assigned to
713 *Flavobacteriales*, *Rhodobacterales*, and *Opitutales*, shows resilience and adaptability within
714 the dynamic environment of the bioreactor, suggesting relationships that support the growth of
715 the focal diatom strain, *P. glacialis*. However, the specific functional roles of these core
716 microbial members remain unexplored. Understanding these roles will unlock insights into
717 optimizing bioreactor conditions and enhancing biomass yield and CO_2 removal from industrial
718 flue gas. Applying targeted experiments using isolation and laboratory coculturing of the
719 members from core community with *P. glacialis* alongside metagenomic and -transcriptomic
720 methods are essential [12, 63]. Also, continuing monitoring the microbiome succession and
721 community profiles over variable production environments and processes enables unraveling
722 the occasional detrimental contaminants in this unique production system. The knowledge will
723 support future research focusing on microbiome engineering to selectively promote beneficial
724 microbial interactions while suppressing antagonistic species that lead to culture crashes [64].
725 By manipulating the microbial community structure through targeted interventions, it may be

726 possible to achieve higher efficiency and stability in CCU processes, thereby advancing the
727 scalability and economic feasibility of these industrial scale marine diatom cultivation systems.
728

729 **5 Conclusion**

730 This work shows how basic concepts of marine microbial ecology can be used to gain a better
731 understanding of, and perhaps even control over, a promising industrial biotechnology for CCU.
732 Nevertheless, it is difficult to determine the mechanistic influence of the microbiome on large-
733 scale biomass production in large industrial-scale photobioreactors since they are difficult and
734 expensive to modify and operate with high replication for experimental purposes only. Each
735 cultivation run of the large vertical photobioreactor is distinct in terms of the cooccurring
736 microbiome community dynamics and maintained growth conditions throughout the *P.*
737 *glacialis* production process because of its location within a highly seasonal coastal factory
738 setting. However, the reactor is home to a relatively diverse microbial ecosystem that
739 transforms into a core microbiome via turnover of taxa introduced by the inoculum as well as
740 continuous seawater intake and air exposure to factory particles. Alterations in the core
741 microbiome may influence or result from declining and or stablized growth of *P. glacialis*.
742 Consequently, the unique signature of the reactor's microbiome is indicative of stable carbon
743 capture and utilization operations, playing a pivotal role in the efficient conversion of industrial
744 CO_2 into marine biomass.
745

746 **Author contributions**

747 HCB and NJA conceptualized the study. IHG, GKE and LI carried out the cultivation and
748 sampling. SK and SP processed molecular samples and amplicon sequence data. NJA and HCB
749 conducted data analysis and wrote and the manuscript draft. All authors contributed to the final
750 version.
751

Aknowledgements

This project was made possible through tight collaboration and technical support from Finnfjord AS, with specific acknolwdegements to Hans Christian Eilertsen, Jo Strømholt, John-Steinar Bergum and Geir-Henning Wintervoll.

Funding

This work was supported by the project AlgScaleUp, funded by the Research Councile of Norway as part of the Green Platform (NFR project number: 328654) and ABSORB – Arctic Carbon Storage from Biomes, which is a strategic funding inniative from UiT – The Arctic University of Norway (https://site.uit.no/absorb/). The bioinformatics computations were performed on resources provided bySigma2 - the National Infrastructure for High-Performance Computing and Data Storage in Norway.

Conflict of interest

The authors have no conflicts of interest to declare.

References

1. Masson-Delmotte, V., et al., *IPCC Special Report: Global Warming of 1.5° C.* Summary for Policymakers. https://www.ipcc.ch/sr15/chapter/spm/, 2018.
2. Schweitzer, H., et al., *Innovating carbon-capture biotechnologies through ecosystem-inspired solutions.* One Earth, 2021. **4**(1): p. 49-59.
3. Field, C.B., et al., *Primary production of the biosphere: integrating terrestrial and oceanic components.* science, 1998. **281**(5374): p. 237-240.
4. Behrenfeld, M.J., et al., *Biospheric primary production during an ENSO transition.* Science, 2001. **291**(5513): p. 2594-2597.
5. Eilertsen, H.C., et al., *Mass cultivation of microalgae: I. Experiences with vertical column airlift photobioreactors, diatoms and CO2 sequestration.* Applied Sciences, 2022. **12**(6): p. 3082.
6. Svenning, J.B., et al., *Lipidome Plasticity Enables Unusual Photosynthetic Flexibility in Arctic vs. Temperate Diatoms.* Marine Drugs, 2024. **22**(2): p. 67.
7. Fulbright, S.P., et al., *Bacterial community changes in an industrial algae production system.* Algal research, 2018. **31**: p. 147-156.

8. Davies, F.K., et al., *Microbiota associated with the large-scale outdoor cultivation of the cyanobacterium Synechococcus sp. PCC 7002.* Algal Research, 2021. **58**: p. 102382.
9. Amin, S.A., M.S. Parker, and E.V. Armbrust, *Interactions between Diatoms and Bacteria.* Microbiology and Molecular Biology Reviews, 2012. **76**(3): p. 667-684.
10. Durham, B.P., et al., *Cryptic carbon and sulfur cycling between surface ocean plankton.* Proceedings of the National Academy of Sciences, 2015. **112**(2): p. 453-457.
11. Buchan, A., et al., *Master recyclers: features and functions of bacteria associated with phytoplankton blooms.* Nature Reviews Microbiology, 2014. **12**(10): p. 686-698.
12. Raina, J.-B., et al., *Chemotaxis shapes the microscale organization of the ocean's microbiome.* Nature, 2022. **605**(7908): p. 132-138.
13. Amin, S.A., et al., *Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria.* Nature, 2015. **522**(7554): p. 98-101.
14. Meseck, S.L., *Controlling the growth of a cyanobacterial contaminant, Synechoccus sp., in a culture of Tetraselmis chui (PLY429) by varying pH: Implications for outdoor aquaculture production.* Aquaculture, 2007. **273**(4): p. 566-572.
15. von Alvensleben, N., et al., *Salinity tolerance of Picochlorum atomus and the use of salinity for contamination control by the freshwater cyanobacterium Pseudanabaena limnetica.* PloS one, 2013. **8**(5): p. e63569.
16. Haines, M., et al., *Pilot-scale outdoor trial of a cyanobacterial consortium at pH 11 in a photobioreactor at high latitude.* Bioresource technology, 2022. **354**: p. 127173.
17. Lian, J., et al., *The effect of the algal microbiome on industrial production of microalgae.* Microbial biotechnology, 2018. **11**(5): p. 806-818.
18. Svenning, J.B., et al., *Temperature dependent growth rate, lipid content and fatty acid composition of the marine cold-water diatom Porosira glacialis.* Algal Research, 2019. **37**: p. 11-16.
19. Thompson, L.R., et al., *A communal catalogue reveals Earth's multiscale microbial diversity.* Nature, 2017. **551**(7681): p. 457-463.
20. Parada, A.E., D.M. Needham, and J.A. Fuhrman, *Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples.* Environmental microbiology, 2016. **18**(5): p. 1403-1414.
21. Apprill, A., et al., *Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton.* Aquatic Microbial Ecology, 2015. **75**(2): p. 129-137.
22. Amaral-Zettler, L.A., et al., *A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes.* PloS one, 2009. **4**(7): p. e6372.
23. Stoeck, T., et al., *Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water.* Molecular ecology, 2010. **19**: p. 21-31.
24. Callahan, B.J., et al., *DADA2: high-resolution sample inference from Illumina amplicon data.* Nature methods, 2016. **13**(7): p. 581-583.
25. Quast, C., et al., *The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.* Nucleic acids research, 2012. **41**(D1): p. D590-D596.
26. Yilmaz, P., et al., *The SILVA and "all-species living tree project (LTP)" taxonomic frameworks.* Nucleic acids research, 2014. **42**(D1): p. D643-D648.
27. Robeson, M.S., et al., *RESCRIPt: Reproducible sequence taxonomy reference database management.* PLoS computational biology, 2021. **17**(11): p. e1009581.

28. Parte, A.C., et al., *List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ.* International Journal of Systematic and Evolutionary Microbiology, 2020. **70**(11): p. 5607.
29. R Core Team, *R: A language and environment for statistical computing, R Foundation for Statistical Computing.* 2021.
30. Liu, C., et al., *microeco: an R package for data mining in microbial community ecology.* FEMS Microbiology Ecology, 2021. **97**(2): p. fiaa255.
31. Aalto, N.J., et al., *Diversity and selection of surface marine microbiomes in the Atlantic-influenced Arctic.* Frontiers in Microbiology, 2022.
32. Hothorn, T., F. Bretz, and P. Westfall, *Simultaneous inference in general parametric models.* Biometrical Journal: Journal of Mathematical Methods in Biosciences, 2008. **50**(3): p. 346-363.
33. Baselga, A., *The relationship between species replacement, dissimilarity derived from nestedness, and nestedness.* Global Ecology and Biogeography, 2012. **21**(12): p. 1223-1232.
34. Brislawn, C.J., et al., *Forfeiting the priority effect: turnover defines biofilm community succession.* The ISME journal, 2019. **13**(7): p. 1865-1877.
35. Baselga, A. and C.D.L. Orme, *betapart: an R package for the study of beta diversity.* Methods in ecology and evolution, 2012. **3**(5): p. 808-812.
36. Oksanen, J., et al., *Community ecology package.* R package version, 2013. **2**(0).
37. Coenen, A.R., et al., *A primer for microbiome time-series analysis.* Frontiers in genetics, 2020. **11**: p. 310.
38. Aalto, N.J., et al., *Microbial Community Dynamics during a Harmful Chrysochromulina leadbeateri Bloom in Northern Norway.* Applied and Environmental Microbiology, 2023. **89**(1): p. e01895-22.
39. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome biology, 2014. **15**(12): p. 1-21.
40. Borchers, H.W. and M.H.W. Borchers, *Package 'pracma'.* accessed on, 2022. **4**.
41. Maechler, M., et al., *Package 'cluster'.* Dosegljivo na, 2013.
42. Lord, E., et al., *Using the stability of objects to determine the number of clusters in datasets.* Information Sciences, 2017. **393**: p. 29-46.
43. Caliński, T. and J. Harabasz, *A dendrite method for cluster analysis.* Communications in Statistics-theory and Methods, 1974. **3**(1): p. 1-27.
44. Guiry, M.D. and G.M. Guiry, *AlgaeBase. World-wide electronic publication, National University of Ireland, Galway.* http://www.algaebase.org, 2022.
45. Aalto, N.J., et al., *Drivers of atmosphere-ocean CO2 flux in northern Norwegian fjords.* Frontiers in Marine Science, 2021: p. 841.
46. Teeling, H., et al., *Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom.* Science, 2012. **336**(6081): p. 608-611.
47. Teeling, H., et al., *Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms.* elife, 2016. **5**: p. e11888.
48. Schäfer, H., et al., *Genetic diversity of 'satellite'bacteria present in cultures of marine diatoms.* FEMS microbiology ecology, 2002. **42**(1): p. 25-35.
49. Eigemann, F., et al., *Phytoplankton Producer Species and Transformation of Released Compounds over Time Define Bacterial Communities following Phytoplankton Dissolved Organic Matter Pulses.* Applied and Environmental Microbiology, 2023. **89**(7): p. e00539-23.
50. Orellana, L.H., et al., *Verrucomicrobiota are specialist consumers of sulfated methyl pentoses during diatom blooms.* The ISME Journal, 2022. **16**(3): p. 630-641.

51. Furusawa, G., et al., *Algicidal activity and gliding motility of Saprospira sp. SS98-5.* Canadian journal of microbiology, 2003. **49**(2): p. 92-100.
52. Mönnich, J., et al., *Niche-based assembly of bacterial consortia on the diatom Thalassiosira rotula is stable and reproducible.* The ISME Journal, 2020. **14**(6): p. 1614-1625.
53. Jackrel, S.L., et al., *Host specificity of microbiome assembly and its fitness effects in phytoplankton.* The ISME journal, 2021. **15**(3): p. 774-788.
54. Paddock, M.B., J.D. Fernández-Bayo, and J.S. VanderGheynst, *The effect of the microalgae-bacteria microbiome on wastewater treatment and biomass production.* Applied microbiology and biotechnology, 2020. **104**(2): p. 893-905.
55. Björk, J.R., et al., *The dynamic core microbiome: Structure, dynamics and stability.* BioRxiv, 2017: p. 137885.
56. Wang, Y., et al., *Identification of harmful protozoa in outdoor cultivation of Chlorella and the use of ultrasonication to control contamination.* Algal research, 2018. **31**: p. 298-310.
57. McIlroy, S.J. and P.H. Nielsen, *The family saprospiraceae.* The Prokaryotes: Other Major Lineages of Bacteria and The Archaea.[4th ed.], 2014: p. 863-889.
58. White, P.A., et al., *The effect of temperature and algal biomass on bacterial production and specific growth rate in freshwater and marine habitats.* Microbial ecology, 1991. **21**: p. 99-118.
59. Arandia-Gorostidi, N., et al., *Warming the phycosphere: Differential effect of temperature on the use of diatom-derived carbon by two copiotrophic bacterial taxa.* Environmental microbiology, 2020. **22**(4): p. 1381-1396.
60. Mühlenbruch, M., et al., *Mini-review: Phytoplankton-derived polysaccharides in the marine environment and their interactions with heterotrophic bacteria.* Environmental microbiology, 2018. **20**(8): p. 2671-2685.
61. Gärdes, A., et al., *Effects of Marinobacter adhaerens HP15 on polymer exudation by Thalassiosira weissflogii at different N: P ratios.* Marine Ecology Progress Series, 2012. **461**: p. 1-14.
62. Hulatt, C.J. and D.N. Thomas, *Dissolved organic matter (DOM) in microalgal photobioreactors: a potential loss in solar energy conversion?* Bioresource technology, 2010. **101**(22): p. 8690-8697.
63. Lian, J., et al., *Different co-occurring bacteria enhance or decrease the growth of the microalga Nannochloropsis sp. CCAP211/78.* Microbial Biotechnology, 2021. **14**(3): p. 1159-1170.
64. Hu, H., et al., *Guided by the principles of microbiome engineering: Accomplishments and perspectives for environmental use.* Mlife, 2022. **1**(4): p. 382-398.