

Genotype and culture condition effects on single-cell diatom microbiomes: enhanced detection of low-abundance taxa with CRISPR-Cas9

Ruben Schulte-Hillen, Jakob K Giesler, Thomas Mock, Nigel Belshaw, Uwe John, Tilmann Harder, Nancy Kühne, Stefan Neuhaus, Sylke Wohlrab

Genotype and culture condition effects on single-cell diatom microbiomes: enhanced detection of low-abundance taxa with CRISPR-Cas9

Ruben Schulte-Hillen ^{1,2,3,†,*}, Jakob K. Giesler ^{2,4,†}, Thomas Mock ⁵, Nigel Belshaw⁵, Uwe John ², Tilmann Harder ^{2,4}, Nancy Kühne ², Stefan Neuhaus ², Sylke Wohlrab ^{2,6}

¹Max Planck Institute for Marine Microbiology, HGF-MPG Joint Research Group for Deep-Sea Ecology and Technology, Celsiusstraße 1, 28359 Bremen, Germany

²Department of Ecological Chemistry, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany

³Albert-Ludwigs-Universität Freiburg, Fakultät für Biologie, Schänzlestraße 1, 79104 Freiburg im Breisgau, Germany

⁴University of Bremen, Faculty of Biology and Chemistry (FB 2), Leobener Straße 1, 28359 Bremen, Germany

⁵School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, United Kingdom

⁶Helmholtz Institute for Functional Marine Biodiversity at the University of Oldenburg (HIFMB), Ammerländer Heerstraße 231, 26129 Oldenburg, Germany

*Corresponding author. Department of Ecological Chemistry, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany. E-mail: ruben.lennart.schulte-hillen@gmail.com

†Ruben Schulte-Hillen and Jakob K. Giesler contributed equally to this work and share first authorship.

Abstract

Primary production in aquatic systems is governed by interactions between microalgae and their associated bacteria. Most of our knowledge about algal microbiomes stems from natural mixed communities or isolated algal monocultures, which therefore does neither address the role of genotypic diversity among the algal host cells nor do they reveal how this host diversity impacts the assembly process of associated bacteria. To overcome this knowledge gap, we developed a single-cell 16S sequencing approach in combination with CRISPR-Cas9 guided depletion of host 16S contaminations from the chloroplast. The validity of this novel method was tested by comparing bacterial communities of 144 single-cells across three genotypes of the Arctic marine diatom *Thalassiosira gravida* grown under different environmental conditions. From these, 62 single-cells were additionally sequenced after CRISPR-Cas9 treatment. Due to the improved sequencing depth, bacterial richness associated with individual diatom cells was increased by up to 56%. By applying this CRISPR-Cas9 treatment we not only revealed intraspecific host-genotype associations but also low-abundance bacterial taxa that were not detected by standard 16S rRNA gene metabarcoding. Thus, the CRISPR-Cas9 assisted single-cell approach developed in this study advances our understanding on how the intraspecific diversity among algal hosts impacts the assembly process of their associated bacteria. This knowledge is essential to understand the co-evolution and adaptation of species in algal microbiomes.

Keywords: holobiont; phycosphere; chloroplast; *Sulfitobacter*; *Colwellia*; 16S; interkingdom interactions; microbiome; vitamin

Introduction

The role and function of interactions between individuals in microbial communities is increasingly recognized in microbiome research [1, 2], which coined the term “holobiont” describing an assemblage of a host (e.g. microalga) and associated microbes (e.g. bacteria). As the main primary producers in aquatic ecosystems, both uni- and multicellular algae have been studied from a holobiont perspective because their associated microbiomes substantially affect algal physiology, growth, and resilience [3, 4]. For example, interactions between host algae and bacteria are underpinned by specific bacterial enzymes as well as reciprocal nutritional requirements for essential trace elements, micro-, and macronutrients [4, 5].

Given the recognized importance of microalgae for diverse ecosystem functions [6, 7], methods to accurately determine their associated microbial communities are of considerable value. Current strategies to assess the bacterial composition associated

with microalgae predominantly rely on laboratory monocultures or natural field communities, introducing uncertainties such as culture artifacts and biases due to single-cell isolation and filtration. Moreover, clonality in microalgal populations is usually short-lived as subpopulations arise due to genetic drift leading to clonal interference. How this intraspecific host diversity impacts the diversity of associated bacterial communities has rarely been studied and is therefore largely unknown. Natural algal microbiomes, on the other hand, are difficult to study in a natural community context due to challenges in assigning bacteria to specific hosts for revealing their interactions [8]. To address these challenges, single-cell-based metabarcoding has recently been developed to assesses the microbial diversity associated with single-cells of host protists such as choanoflagellates and ciliates [9–11].

However, this single-cell sequencing strategy comes with several challenges. For instance, sequencing small amounts of DNA isolated from a proximity of the host protist might introduce a bias

Received: 2 October 2024. Revised: 12 September 2025. Accepted: 24 October 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of the International Society for Microbial Ecology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

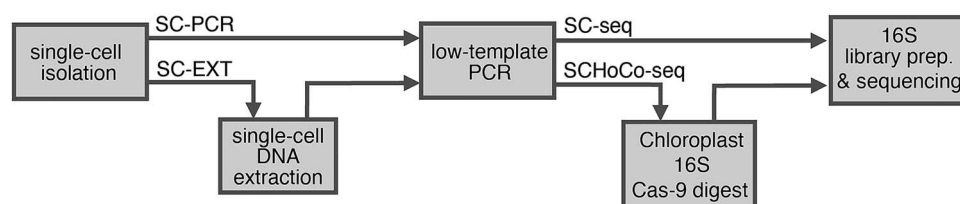


Figure 1. Schematic overview of the laboratory workflow. After single-cell isolation, samples were either subjected to DNA extraction (SC-EXT) or processed directly to low template PCR without prior DNA extraction (SC-PCR). A subset of low-template PCR products from all samples (62 out of 144) were treated with Cas9 to dissect the chloroplast 16S rRNA gene (SCHoCO-seq) prior to library preparation.

due to significant variability between single host cells in terms of the success of DNA extraction and amplification [12]. Although microbiomes on single sand grains have been successfully studied despite low DNA concentrations [13], similar microbiome analyses using single-celled algal microbiomes have been challenging due to the presence of 16S rRNA gene copies encoded in the algal chloroplast genomes, which can account for up to 99% of all sequence reads [14, 15].

This predominance of 16S rRNA gene copies encoded in chloroplast genomes reduces the sequencing depth of the 16S copies from the associated microbiomes. This caveat is adding to challenges imposed by overall low quantities of DNA, spatial heterogeneity and potential other methodological biases (see above). One strategy to reduce contamination by host 16S rRNA gene sequences from chloroplasts is PCR clamping, a method to suppress a particular allele during PCR [16]. Yet, this strategy has its own biases, i.e. species-specific suppression of PCR amplification [17, 18]. Here, we address these biases with a novel approach, Single-Cell Holobiont Cas9-Optimised-Sequencing (SCHoCO-Seq), which employs a Cas9 nuclease together with a target-specific guide RNA to dissect the 16S rRNA gene sequences of oceanic microalgae. Our approach is thus methodologically complementary to the approach developed by Song and Xie [19], who used Cas9 to dissect the 16S rRNA gene copies of rice chloroplasts from bulk leaf samples. However, we further extended this method to the single-cell level. In addition, as our approach is not restricted to a single species, it can be used to gain insight into a wide range of natural microbiomes associated with diverse single-cell host organisms.

We have compared this novel SCHoCO-seq approach with CRISPR/Cas9 untreated single-cell control microbiomes (further referred to as SC-Seq) of the diatom *Thalassiosira gravida* under different culture conditions. The SCHoCO-seq workflow is based on single-cell isolation, optional DNA extraction, subsequent nested low template PCR and Cas9-mediated, target-specific cleavage of the diatom chloroplast 16S rRNA gene amplicons, followed by amplicon metabarcoding of the associated prokaryotic community. Furthermore, controls were conducted to test for contaminations and off-target activity.

Material and Methods

Three strains of the Arctic diatom *T. gravida* were grown under different culture conditions. Single-cells were manually isolated and washed to remove free-living bacteria leaving mainly diatom-associated bacteria. The isolated single diatom cells were processed according to the workflow in Fig. 1 and described in detail below. Two different preprocessing protocols are presented: a direct PCR method, which is less labor intensive, and a method involving a single-cell DNA extraction step, which has the advantage of retaining the DNA for additional analysis.

Culture conditions

Diatom cultures of *T. gravida* strains (UIO 478; UIO 483; UIO 448, Table S1) were obtained from the Norwegian Culture Collection of Algae (NORCCA). Here, we refer to UIO 478 as A1, UIO 483 as A2, and UIO 448 as A5. Stock cultures were maintained in three culture media based on a modified and silica-enriched K-medium [20] (Table S2), a nitrogen-deplete medium (Table S3), and a vitamin-deplete medium (Table S4). Cultures were grown at 4°C and 50 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ under a light:dark period of 16:8 h.

Single-cell isolation

To study the diversity of diatom-associated microbiomes among conspecific host cells across genotypes, single-cells of *T. gravida* strains A1, A2, and A5 cultured under different conditions (see above) were isolated ($n = 8$). To remove the free-living bacteria, individual diatom cells were picked and sequentially washed five times by pipetting 3 μL into two separate 30 μL drops of PBS buffer (PBS CellPure, Roth, Karlsruhe, Germany), followed by three washings in 30 μL of PCR-grade water under an inverted light microscope. To validate the absence of free-living bacteria, 3 μL culture blanks (without diatom cells) were sampled from all strains and treated identical to the single-cell samples undergoing single-cell DNA extraction (see below). Cells were stored either in 30 μL 2xT&C lysis buffer of the MasterPure Complete DNA & RNA Purification Kit (Lucigen, Wisconsin, USA) or in 10 μL PCR-grade water. Cells in PCR-grade water were used directly as DNA template for a nested, low template PCR (see below). Cells stored in 2xT&C lysis buffer were used to generate a PCR product after DNA extraction (see below).

DNA extraction from isolated single-cells

To extract DNA from the single-cellular diatom microbiomes, a modified protocol of the DNA MasterPure Complete DNA & RNA Purification Kit was employed (SC-EXT). The changes to the manufacturer instructions were: (i) smaller buffer volumes, (ii) added pellet dye, and (iii) rinsing of pipette tips. (details in Table S5).

Nested low template PCR of single diatom cell Holobionts

During the low template PCR, the hypervariable V1-V9 region of the 16S rRNA gene was targeted by Primer 27F (5'-AGAGTTTGATY MTGGCTCAG-3') [21] and 1492R (5'-TACGGYTACCTTGTACGACT T-3') [22]. PCR reactions for single-cell samples were set up to 50 μL per reaction containing 5 μL 10 × Hotmaster Taq Buffer (Qiagen, Hilden, Germany), 0.5 μL Primer 27F (0.1 μM), 0.5 μL Primer 1492R (0.1 μM), 0.5 μL dNTPs (0.2 μM), 0.1 μL bovine serum albumin (BSA, 10 mg/ml), 0.5 Taq Polymerase (Qiagen), 32.9 μL H₂O and 10 μL Single-Cell Sample (SC PCR) or 2.5 μL extracted DNA (SC-EXT) plus 7.5 μL H₂O. PCR protocols for SC-PCR and SC-EXT differed only in the duration of the initial denaturation phase of 94°C for 1.5 min (SC-EXT) and 10 min (SC-PCR), respectively. Both protocols

consisted of 35 cycles; denaturation: 94°C for 1.5 min primer annealing: 55°C for 1.5 min, elongation: 68°C for 2 min, final elongation: 68°C for 10 min. PCR products were cleaned with the MinElute PCR Purification kit (Qiagen). All culture blanks (without diatom cell) showed no bands during the gel-electrophoresis (Fig. S1).

Cas9 digest of chloroplast-derived 16S-rRNA gene amplicons

The 16S rRNA gene amplicons generated by nested low template PCR were cloned into pCR 4-TOPO vectors and transferred into competent *Escherichia coli* cells of the TOPO TA cloning kit (Thermo Fisher, Massachusetts, USA). Plasmids of single colonies were purified with the Plasmid Plus Midi Sample Kit (Qiagen). PCR products were sequenced on an ABI 3130xl Genetic Analyzer; obtained sequences were searched for chloroplast 16S rRNA gene sequences by BLAST search in the PhytoRef database [23]. Plasmids containing the respective *Thalassiosira* chloroplast 16S rRNA gene and respective sequences were used for the design of the gRNA.

Design of gRNA

Guide RNAs (gRNAs) were designed for the *T. gradata* chloroplast 16S rRNA gene according to Hopes *et al.* [24]. In their study, Hopes *et al.* developed a modular Golden Gate cloning system for genome editing in diatoms, which streamlines the process of designing and validating gRNAs for targeted genome modifications. Their protocol covers both in silico gRNA selection, construct assembly, and experimental validation for high editing specificity in diatoms, which we adapted for efficient Cas9-mediated removal of chloroplast sequences in our single-cell workflow. Five gRNA sequences were selected for in vitro testing by comparison with endosymbiotic derived 16S rRNA gene amplicons and gRNAs lacking homology in the PAM sequences and therefore not cut by the gRNA mediated Cas9 [25]. Subsequently, the selected gRNAs were tested for their efficiency in directing Cas9 digestion of the *T. gradata* chloroplast 16S rRNA gene by OmicronCR, Norwich, UK.

Selection of the most suitable gRNA

The five gRNA sequences were aligned to the chloroplast sequence of a total of 15 diatom families retrieved from the PhytoREF plastid 16S rRNA database [23]. The gRNA which cut the central V4 region of chloroplast ASV-16S rRNA sequences in a highly conserved region was selected (Fig. S2). To determine the general utility of SCHO-co-seq for diverse diatom taxa, we searched for matching crRNA target sequences in the full PhytoREF database [23]. In addition, we searched the SILVA database V138 [26] for potential off-target sequences to further test whether a loss of microbial 16S rRNAs occurred due to unwanted Cas9 digestions. We allowed two mismatches to include potential off-targets with a 4% probability given the low mismatch tolerance of CRISPR-Cas9 [27]. With off-targets we refer to the potential unintended cleavage at nontarget DNA sites by CRISPR-Cas9. Furthermore, we evaluated in silico the phylogenetic range of the diatom chloroplast sequences not only regarding matching gRNA, but also regarding the outer and inner primers of the SCHO-co-seq approach. For this purpose, we also relied on the PhytoREF database. Chloroplast sequences with 100% matches to all primers and the gRNA were aligned with kalign [28] and an unrooted phylogenetic tree was built with FastTree [29]. Taxonomic annotations were transferred from the PhytoREF database [23].

Synthesis of sgRNA

The single guide RNA (sgRNA) comprised two specific sequences: the CRISPR RNA (crRNA) containing the complementary sequence to the target sequence, and the trans-activating-crRNA (tracrRNA) guiding and stabilizing the Cas9 nuclease [30]. Both parts were derived from respective oligonucleotides (universal oligo: 5'-AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGAC TAGCCTTATTTTAAGTTGCTATTCT-3', specific oligo: 5'-TAATACG ACTCACTATAGGAAGTCAACTGTAAATCTTGGTTTTAGAGCTAGA AATAG-3') (Eurofins, Hamburg, Germany) which were grafted together and amplified in a PCR, forming the double-stranded sgRNA template. For this reaction to succeed the two oligonucleotides must share a common overlapping sequence (linker) as well as binding sequences for a forward and reverse primer (Fig. S2).

A PCR was set up to 50 µL, containing 5 × Phusion Plus buffer (10 µL), a sgRNA specific oligo (1 µL, 0.1 µM), T7 sgRNA oligo 2 (1 µL, 0.1 µM), T7sgRNA Forward (3.75 µL, 10 µM), T7sgRNA Reverse (3.75 µL, 10 µM), dNTPs (1.5 µL, 10 µM), Phusion polymerase (0.5 µL) and PCR grade water (28.5 µL) (all Jena Bioscience, Germany). PCR conditions were: initial denaturing (98°C, 30 s), 30 cycles of denaturing (98°C, 10 s), annealing (51°C, 10 s), and elongation (72°C, 15 s); final elongation (72°C, 10 min.). The sgRNA was synthesized by in-vitro transcription as followed: 20 µL containing High Yield T7 reaction buffer (2 µL), DTT (2 µL, 100 mM), ATP, GTP, CTP, UTP (each 1.5 µL, 100 mM), High Yield T7 RNA Polymerase Mix (2 µL), template DNA (2 µL) and PCR grade water (4 µL) (Jena Bioscience, Germany). The reaction mix was incubated for 11 h at 37.5°C, followed by a DNase I digest (1 µL RNase free DNase I, Jena Bioscience) for 15 min at 37°C. The sgRNA was purified using the RNA clean & concentrator-5 kit (Zymo Research, USA) and the gRNA synthesis was checked for correct product size with an RNA Nano Chip Assay on a 2100 Bioanalyzer device (Agilent Technologies, Santa Clara, California, USA). The expected amplicon length of the sgRNA template PCR product is 127 bp, including both primer regions and subtracting the overlapping sequence.

In vitro digestion of 16S amplicons using Cas9 and gRNA

The Cas9 digest was performed after the amplification and purification of the V1-V9 region of the prokaryotic 16S rRNA and prior to the 16S rRNA V4 amplicon sequencing library preparation. The Cas9 Nuclease (New England Biolabs, USA) and the gRNA were assembled in a 27 µL reaction containing PCR grade Water (20 µL), Cas9 Buffer (3.5 µL), gRNA (150 nM, 3 µL) and Cas9 (1 µM, 0.5 µL) by incubating at 25°C for 10 min. The 16S rRNA gene was digested by adding 3 µL of 15 nM target DNA to the mixture and incubating it at 37°C overnight (~11 h). The sample was cleaned with the MinElute PCR Purification kit (Qiagen).

Illumina sequencing of partial 16S rRNA genes

The amplicon library was constructed according to the manufacturer's protocol for 16S metagenomic sequencing library preparation, described in Ahme *et al.* [31]. The library was sequenced on the MiSeq System (Illumina, San Diego, USA) according to the MiSeq Reagent Kit v3 (600 cycles) (Illumina). Forward Primer: MS_v4_515F_N:

5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG+GTGCCAGC MGCCGCGGTAA-3' [32], Reverse Primer: MS_v4_806R_1:

5'-GTCGTCGGCAGCGTCAGATGTGTATAAGAGACAG+GGACTA CHVGGGTWCTAAT-3' [33].

Processing of sequence data

FASTQ files were demultiplexed according to the “Generate FASTQ workflow” of the MiSeq sequencer software. Primers were removed with cutadapt v2.8 [34], and sequence data were processed with the DADA2 R package (quality trimming, denoising, merging, removal of chimeras, [35]). The resulting Amplicon Sequence Variants (ASVs) were taxonomically annotated with the SILVA v138 database [26]. Details about the FASTQ and ASV data processing pipeline are described in Ahme et al. [31]. Statistical analyses were performed in R version 4.1.2 (R Studio, Inc. 2021; www.r-project.org).

Analysis of single-cell diversity and bacterial community composition

ASVs were excluded from a sample if (i) their relative contribution to the respective sample was smaller than 0.25% as recommended by Reitmeier et al [36] (ii) ASVs represent common contaminants according to Sheik et al. [37] (iii) ASVs were known to be derived from nonmarine sources (see supplemental Table S6). Samples with sequencing depths in the lower 10% quantile of the whole dataset, or below 10 000 reads, were excluded from the analysis. ASV count tables were scaled by subsampling using the SRS package [38] and normalized by power transformation ($X^{0.25}$). Graphical representation was done with the Phyloseq package [39].

Statistical analysis of single-cell samples

Richness and Shannon diversity of the *T. gravis* single-cell samples were determined using the “microbiome” package, [40]. To test for significant effects on ASV richness and Shannon diversity as a response to strain identity (A1; A2; A5), culture condition (full medium; -N; -Vit) and DNA processing method (SC-EXT; SC-PCR), three-way ANOVAs were calculated with subsequent Tukey's post hoc tests. To evaluate differences in microbiome community composition across *T. gravis* strains and culture conditions, a Bray-Curtis distance matrix was calculated from the down-sampled and normalized ASV count data, serving as input for ordination of a principal component analysis (PCA). PERMANOVA was applied on the Bray-Curtis distance matrix as dependent variable, and strain identity and culture condition as independent variables using the “adonis2” function of the “vegan” package [41]. To obtain a more detailed analysis on microbiome composition and its drivers, bacterial ASVs were aggregated and subset to the genus level. Normalized count subsets for each genus were tested for differences across strains and culture conditions using two-way ANOVAs.

Analysis of SCHO-seq samples

Bacterial ASV data of single-cells of *T. gravis* strains A1, A2, and A5 were quality-filtered, down-sampled and normalized as described above. After richness calculation, an ANOVA was conducted to test for differences in bacterial ASV richness as a response to SCHO-seq treatment (compared to SC-seq control treatment). The respective diatom single-cell ID (of which the DNA product for both treatments originated) was treated as a random factor. To test for differences in relative proportions of bacterial ASVs, chloroplast ASVs and mitochondrial ASVs as a response to SCHO-seq, a t-test was conducted to compare the relative proportion of each ASV category compared to the control treatment. To evaluate the effect of sequencing depth on the relative proportion of bacterial, chloroplast and mitochondrial ASVs, 1000 replicates of random subsamples (each containing

500, 1000, 5000, and 10 000 ASVs) were generated from each sample.

The incidence rate ratio (IRR) represents the probability of detecting a bacterial family with SCHO-seq compared to SC-seq. The IRR was estimated by comparing the ASV counts per family using a generalized linear model with Poisson regression and log-link function in R at a significance level of $P < .05$. Positive values indicate a significant increase in the detection probability, while negative values indicate a significant decrease in the detection probability.

Results

Single-cell diversity and bacterial community composition analysis

The 16S rRNA ASV richness and Shannon diversity of the *T. gravis* single-cell microbiomes was strain specific and impacted by culture condition, but not affected by single-cell DNA processing methods (Table S7, Fig. S3). The differences were mainly caused by strain A1 which showed a significantly higher richness and Shannon diversity in the nitrogen- and vitamin-deplete media compared to the full medium (Fig. S4).

Moreover, the 16S rRNA gene ASV community composition of the single-cell *T. gravis*-associated microbiomes differed significantly in response to strain identity and culture conditions (Table S8). The genotypic specificity regarding microbiome community composition was reflected on the single-cell level, particularly for highly abundant bacterial groups. This is supported by distinct genotype-specific clustering in the PCA (Fig. 1). However, normalized read counts for multiple bacterial genera differed significantly between diatom strains and culture conditions (Table S9). Furthermore, no bacterial genus was consistently present across all analyzed single-cell replicates and strains (Fig. 2). Yet, eight of the 48 detected bacterial genera occurred across all tested strains, but with differences in abundance and frequency between single-cells. Within the fraction of shared bacterial genera, the *Sulfitobacter*, *Colwellia*, and *Roseobacter* clades showed the highest normalized reads and were particularly abundant in strain A1. Moreover, strain A1 shared three bacterial genera exclusively with strain A2 (most importantly *Glaciecola* and *Paraglaciicola*) and four bacterial genera exclusively with strain A5, of which *Polaribacter* was the most abundant genus. In strain A1, eight unique bacterial genera were found, with *Aurantivirga* as the most abundant genus. For strain A2 and A5, 10 unique bacterial genera were found each, of which *Lentilitoribacter* showed the highest abundance in strain A2 and *Octadecabacter* clade in strain A5.

Multiple bacterial groups showed significantly different abundance under specific culture conditions (Table S9, Fig. 2). This was evident for *Marinobacter*, which increased in abundance in vitamin-deplete treatments, as well as *Colwellia* showing an increase in abundance under nitrogen-deplete culture conditions.

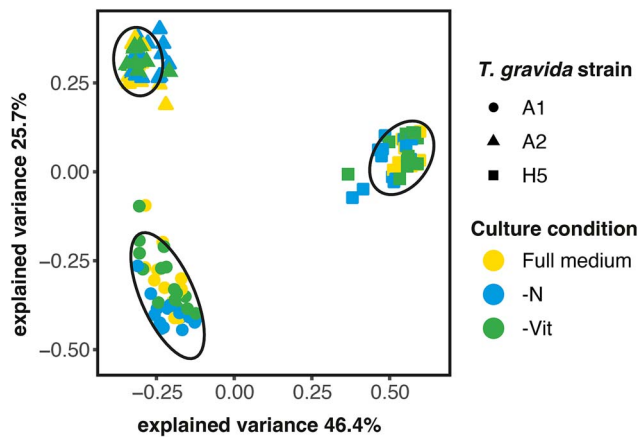
Single-cell SCHO-seq results

The SCHO-seq treatment significantly affected the relative proportions of all ASV categories (bacterial, chloroplast, and other ASVs) of the single-cell microbiomes (Table 1, Fig. 3). On average, chloroplast ASVs in a given single diatom-cell accounted for $71.4 \pm 14.3\%$ of total reads in SC-seq control treatments, but were reduced in the SCHO-seq treatment to $17.5 \pm 12.8\%$. Bacterial ASVs, which on average accounted for $25.9 \pm 14.7\%$ in the SC-seq control, increased to $71.0 \pm 16.7\%$ in the SCHO-seq treatment. Random subsampling across ranges of sequencing depth (Fig. 3)

Table 1. Results of Welch t-tests on relative proportions of the ASV categories bacteria, chloroplasts and other 16S rRNA gene ASVs as a response to SCHoCO-seq.

	% in SCHoCO-seq sample	% in SC-seq control	P-value	t-value
Bacteria	71.0 ± 16.7	25.9 ± 14.7	<.001	15.94
Chloroplast	17.5 ± 12.8	71.4 ± 14.3	<.001	-22.11
Other	11.5 ± 8.3	2.7 ± 2.1	<.001	8.17

Bacteria refer to all bacterial ASVs after removal of (A) potential contamination and (B) ASVs without annotations at least at the order level. Chloroplasts refer to all ASVs annotated as chloroplasts, and "Other" refer to mitochondria-derived ASVs and ASVs with no annotation at least at the order level. Mean proportions of ASV-categories are given in percent ± standard deviation. T- and P-values are reported for the treatment effect for each ASV category. Values marked with an asterisk (*) indicate significant effects ($P < .05$).

**Figure 2.** Principal component analysis of Bray-Curtis distances from *T. gravis* single-cell microbiomes in different culture media, including full K medium, nitrogen -deplete medium (-N), and vitamin-deplete medium (-Vit) based on normalized 16S ASV data. Culture conditions indicated by colour. Ellipses (confidence level = 0.95) show clusters of the respective *T. gravis* strain (indicated by shape).

showed that with increasing sequencing depth mostly bacterial 16S ASVs were detected in a theoretical ASV composition in the SCHoCO-seq treatment, whereas increasing sequencing depth in the SC-seq control treatment mainly resulted in more chloroplast ASV reads.

By increasing the overall proportion of bacterial ASV reads (Fig. 3), SCHoCO-seq significantly increased the probability of detecting several bacterial families compared to SC-seq (Fig. 4A). There was no significant decrease in the detection probability for any bacterial family. The increase in IRR values was particularly pronounced for *Sphingomonadaceae*, where the ASV counts were ~16 times higher compared to SC-seq. *Methylophilaceae*, *Rhizobiaceae*, *Parvibaculaceae*, and *Cryomorphaceae* also showed comparably high IRR values with IRRs >5 (Fig. 4A). The SCHoCO-seq approach additionally identified the bacterial phylum *Verrucomicrobiota* (Fig. 4A), which was not covered by the SC-seq control (Fig. 2). Consistently, there was a significant increase in the IRR value for a total of six families within these phyla.

In line, the ANOVA results indicated a significant effect of SCHoCO-seq treatment on bacterial ASV richness (Fig. 4B, Table S10). Single-cell samples without Cas9 digest (SC-seq) had a mean ASV richness of 10.4 ± 3.2 (strain A1), 6.5 ± 3.3 (strain A2), and 5.6 ± 1.8 (strain A5). In SCHoCO-seq treatments, richness increased to 15.0 ± 3.7 (strain A1), 8.4 ± 3.4 (strain A2), and 8.1 ± 1.5 (strain A5), corresponding to a richness increase of 51%, 53%, and 56%, respectively.

In silico gRNA specificity

The selected gRNA and the search for corresponding crRNA target sequences in the full PhytoREF databases resulted in a perfect match (100%) with 969 of the 1068 available diatom sequences (~90%), spanning various taxonomic lineages (Fig. 5A). Of 99 sequences not matching the crRNA, 85 had at least no genus-level annotation, rendering their taxonomic identification elusive. Identification of potential Cas9 off-targets based on the 16S rRNA gene SSU sequences obtained from the SILVA V138 database resulted in 100% matches to only 644 sequences annotated as chloroplast and one mitochondrial hit. However, there were no other matches to any of the 451 459 16S rRNA gene SSU sequences, indicating no potential Cas9 off-targets for the selected gRNA. Low-probability potential off-targets (i.e. considering a gRNA sequence match of 96%, [25]) were identified for 117 16S rRNA gene SSUs, representing only a fraction of 0.025% of the sequence representatives in the SILVA V138 database. The evaluation of the phylogenetic range of the diatom chloroplast sequences based on the crRNA target sequence and the outer and inner primers used for SCHoCO-seq resulted in 62 matches, spanning at least three families and 12 orders (Fig. 5B). It should be noted that this relatively low number compared to the gRNA matches is due to the fact that only ~20% of sequences in PhytoREF are long enough to fully cover both primer binding sites (Fig. 6).

Discussion

This study demonstrated that the diatom microbiomes under investigation were mostly strain-specific and that the different micro-/macronutrient limitations treatments modulated the abundance of single microbiome members. The SCHoCO-seq treatment enabled in depth assessment of the microbial diversity in algal microbiomes on single-cell host level by reducing chloroplast ASV reads while increasing the sequencing depth for bacterial ASVs. Consequently, SCHoCO-seq led to a higher observed microbiome richness and detected low-abundance microbial taxa which were not found with standard 16S rRNA gene single-cell metabarcoding. Our *in silico* analyses identified a 100% match to 969 of the potential 1068 crRNA target sequences in PhytoREF databases, suggesting the newly developed protocol to be applicable to a wide range of diatom species with no apparent 16S rRNA gene off-target effects.

With respect to genotype-specific microbiomes of *T. gravis*, strain A1 and A2 were more similar to each other than to strain A5. Despite differing dates of isolation, variations in microbiome composition were most likely explained by the biogeographic distance of their sampling origin (Table S1), as microbiomes of laboratory cultures have been shown to be stable over long periods of time [42]. Yet, although strain A1 and A2 originated from



Figure 3. Normalized (subsamped and power-transformed [$X^{0.25}$]) 16S rRNA gene ASV read counts (indicated by continuous color scale, gray = not detected) summarized on the genus level (y-axis) across *T. gravida* single-cell samples (x-axis). *T. gravida* strains and culture conditions are indicated by vertical facets. Genera are ordered by Phylum (displayed as discrete color scale (right)). Genera occurring in <4 samples were excluded from the plot. Single-cell sample treatments indicated with; full medium = modified K medium, -Vit = vitamin limited medium, -N = nitrogen limited medium, EXT = single-cell DNA extraction (SC-EXT), PCR = direct single-cell PCR (SC-PCR).

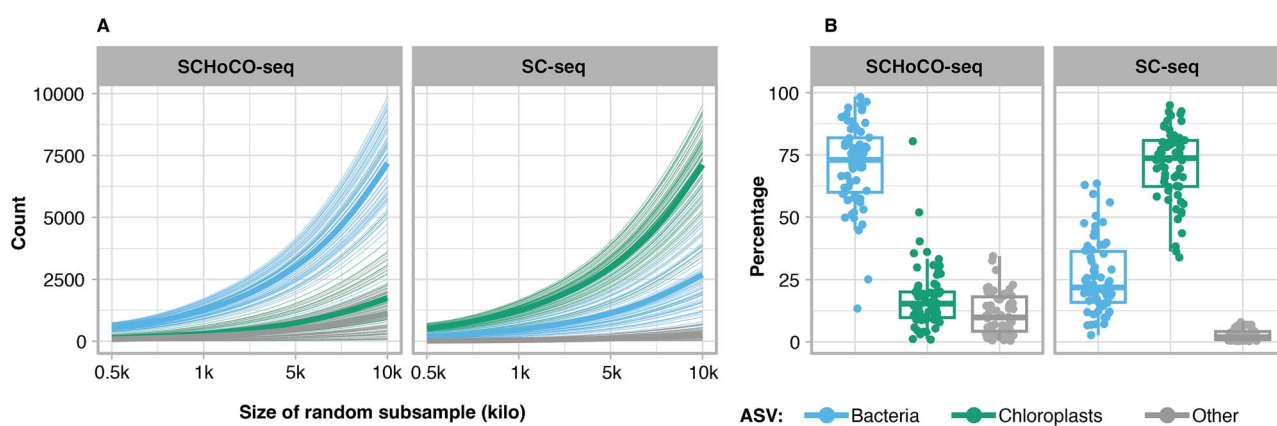


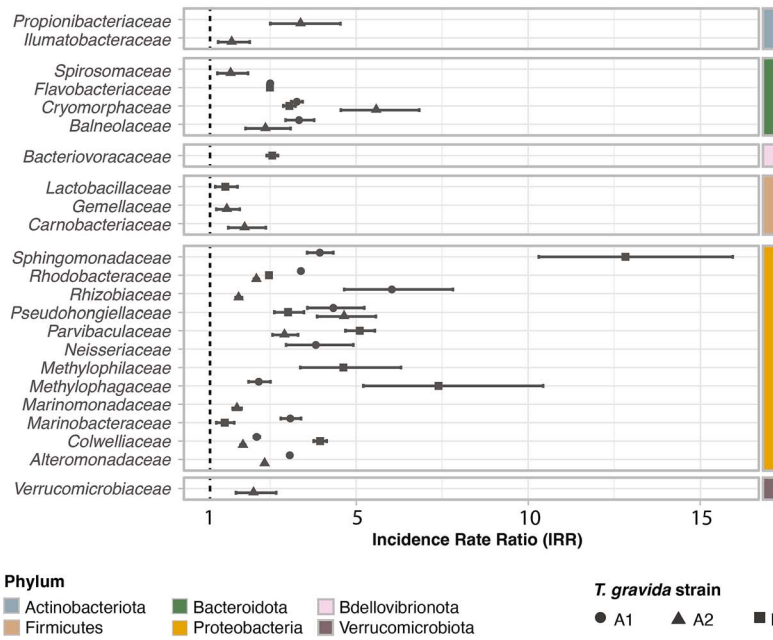
Figure 4. Effect of Cas9 plasmid 16S rRNA gene amplicon digest on total read composition and as a function of a range of assumed sequencing depths. The line plots show the theoretical composition of each sample if it was sequenced at different depths (obtained from random subsamples of the original dataset). Thin lines represent the ASV composition of each of the 62 samples after (SCHoCO-seq) and prior (SC-seq) to the cas9 digest. Bold lines are means obtained from the respective ASV compositions (A). Boxplots represent the percentage of the actual read composition of the ASVs, summed by annotation. “Bacteria” refers to all bacterial ASVs after removal of (A) potential contamination and (B) ASVs without annotations at least at the order level. “Chloroplasts” refers to all ASVs annotated as chloroplasts, and “others” refers to mitochondria-derived ASVs and ASVs with no annotation at least at the order level (B).

the same location, their microbiomes still showed differences in richness and diversity, highlighting the necessity for a single-cell perspective when profiling natural phytoplankton microbiomes.

Previous studies discussed the possibility of a core microbiome, i.e. a fraction of microbiome bacteria consistently present across diatom genotypes [43, 44]. However, the probability to detect a core microbiome likely decreases with increasing sampling size, biogeographic distance and region. For example, Ahern et al. [45] found no shared ASVs across 85 *T. rotula* strains from various ocean provinces. Yet, for the *T. gravida* strains in this study, a shared core community across the studied

genotypes was indeed found, although not consistently across all single-cell replicates. Remarkably, the shared core microbial community consisted of the genera *Colwellia*, *Roseobacter* clade NAC11-7 lineage as well as *Sulfotobacter*, also belonging to the *Roseobacter* group. Particularly the *Sulfotobacter* genus evolved complex mutualistic-symbiotic interactions with phytoplankton of which the underlying mechanisms are comparably well understood and include bacteria-derived provision of phyto-hormones and ammonium for their host [46, 47], as well as specific antibiotics to protect the holobiont from parasitic bacteria [48]. In exchange, *Sulfotobacter* receive various forms of

A: Incidence Rate Ratios with Confidence Intervals



B: Comparison of ASV richness

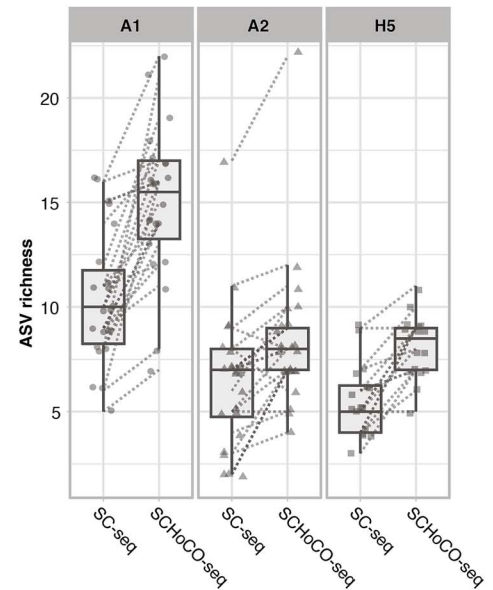
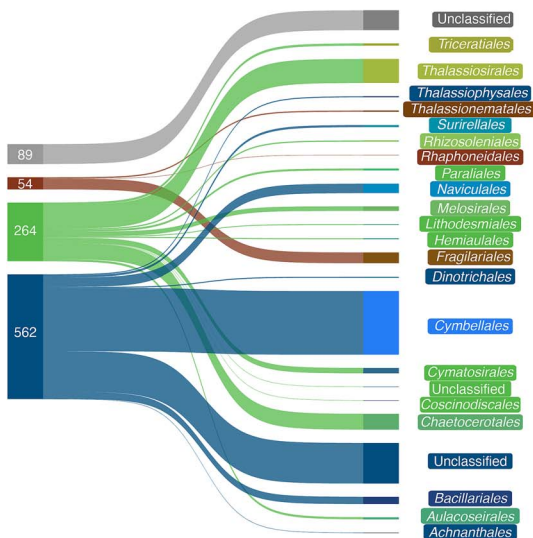


Figure 5. Performance comparison of 62 SCHO-seq and SC-seq sample pairs on (A) the differences in the probability of detecting a given bacterial family level based on ASV counts per sample and (B) the actual total ASV richness. In A, bacterial families for which a significant change in the incidence rate ratio could be detected are shown, with error bars representing the 95% confidence interval. The dotted line marks the boundary between a decrease in IRR (<1) and an increase in IRR (>1).

A: gRNA matches (100%)



B: verified primer & gRNA matches (100%)

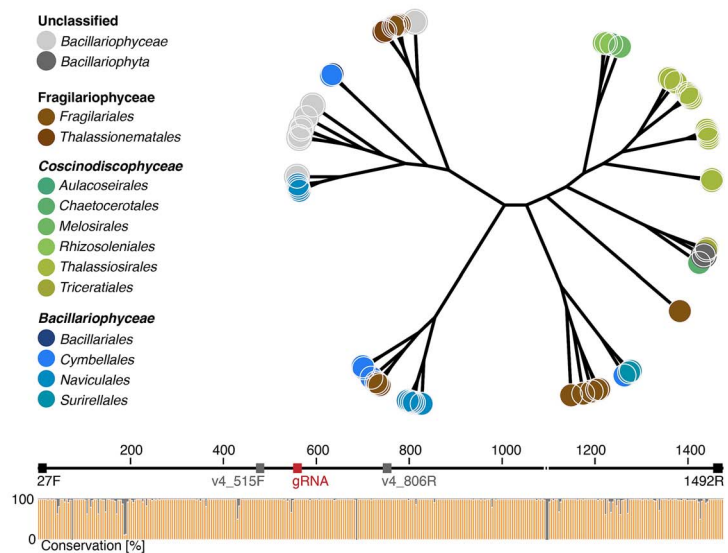


Figure 6. Coverage of taxonomic lineages with perfect gRNA sequence matches (A) and perfect matches with all primer pairs used in this study (B). Numbers in (A) refer to the gRNA chloroplast sequence matches for each diatom family. The bottom panel in (B) shows the primer binding sites and nucleotide conservation of 62 diatom chloroplast sequences; the corresponding unrooted phylogenetic tree shows the order-level annotations as indicated in the PhytoREF database.

exudated organic carbon [8] from the diatom host, as well as dimethylsulfoniopropionate (DMSP) for sulfur oxidation, and amino acids like tryptophan and taurine as organic nitrogen source [46, 47, 49].

Moreover, the *T. gravida* microbiomes responded to nitrogen- or vitamin-limited culture conditions with compositional changes of few specific bacterial genera. Considering that the nitrogen- and vitamin-limited cultures were inoculated from a full medium culture, the fact that certain bacterial groups were only retrieved in the limitation treatments suggests that these low-abundance

microbiome taxa may become abundant only during abiotic stress. This is exemplified for strain A1, where *Marinobacter* increased in abundance in the vitamin-limited treatments. *Marinobacter* are known for their ability to produce siderophores, potentially providing iron to their host [50, 51]. *Marinobacter* have also been associated with potential vitamin supply for the marine dinoflagellate *Lingulodinium polyedrum* [52]. For the genus *Balneola*, an increase in the nutrient-limited treatments was observed for *T. gravida* strain A1. However, these increases were not necessarily due to a mutualistic interaction but may

also indicate opportunistic mechanisms. In fact, Zhu et al. [53] hypothesized that *Balneo* has the ability to degrade specific organic compounds released by plants and potentially also phytoplankton upon encounter of nutrient stress. Moreover, diatoms have been shown to regulate their microbiome by target-specific secondary metabolites like rosmarinic acid or azelaic acid [54], yet whether this was the case in our experiments remains elusive.

Studying microbiome diversity in laboratory cultures is influenced not only by the fact that most heterotrophic bacteria cannot be cultured [55], but also by the dynamics of the microbiome community itself. Certain culture conditions may favor the growth of specific bacterial groups, thus underestimating bacterial richness. Conversely, shifts in evenness may allow the detection of bacterial groups that perform essential services for their host and represent an insurance for the holobiont in case of perturbations, while others remain undetected. This highlights the importance of considering microbiome community dynamics and the potential for different culture conditions to reveal distinct aspects of the microbiome's diversity and functional roles. Additionally, considering the isolation bias in terms of species- and even genotype-specificity of microbiome composition [18, 45], our findings underline the necessity for *in-situ* single-cell data on microbiome composition.

The bacterial richness in *T. gravida* microbiomes obtained in this study at the single-cell level by SCHO-co-seq agrees with richness data obtained from bulk diatom samples ranging from 1 to 125 ASVs or OTUs, respectively [18, 43, 45, 56]. To recover the full bacterial community richness on a given single diatom cell (preceding the SCHO-co-seq protocol), both the SC-PCR and SC-EXT methods developed in this study have been equally accurate as reflected by insignificant statistical test results. Both single-cell methods are therefore applicable depending on the research objective. For example, the SC-EXT approach offers the advantage of enabling multiple PCR reactions for further genomic DNA-based investigations, such as metagenomics or host identification via 18S sequencing, and the long-term storage of the sample [57].

The SCHO-co-seq approach is a promising strategy to accurately analyze diatom host-associated microbiome ASVs. This is largely due to the reduction of chloroplast reads, which in turn boost the detection and resolution of the microbiome richness. Achieving only a partial digestion of the target sequences is consistent with recent reports demonstrating that Cas9-mediated digestion is often incomplete under standard conditions. Accomplishing complete cleavage may require careful balancing of sgRNA, enzyme, and substrate concentrations; however, even then, technical limitations frequently persist [58].

Furthermore, our subsampling approach showed that the increase of microbiome ASVs over various sequencing depths by SCHO-co-seq was higher than the corresponding increase in chloroplast ASVs (Fig. 3). Importantly, this pattern was the opposite without Cas9 treatment, suggesting that merely increasing the sequencing depth is not an efficient strategy to recover low-abundance microbiome ASVs. Therein, SCHO-co-seq is particularly valuable in resolving the bacterial diversity of diatom microbiomes that harbor low bacterial abundance in their phycosphere. Thus, especially due to its potential to reveal higher community richness and identify potentially low-abundance bacterial taxa in the phycosphere, SCHO-co-seq has the potential to become the preferred method in diatom microbiome diversity research. Indeed, by identifying low-abundance bacterial taxa, taxon shifts, gains and losses, the overall genetic diversity in diatom microbiomes can be more accurately assessed, thereby increasing our knowledge of microbiome community assembly

[44], regulation and modulation. As microbiome genetic diversity and variation are aligned to selection [59], these capabilities are particularly relevant under current global change conditions. SCHO-co-seq therefore bears potential to assess and distinguish diatom microbiome responses to environmental change, e.g. plasticity, resilience, functional redundancy, and dysbiosis of diatom holobionts [59, 60].

Conclusion

This study revealed (i) genotype-specific diatom microbiomes identified through single-cell isolation and (ii) that diatom holobionts respond to macro- and micro-nutrient limitations with compositional changes of their microbiome. To increase the discovery of low-abundance bacterial taxa as part of diatom microbiomes, we have developed SCHO-co-seq which increases the bacterial biodiversity (richness) by reducing the chloroplast 16S ASVs. Thus, this novel single-cell microbiome analysis may serve as an important steppingstone towards *in-situ* single-cell microbiome analysis. The latter will help to understand how phytoplankton-associated microbiomes function in natural systems. In addition, the method presented here for reducing host DNA contamination with the aid of Cas9 can be applied to various other host microbiome systems by adapting the sequence of the guide RNA.

Acknowledgements

The authors thank the working group for their work.

Author contributions

Sylke Wohlrab: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing—original draft, Writing—review & editing. Ruben Schulte-Hillen: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing—original draft, Writing—review & editing. Jakob K. Giesler: Conceptualization, Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing—original draft, Writing—review & editing. Nancy Kühne: Investigation. Uwe John: Funding acquisition, Supervision, Writing—review & editing. Stefan Neuhaus: Data curation. Tilmann Harder: Conceptualization, Writing—review & editing. Thomas Mock: Conceptualization, Funding acquisition, Resources, Writing—review & editing. Nigel Belshaw: Resources.

Ruben Schulte-Hillen and Jakob Giesler contributed equally to this work and share first authorship. For the purposes of applications, proposals and outreach, either author may list themselves first.

Supplementary material

Supplementary material is available at ISME Communications online.

Conflicts of interest

The authors declare there is no conflict of interest.

Funding

This research was funded by the Helmholtz research program 'Changing Earth, Sustaining our Future' (subtopic 6.2 'Adaptation of marine life: from genes to ecosystems' in topic 6 'Marine

and Polar Life') of the Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Germany. Furthermore, we acknowledge the support by the Open Access Publication Funds of Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung.

Data availability

All raw sequencing data supporting the findings of this study have been deposited in the NCBI BioProject database under the accession number (BioProject ID): PRJNA1335164.

References

- Margulis L. Evolution of eukaryotes. *Am Sci* 1991;**79**:187–8.
- Dittami SM, Arboleda E, Auguet JC. et al. A community perspective on the concept of marine holobionts: current status, challenges, and future directions. *PeerJ* 2021;**9**:e10911. <https://doi.org/10.7717/peerj.10911>
- Giesler JK, Harder T, Wohlrab S. Microbiome and photoperiod interactively determine thermal sensitivity of polar and temperate diatoms. *Biol Lett* 2023;**19**:20230151. <https://doi.org/10.1098/rsbl.2023.0151>
- Amin SA, Parker MS, Armbrust EV. Interactions between diatoms and bacteria. *Microbiol Mol Biol Rev* 2012;**76**:667–84. <https://doi.org/10.1128/MMBR.00007-12>
- Kodama Y, Fujishima M. Endosymbiotic *Chlorella variabilis* reduces mitochondrial number in the ciliate *Paramecium bursaria*. *Sci Rep* 2022;**12**:8216. <https://doi.org/10.1038/s41598-022-12496-8>
- Cirri E, Pohnert G. Algae-bacteria interactions that balance the planktonic microbiome. *New Phytol* 2019;**223**:100–6. <https://doi.org/10.1111/nph.15765>
- Kuhlisch C, Shemi A, Barak-Gavish N. et al. Algal blooms in the ocean: hot spots for chemically mediated microbial interactions. *Nat Rev Microbiol* 2024;**22**:138–54. <https://doi.org/10.1038/s41579-023-00975-2>
- Martin K, Schmidt K, Toseland A. et al. The biogeographic differentiation of algal microbiomes in the upper ocean from pole to pole. *Nat Commun* 2021;**12**:5483. <https://doi.org/10.1038/s41467-021-25646-9>
- Needham DM, Poirier C, Bachy C. et al. The microbiome of a bacterivorous marine choanoflagellate contains a resource-demanding obligate bacterial associate. *Nat Microbiol* 2022;**7**:1466–79. <https://doi.org/10.1038/s41564-022-01174-0>
- Boscaro V, Manassero V, Keeling PJ. et al. Single-cell Microbiomics unveils distribution and patterns of microbial symbioses in the natural environment. *Microb Ecol* 2023;**85**:307–16. <https://doi.org/10.1007/s00248-021-01938-x>
- Rossi A, Bellone A, Fokin SI. et al. Detecting associations between ciliated Protists and prokaryotes with culture-independent single-cell Microbiomics: a proof-of-concept study. *Microb Ecol* 2019;**78**:232–42. <https://doi.org/10.1007/s00248-018-1279-9>
- Schmitz TC, Dede Eren A, Spierings J. et al. Solid-phase silica-based extraction leads to underestimation of residual DNA in decellularized tissues. *Xenotransplantation* 2021;**28**:e12643. <https://doi.org/10.1111/xen.12643>
- Probandt D, Eickhorst T, Ellrott A. et al. Microbial life on a sand grain: from bulk sediment to single grains. *ISME J* 2018;**12**:623–33. <https://doi.org/10.1038/ismej.2017.197>
- Lundberg DS, Lebeis SL, Paredes SH. et al. Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 2012;**488**:86–90. <https://doi.org/10.1038/nature11237>
- Zarraonaindia I, Owens SM, Weisenborn P. et al. The soil microbiome influences grapevine-associated microbiota. *MBio* 2015;**6**:1–10. <https://doi.org/10.1128/mBio.02527-14>
- Orum H, Nielsen PE, Egholm M. et al. Single base pair mutation analysis by PNA directed PCR clamping. *Nucleic Acids Res* 1993;**21**:5332–6. <https://doi.org/10.1093/nar/21.23.5332>
- Jackrel SL, Owens SM, Gilbert JA. et al. Identifying the plant-associated microbiome across aquatic and terrestrial environments: the effects of amplification method on taxa discovery. *Mol Ecol Resour* 2017;**17**:931–42. <https://doi.org/10.1111/1755-0998.12645>
- Baker LJ, Kemp PF. Exploring bacteria–diatom associations using single-cell whole genome amplification. *Aquat Microb Ecol* 2014;**72**:73–88. <https://doi.org/10.3354/ame01686>
- Song L, Xie K. Engineering CRISPR/Cas9 to mitigate abundant host contamination for 16S rRNA gene-based amplicon sequencing. *Microbiome* 2020;**8**:80. <https://doi.org/10.1186/s40168-020-00859-0>
- Keller MD, Selvin RC, Claus W. et al. Media for the culture of oceanic ULTRAPHYTOPLANKTON1,2. *J Phycol* 2007;**23**:633–8. <https://doi.org/10.1111/j.1529-8817.1987.tb04217.x>
- Frank JA, Reich CI, Sharma S. et al. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* 2008;**74**:2461–70. <https://doi.org/10.1128/AEM.02272-07>
- Lane DJ. 16S/23S rRNA sequencing. *Nucleic Acid Techniques in Bacterial Systematics* 1991;115–75.
- Decelle J, Romac S, Stern RF. et al. PhytoREF: a reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy. *Mol Ecol Resour* 2015;**15**:1435–45. <https://doi.org/10.1111/1755-0998.12401>
- Hopes A, Nekrasov V, Belshaw N. et al. Genome editing in diatoms using CRISPR-Cas to induce precise Bi-allelic deletions. *Bio Protoc* 2017;**7**:e2625. <https://doi.org/10.21769/BioProtoc.2625>
- Hsu PD, Scott DA, Weinstein JA. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 2013;**31**:827–32. <https://doi.org/10.1038/nbt.2647>
- Quast C, Pruesse E, Yilmaz P. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;**41**:D590–6. <https://doi.org/10.1093/nar/gks1219>
- Anderson EM, Haupt A, Schiel JA. et al. Systematic analysis of CRISPR-Cas9 mismatch tolerance reveals low levels of off-target activity. *J Biotechnol* 2015;**211**:56–65. <https://doi.org/10.1016/j.jbiotec.2015.06.427>
- Lassmann T. Kalign 3: multiple sequence alignment of large data sets. *Bioinformatics* 2019;**36**:1928–9. <https://doi.org/10.1093/bioinformatics/btz795>
- Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 2009;**26**:1641–50. <https://doi.org/10.1093/molbev/msp077>
- Hiranniramol K, Chen YH, Liu WJ. et al. Generalizable sgRNA design for improved CRISPR/Cas9 editing efficiency. *Bioinformatics* 2020;**36**:2684–9. <https://doi.org/10.1093/bioinformatics/btaa041>
- Ahme A, Von Jackowski A, McPherson RA. et al. Winners and losers of Atlantification: the degree of ocean warming affects the structure of Arctic microbial communities. *Genes (Basel)* 2023;**14**:623. <https://doi.org/10.3390/genes14030623>
- Apprill A, McNally S, Parsons R. et al. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11

- bacterioplankton. *Aquat Microb Ecol* 2015;**75**:129–37. <https://doi.org/10.3354/ame01753>
33. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* 2016;**18**:1403–14. <https://doi.org/10.1111/1462-2920.13023>
 34. Martin M. CUTADAPT removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 2011;**17**:17. <https://doi.org/10.14806/ej.17.1.200>
 35. Callahan BJ, McMurdie PJ, Rosen MJ. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;**13**:581–3. <https://doi.org/10.1038/nmeth.3869>
 36. Reitmeier S, Hitch TCA, Treichel N. et al. Handling of spurious sequences affects the outcome of high-throughput 16S rRNA gene amplicon profiling. *ISME. Communications* 2021;**1**:31. <https://doi.org/10.1038/s43705-021-00033-z>
 37. Sheik CS, Reese BK, Twing KI. et al. Identification and removal of contaminant sequences from ribosomal gene databases: lessons from the census of deep life. *Front Microbiol* 2018;**9**:840. <https://doi.org/10.3389/fmicb.2018.00840>
 38. Beule L, Karlovsky P. Improved normalization of species count data in ecology by scaling with ranked subsampling (SRS): application to microbial communities. *PeerJ* 2020;**8**:e9593. <https://doi.org/10.7717/peerj.9593>
 39. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013;**8**:e61217. <https://doi.org/10.1371/journal.pone.0061217>
 40. Lahti L, Shetty S. Tools for Microbiome Analysis in R. Vienna, Austria: The R Foundation for Statistical Computing; 2017. Available at: <http://microbiome.github.io/microbiome>
 41. Oksanen J, Blanchet FG, Kindt R. et al. vegan: Community Ecology Package. R package version 2.0-10. Vienna, Austria: The R Foundation for Statistical Computing; 2013. Available at: <https://CRAN.R-project.org/package=vegan>
 42. Barreto Filho MM, Walker M, Ashworth MP. et al. Structure and long-term stability of the microbiome in diverse diatom cultures. *Microbiol Spectr* 2021;**9**:e0026921. <https://doi.org/10.1128/Spectrum.00269-21>
 43. Behringer G, Ochsenkuhn MA, Fei C. et al. Bacterial communities of diatoms display strong conservation across strains and time. *Front Microbiol* 2018;**9**:659. <https://doi.org/10.3389/fmicb.2018.00659>
 44. Monnich J, Tebben J, Bergemann J. et al. Niche-based assembly of bacterial consortia on the diatom *Thalassiosira rotula* is stable and reproducible. *ISME J* 2020;**14**:1614–25. <https://doi.org/10.1038/s41396-020-0631-5>
 45. Ahern OM, Whittaker KA, Williams TC. et al. Host genotype structures the microbiome of a globally dispersed marine phytoplankton. *Proc Natl Acad Sci U S A*. 2021;**118**:e2105207118. <https://doi.org/10.1073/pnas.2105207118>
 46. Amin SA, Hmelo LR, van Tol HM. et al. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* 2015;**522**:98–101. <https://doi.org/10.1038/nature14488>
 47. Segev E, Wyche TP, Kim KH. et al. Dynamic metabolic exchange governs a marine algal-bacterial interaction. *elife* 2016;**5**:5. <https://doi.org/10.7554/eLife.17473>
 48. Beiralas R, Ozer N, Segev E. Abundant Sulfatobacter marine bacteria protect *Emiliania huxleyi* algae from pathogenic bacteria. *ISME Commun* 2023;**3**:100. <https://doi.org/10.1038/s43705-023-00311-y>
 49. Yang Q, Ge YM, Iqbal NM. et al. Sulfatobacter alexandrii sp. nov., a new microalgae growth-promoting bacterium with exopolysaccharides bioflocculating potential isolated from marine phytoplankton. *Antonie Van Leeuwenhoek* 2021;**114**:1091–106. <https://doi.org/10.1007/s10482-021-01580-0>
 50. Amin SA, Green DH, Kupper FC. et al. Vibrioferrin, an unusual marine siderophore: iron binding, photochemistry, and biological implications. *Inorg Chem* 2009;**48**:11451–8. <https://doi.org/10.1021/ic9016883>
 51. Butler A, Harder T, Ostrowski AD. et al. Photoactive siderophores: structure, function and biology. *J Inorg Biochem* 2021;**221**:111457. <https://doi.org/10.1016/j.jinorgbio.2021.111457>
 52. Cruz-Lopez R, Maske H. The vitamin B1 and B12 required by the marine dinoflagellate *Lingulodinium polyedrum* can be provided by its associated bacterial community in culture. *Front Microbiol* 2016;**7**:560. <https://doi.org/10.3389/fmicb.2016.00560>
 53. Zhu Z, Meng R, Smith WO Jr. et al. Bacterial composition associated with Giant colonies of the harmful algal species *Phaeocystis globosa*. *Front Microbiol* 2021;**12**:737484. <https://doi.org/10.3389/fmicb.2021.737484>
 54. Shibl AA, Isaac A, Ochsenkuhn MA. et al. Diatom modulation of select bacteria through use of two unique secondary metabolites. *Proc Natl Acad Sci USA* 2020;**117**:27445–55. <https://doi.org/10.1073/pnas.2012088117>
 55. Vartoukian SR, Palmer RM, Wade WG. Strategies for culture of ‘unculturable’ bacteria. *FEMS Microbiol Lett* 2010;**309**:1–7. <https://doi.org/10.1111/j.1574-6968.2010.02000.x>
 56. Sison-Mangus MP, Jiang S, Tran KN. et al. Host-specific adaptation governs the interaction of the marine diatom, *Pseudo-nitzschia* and their microbiota. *ISME J* 2014;**8**:63–76. <https://doi.org/10.1038/ismej.2013.138>
 57. Hallmaier-Wacker LK, Lueert S, Roos C. et al. The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis. *Sci Rep* 2018;**8**:6292. <https://doi.org/10.1038/s41598-018-24573-y>
 58. Ma X, Kthiri D, Gill MS. et al. Cas9 endonuclease: a molecular tool for in vitro cloning and CRISPR edit detection. *Front Genome Ed* 2025;**7**:1565297. <https://doi.org/10.3389/fgeed.2025.1565297>
 59. Baltar F, Bayer B, Bednarsek N. et al. Towards integrating evolution, metabolism, and climate change studies of marine ecosystems. *Trends Ecol Evol* 2019;**34**:1022–33. <https://doi.org/10.1016/j.tree.2019.07.003>
 60. Graham EB, Knelman JE, Schindlbacher A. et al. Microbes as Engines of Ecosystem Function: when does community structure enhance predictions of ecosystem processes? *Front Microbiol* 2016;**7**:214. <https://doi.org/10.3389/fmicb.2016.00214>