PNAS
www.pnas.org

Supplementary Information for

Diatom Modulation of Select Bacteria Through Use of Two Unique Secondary Metabolites

Ahmed A. Shibl, Ashley Isaac, Michael A. Ochsenkühn, Anny Cárdenas, Cong Fei, Gregory Behringer, Marc Arnoux, Nizar Drou, Miraflor P. Santos, Kristin C. Gunsalus, Christian R. Voolstra, Shady A. Amin

Shady A. Amin
Email: samin@nyu.edu

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.2012088117

Supplementary Methods

Diatom isolation and growth. *Asterionellopsis glacialis* strain A3 was isolated from the Persian Gulf and identified as described previously (1). All cultures were maintained in *f/2*+Si medium (2) in semi-continuous batch cultures (3) and incubated in growth chambers (Percival, Perry, IA) at 22°C, 125 µE m^{-2} s^{-1}, and a 12:12 light/dark cycle. Light flux was measured using a QSL-2100 PAR Sensor (Biospherical Instruments Inc., San Diego, CA). Growth was monitored by measuring *in vivo* fluorescence using a 10-AU fluorometer (Turner Designs, San Jose, CA). All cultures were acclimated throughout the experiments as described below for at least three transfers using semi-continuous batch cultures. Cultures were considered acclimated if the growth rates of three consecutive transfers of triplicate cultures did not vary by more than 15%. Specific growth rates (μ) were calculated from the linear regression of the natural log of *in vivo* fluorescence versus time during the exponential growth phase of cultures. Standard deviation of μ was calculated using μ values from biological replicates over the exponential growth period.

Microbial consortium reseeding experimental design. To examine the interactions between *A. glacialis* A3 and its bacterial consortium, the diatom was first made axenic as described previously (1). In brief, approximately 25 mL of a late-exponential phase growing *A. glacialis* A3 culture was gravity filtered onto a 0.65-µm pore-size polycarbonate membrane filter (Millipore). Cells were quickly rinsed with sterile *f/2*+Si media. Using sterile tweezers, the filter was removed from the filtration unit and washed for ~1 minute in sterile media containing 20 mg/mL Triton X-100 detergent to remove surface-attached bacteria. The filter was discarded after re-suspension of cells by gentle shaking in sterile detergent-free media. Cells were again gravity filtered onto a fresh 0.65-µm pore-size polycarbonate membrane filter and rinsed with sterile media. Subsequently, cells were washed off the filter by gentle shaking into sterile media containing a suite of antibiotics (per mL: 50 µg streptomycin, 66.6 µg gentamycin, 20 µg ciprofloxacin, 2.2 µg chloramphenicol, and 100 µg ampicillin). Cells were then incubated in antibiotic-containing media for 48 hours under regular growth conditions. Finally, 0.5–1.0 mL of antibiotics-treated cells was transferred to antibiotic-free media. Cultures were regularly monitored for bacterial contamination by checking for bacterial growth in Zobell marine broth 2216 (HiMedia) (4) in addition to filtering 2-3 mL of exponential-phase growing culture and using Sybr Green I (Invitrogen) staining and epifluorescence microscopy (Nikon Eclipse 80i) as described previously (5). This axenic *A. glacialis* A3 culture was left to acclimate to no bacteria for ~170 generations and was subsequently used for the reseeding experiment.

To conduct the reseeding experiment, axenic and xenic *A. glacialis* A3 cultures were acclimated to growth in 1 L batch cultures. Xenic and axenic *A. glacialis* A3 cultures were grown side by side to allow the harvesting of the true bacterial consortium composition and adding it to the axenic *A. glacialis* A3. To begin the experiment, 6 L of axenic and 9 L of xenic *A. glacialis* A3 culture batches were inoculated at the same cell density (~ 5,000 diatom cells/mL) and time. Once both cultures reached a diatom cell density of ~1×10^5 cells/mL, the xenic cultures were pooled, gently sonicated to detach diatom-attached bacteria and filtered through a sterile 3-µm polycarbonate filter (25 mm, Whatman, NJ, United States) to remove diatom cells; the filtrate containing the microbial community was collected in a sterile flask and was used for subsequent steps. The filtrate was subsequently centrifuged at 4,000 rpm for 20 minutes using an Avanti J-26 XPI centrifuge (Beckman Coulter, Inc.) to concentrate the bacterial consortium and remove residual organic carbon from the media. The bacterial pellet was washed with sterile *f/2*+Si once, centrifuged and subsequently reconstituted in 3 mL of sterile media. Bacterial cell density was enumerated using epifluorescence microscopy (Nikon Eclipse 80i) as described previously (5). This bacterial

consortium stock was divided into three parts, each containing ~ 9 × 10^5 cells/mL; 1) a third of the sample was used to isolate DNA for bacterial consortium metagenomics; 2) a third of the sample was incubated in triplicate 250 mL sterile *f/2*+Si media for 0.5 hours [this sample was used for the bacterial consortium RNA control with no diatom (bacterial consortium control)]; 3) the remainder of the sample was added to triplicate 1 L bottles of the acclimated axenic *A. glacialis* A3 (reseeded diatom). The remainder of the axenic *A. glacialis* A3 culture (3 L) served as triplicate axenic control (axenic diatom). This scheme ensured that the reseeded diatom samples contained similar diversity and density of bacteria and diatom relative to the original xenic culture. We avoided adding bacteria from natural seawater to ensure our experiments included originally isolated diatom symbionts. The beginning of the reseeding experiment (*t*=0) is marked by the addition of the bacterial consortium to the axenic diatom. A simplified schematic of the experimental design and diatom-bacterial consortium growth is shown in Fig. S1A.

Diatom DNA and RNA isolation and sequencing. DNA was isolated from axenic *A. glacialis* A3 by filtering cells onto a 3-µm polycarbonate filter and using the Wizard SV Genomic DNA Purification kit (Promega) following the manufacturer's instructions and quantified on a Qubit 3.0 fluorometer using the DNA high sensitivity assay kit (Thermo-Fisher Scientific). The diatom DNA library was prepared with 200 ng of starting material using the TruSeq DNA Nano kit (Illumina, San Diego, CA, USA).

For axenic diatom RNA samples, cultures were filtered through 3-µm polycarbonate filters (25 mm, Whatman, NJ, United States) at 0.5 hours and 24 hours after the beginning of the reseeding experiment. For reseeded diatom samples, cultures were filtered through 3-µm polycarbonate filters to obtain diatom-enriched samples then through 0.2-µm polycarbonate filters to obtain bacterial consortium samples (Fig. S1A). All filters were flash frozen in liquid nitrogen and later stored in -80°C until further processing. From the 3-µm filters, cells were lysed by bead beating with sterile beads (Sigma) for 10 minutes followed by total RNA isolation using the ToTALLY RNA total RNA Isolation kit (Ambion) according to the manufacturer's instructions. Samples were treated with two rounds of DNase to remove contaminating DNA using Turbo-DNase (Ambion). Ribosomal RNA (rRNA) were removed using the Poly(A)Purist MAG kit (Thermo-Fisher Scientific) following the manufacturer's instructions. The mRNA was amplified using a MessageAmp II aRNA Amplification kit (Invitrogen) according to the manufacturer's instructions. RNA libraries were prepared with a maximum of 50 µL starting material, as per protocol instructions, using the TruSeq RNA v2 kit (Illumina, San Diego, CA, USA).

The resulting libraries' concentrations and size distributions were assessed on a Qubit 3.0 fluorometer using the DNA high sensitivity assay kit (Thermo-Fisher Scientific) and a Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA). Following this, libraries were normalized, pooled and quantified by qPCR with the KAPA Library quantification kit for Illumina platforms (Kapa Biosystems, Wilmington MA, USA) on a StepOnePlus qPCR system (Thermo-Fisher Scientific). One replicate library, B7, belonging to the reseeded bacterial samples at 24 hours was discarded due to low quality. Finally, samples were loaded at 12 pM with 2% phiX on a High Output FlowCell and paired-end sequenced (2x100 bp) on the Illumina HiSeq 2500 platform available at the NYU Abu Dhabi Center for Genomics and Systems Biology (Table S8).

Diatom genome. Raw genomic reads were assessed with the FastQC v0.11.5(6) tool. Low-quality bases and sequencing adaptor contaminants were removed by the Trimmomatic v0.36 tool (7) with the following parameters: "ILLUMINACLIP:adapter.fa:2:30:10 TRAILING:3 LEADING:3 SLIDINGWINDOW:4:15 MINLEN:36". Quality trimmed reads were then *de novo* assembled on Platanus v1.2.4 (8) to yield 1,840 scaffolds of size >10kb out of 6,925 scaffolds with N50=21,686

and a total size of 66.5 Mbp. The final assembly was assessed for accuracy and completeness with QUAST v5.0.2 (9), and BUSCO v3 (10).

Diatom transcriptome. Raw RNAseq reads were quality trimmed as described for the genome. HISAT2 v2.0.4 (11) was used to map the reads to the assembled genome. Generated SAM files were converted to alignment files in BAM format and sorted by coordinates with SAMtools v1.5 (12). Using StringTie v1.3.0 (13), GTF files per sample were created then merged into one file representing the transcriptome. Transcript assemblies were annotated on the Trinotate pipeline (http://trinotate.github.io) following Bryant *et al* (14). Significant differences in gene expression between the samples were evaluated with DESeq2 v1.14.1 (15) at a false discovery rate (FDR) of 0.1 and a minimum $\log_2$-fold change of 0.5. Subcellular localization of gene products was determined on DeepLoc (16). Differentially expressed genes across different timepoints were visualized using the Circos package (17).

Bacterial consortium DNA and RNA isolation and sequencing. For bacterial consortium metagenomic samples, bacterial DNA was isolated from the bacterial consortium control sample using bead-beating with sterile beads (Sigma) for 10 minutes followed by the EZNA Bacterial DNA kit (Omega Bio-Tek) according to the manufacturer's instructions and then quantified on the Qubit 3.0 fluorometer using the DNA high sensitivity assay kit (Thermo-Fisher Scientific). The consortium metagenomic library was prepared with 200 ng of starting material using the TruSeq DNA Nano kit (Illumina, San Diego, CA, USA) and paired-end sequenced (2x100 bp) on the Illumina HiSeq 2500 platform.

For bacterial consortium RNA samples, consortium control samples were filtered through 0.2-µm polycarbonate filters (25 mm, Whatman, NJ, United States) at 0.5 hours after the beginning of the reseeding experiment. Reseeded *A. glacialis* A3 cultures were filtered through 3-µm then 0.2-µm polycarbonate filters at 0.5 hours and 24 hours after the beginning of incubation (Figure S1B). All filters were flash frozen in liquid nitrogen and later stored in -80°C until further processing. From the 0.2-µm filters, cells were lysed by bead beating with sterile beads (Sigma) for 10 minutes followed by total RNA isolation using the RNeasy Mini kit (Qiagen, Germantown, MD) according to the manufacturer's instructions. Samples were treated with two rounds of DNase to remove contaminating DNA using Turbo-DNase (Ambion). Ribosomal RNA (rRNA) were removed using the MicrobExpress Bacterial mRNA enrichment kit (Ambion) following the manufacturer's instructions. mRNA was amplified using the MessageAmp II-Bacteria RNA Amplification kit (Invitrogen) according to the manufacturer's instructions. RNA libraries were prepared with a maximum of 50 µL starting material, as per protocol instructions, using the TruSeq RNA v2 kit (Illumina, San Diego, CA, USA). The resulting library sizes and distributions were assessed on the Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA). Following this, libraries were normalized, pooled and quantified by qPCR with the KAPA Library quantification kit (Illumina, San Diego, CA, USA). Finally, samples were loaded at 12 pM with 2% phiX on a High Output FlowCell and paired-end sequenced (2x100 bp) on the Illumina HiSeq 2500 platform (Table S8).

Bacterial consortium metagenome, binning and assembly. Raw metagenomic reads were quality trimmed on Trimmomatic v0.36, with a minimum length of 75 bp. Quality-checked reads were mapped to the *A. glacialis* A3 genome on BBtools using the BBmap package v37.10 (http://sourceforge.net/projects/bbmap/) with default parameters. Reads that did not map to the diatom were then used as input for Kaiju v1.5.0 (18) to determine the taxonomic profile and abundance of the microbial community at the protein level. De novo assembly was done using MEGAHIT v1.0.2 (19) with a k-*mer* size of 127 and scaffolds were binned into metagenomically-assembled genomes (MAGs) on MetaBAT v0.25.4 (20). The MAGs were assessed for

completeness and contamination with CheckM v1.0.7 (21) then visualized and refined on Anvi'o v3 (22) until contamination values dropped below 5%. The closest genomic neighbor was determined by performing whole-genome comparisons of amino acid identities (AAI) using the Microbial Genomes Atlas (MiGA) (23) (Table S1). Functional annotation of the MAGs was performed on Prokka v1.12 (24).

Bacterial consortium metatranscriptomes. Raw RNAseq reads were quality trimmed as described above. Paired-end reads were merged on Flash v1.2.11 (25) and rRNA fragments were identified and removed using SortMeRNA v2.0 (26). Non-rRNA reads were mapped to protein-coding genes of the MAGs with Bowtie2 v2.3.3 (27) (Table S2). Resulting SAM files were used to quantify gene expression levels using eXpress v1.5.1 (28) considering only genes with a minimum read count of 10 per group. Significant differences in gene expression between the samples were evaluated with DESeq2 v1.14.1 at a false discovery rate (FDR) of 0.1 and a minimum $\log_2$-fold change of 0.5. To infer the functional potential of the entire bacterial consortium, functional profiling against UniRef50 (29) was performed using HUMAnN2 v0.11.2 (30). Gene families were further mapped to Gene Ontology (GO) terms (31) and structured into pathways with MetaCyc (32) to generate "copies per million (CPM)" values across the different conditions. Data plots were generated in R v3.4.3 (33) with RStudio v1.2.1335 (RStudio Inc., Boston, MA, USA) and packages ggplot2 v3.1.1 (34) and ggtern v3.1.0 (35).

Exometabolite extraction. All glassware used was acid washed (1.2 M HCl), rinsed with MilliQ-H_2O, furnace-baked at 420°C and sterilized for a minimum of 12 hours to eliminate residual organic carbon contamination. All solutions were made with either MilliQ-H_2O or LC-MS grade methanol (Thermo-Fisher). Cell-free filtrates from the axenic diatom and reseeded diatom samples at all timepoints (i.e. 0.5, 4, 24 and 48 hours, figure S1) were placed in 500-mL dark glass bottles (Thermo-Fisher), acidified to pH ~3 using 100% formic acid (Sigma). No bacterial consortium control was used because the consortium stock culture was free of carbon and would not survive. For QToF-MS, organic molecules were extracted by passing each replicate onto 500 mg Oasis HLB solid-phase extraction (SPE) cartridges (Waters, USA) using a peristaltic pump (MasterFlex Easy-Load 3, USA) at a flowrate of ~5 mL/min. All SPE cartridges were pre-conditioned according to the manufacturer's instructions. Salts were washed from the SPE cartridges using 0.1% trifluoroacetic acid in MilliQ-H_2O. Organic molecules were eluted into 5-mL borosilicate tubes (Thermo-Fisher) using 5% ammonium hydroxide in methanol. For FT-ICR-MS, organic molecules were extracted by passing samples as described for Q-ToF-MS except for the use of PPL Bond-Elut solid-phase extraction columns (Agilent Technologies, US), according to the manufacturer's instructions, instead of Oasis HLB. All extracts were immediately dried using a Savant SC210A SpeedVac concentrator (Thermo-Fisher) and stored at -80°C until analysis.

***UHPLC-QToF-MS*.** Metabolites were analyzed on a Bruker Impact II HD quadrupole time-of-flight mass spectrometer (QToF-MS, BrukerDaltonik GmbH, Bremen, Germany) coupled to an Agilent 1290 UHPLC system (Agilent, US). Metabolites were separated using a reversed-phase (RP) method, where medium-polarity and non-polar metabolites were separated using an Eclipse Plus C_{18} column (50mm × 2.1mm ID) (Agilent, US). Chromatographic mobile phases consisted of MilliQ-H_2O + 0.2% formic acid (buffer A), Acetonitrile + 0.2% formic acid (buffer B). The gradient started with 95% A and 5% B, with a gradient of 18 minutes to 100% B and 2 minutes at 100% B. Every run was followed by a 5-min wash step from buffer B to buffer A to isopropanol and back to the initial condition, where the column was equilibrated for another 2 minutes. Detection was carried out in positive and negative ionization modes with the following parameters: ESI settings: dry gas temperature = 220 °C, dry gas flow = 8.0 L/min, nebulizer pressure = 2.2 bar, capillary voltage = 4500 V, end plate offset = (-)500 V; MS-ToF setting: Funnel 1 RF = 150 Vpp, Funnel 2 RF = 200

Vpp, Hexapole RF = 50, Quadrupole Ion Energy = 1 eV, Collision Energy = 7 eV, untargeted MS/MS = stepping 30 - 50 eV; Acquisition Setting: mass range = 50 - 1300 m/z, Spectra rate = 6.0 Hz spectra/s, 1000 ms/spectrum. Auto MS/MS was performed in stepping mode, splitting each fragmentation scan equally into 25 and 50 eV.

Calibration, retention time alignment and peak picking of individual LC-MS runs were performed using the T-Rex 3D algorithm of Metaboscape v4.0 (BrukerDaltonik GmbH, Bremen, Germany). Background noise was removed by applying an intensity threshold of 1000. Peak-picking and integration were accompanied by ^{13}C cluster detection to verify molecular features and remove those which appear in less than 40% of samples. Peak annotation was performed using an in-house generated spectral library of 668 biomolecules from the Mass Spectrometry Metabolite Library of Standards (IROA Technologies, US) and further using the Bruker Personal MS/MS Library (BrukerDaltonik GmbH, Bremen, Germany). The acquired LC-MS data was normalized according to sample volume, scaled across all samples and log_{10}-transformed. Multivariate statistical analysis was performed using MetaboAnalyst v3.0 (36) on >1,200 molecular features (Dataset S1), including confirmed metabolites, to generate principal component analysis (PCA) plots for axenic vs. reseeded conditions at all timepoints. Mahalanobis distances were calculated on R v3.4.3. The heatmap of confirmed metabolites (Table S3) was visualized using the ComplexHeatmap package (37). Significance in relative abundance between different time points in reseeded and axenic samples were calculated using a Student's *t*-test (Bonferonni-adjusted $p<0.05$).

FT-ICR-MS. Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) was used to determine the molecular composition of dissolved organic matter (DOM) components in the exometabolome. High-resolution mass spectra were acquired on a SolariX FT-ICR-MS (BrukerDaltonik GmbH, Bremen, Germany) equipped with a 7 Tesla superconducting magnet and ParaCell analyzer. Samples were directly injected into an electrospray ionization (ESI) source (BrukerDaltonik GmbH, Bremen, Germany) at a flow rate of 2 µL/min operated in negative ionization mode with a capillary voltage = 4500 V, end plate offset = (-)500 V, nebulizer pressure = 2 bar, dry gas flow = 10 L/min, dry gas temperature = 220°C. Spectra were acquired with a time domain of four mega words in 2ω resonance mode over a mass range of m/z 80 to 1000, with an optimal mass range from 200-600 m/z. Three-hundred scans were accumulated for each sample. Spectra were internally calibrated with a fatty acids reference list on the DataAnalysis 5.0 software (Bruker, Germany). Peak alignment was performed with maximum error thresholds of 0.01 ppm. The FT-ICR-MS spectra were exported to peak lists with a cut-off signal/noise ratio of 3 and a minimal signal intensity of 10^6. Chemical formulae calculation was performed with an error threshold of 0.5 ppm from the exact mass for the chemical formula and isotopic fine structure. Chemical formulae were only generated if all theoretical isotope peaks (100%) were found in spectra (Datasets S2 and S3).

***Bacterial isolation, genomic DNA extraction, sequencing and assembly*.** To isolate individual bacterial strains from the bacterial consortium, 200 µL of the xenic *A. glacialis* A3 culture in log-phase diluted in sterile seawater were spread evenly on Zobell marine agar 2216 (HiMedia) plates and incubated at 25°C in the dark. For further purification, single colonies were picked, restreaked onto new agar plates and incubated as before. Cells were subsequently inoculated into marine broth and incubated at 28°C in a shaker incubator at 180 rpm. 2 mL of three bacterial cultures at an OD_{600} of 1 were centrifuged at 5000 rpm for 10 minutes to pellet the cells. Genomic DNA was extracted with the EZNA Bacterial DNA kit (Omega Bio-Tek) following the manufacturer's instructions and quantified on the Qubit 3.0 fluorometer using the DNA high sensitivity assay kit (Thermo-Fisher Scientific). Genomes of the bacterial isolates were sequenced using Illumina

MiSeq and PacBio platforms at either Apical Scientific (Selangor, Malaysia) or Novogene Bioinformatics Technology Co., Ltd (Beijing, China). PacBio reads were assembled into contigs using Canu v1.7 (38) after trimming and filtering. Raw reads were further mapped to the primary assemblies to identify and correct errors with BLASR v5.3 (39) and Arrow v2.2.1 (SMRT Link v7.0, www.pacb.com). Illumina paired-end 150 bp reads were trimmed using BBDuk and aligned with BBmap (https://sourceforge.net/projects/bbmap/) for further polishing of the PacBio assemblies. Resulting datasets were used as input to Pilon (40) for error correction and genome assembly improvement. The final consensus reference genomes *Sulfitobacter pseudonitzschiae* F5, *Phaeobacter* sp. F10 (Rhodobacteraceae) and *Alteromonas macleodii* F12 (Alteromonadaceae) were annotated on Prokka v1.12 and checked for completeness using BUSCO v3. Metagenomic read recruitment using the reference genomes was done using Bowtie2 v2.3.3. Average nucleotide identities shared between the isolates and MAGs of the same families were calculated using the enveomics toolbox (http://enve-omics.ce.gatech.edu/ani/).

Phylogenomics. To investigate the phylogenetic placement for roseobacter MAGs and the two isolated strains *S. pseudonitzschiae* F5 and *Phaeobacter* sp. F10, 43 complete genomes from the Rhodobacteraceae family and *Agrobacterium tumefaciens* Ach5, used as an outgroup, were downloaded from NCBI (Table S6). To investigate the phylogenetic placement for Alteromonadaceae MAGs and the isolated strain *A. macleodii* F12, 20 complete genomes from the Alteromonadaceae family and *Pseudomonas syringae* CC1557, used as an outgroup, were downloaded from NCBI (Table S7). First, bcgTree (41) was used to concatenate sequences of 107 single-copy core genes, located by HMMER v3.1b2 (42). MUSCLE v3.8.31 (43) and Gblocks 0.91b (44) were used to create and refine a multiple sequence alignment, respectively. The ETE3 package (45) was implemented on the final alignment using RAxML (46) with a JTT+GAMMA substitution model and 1000 bootstraps to generate the phylogenomic trees.

Bacterial growth assays. To assess the effects of diatom metabolites on the growth of bacteria, a representative of Rhodobacteraceae strains (*S. pseudonitzschiae* F5) and *A. macleodii* F12 were tested for growth on citrulline, norvaline, azelaic acid, leucine, threonine, hippurate, carnosine, 3-quinolinecarboxylic acid, salicylic acid, suberic acid, phenylacetic acid, 4-hydroxybenzaldehyde, 1-methylhistidine, 3-phosphoglyceric acid and phenyl acetate. Stock solutions of the assay compounds were prepared by dissolving each into Milli-Q water and subsequently filter-sterilizing through 0.2-µm membrane Nalgene syringe filters (Thermo Scientific, NY, USA). Liquid cultures were grown from single colonies in marine broth until an OD_{600} ~0.3 was reached. One milliliter aliquots of each culture were then centrifuged at 15,000 rpm (Eppendorf Centrifuge 5424) for 1 minute and the pellets were resuspended in 1 mL 10% marine broth diluted with sterile seawater. 5 µL of this bacterial stock were subsequently used to inoculate triplicate tubes containing 100 µM of each molecule in sterile 10% marine broth at a ratio of 1:1000. Growth in 10% marine broth without adding metabolites served as negative control. Absorbance at 600 nm of all cultures was measured every 24 hours from 100 µL aliquots dispensed into 96-well flat-bottom plates using an Epoch microplate spectrophotometer (BioTek Instruments Inc. Winooski, VT, USA). Sterile 10% marine broth was used as blank to correct for background media absorbance. Absorbance readings were normalized against the highest value for each bacterial isolate over the assay period. Significant differences in growth were determined by Student's *t*-test ($p < 0.05$).

Bacterial motility assay. Semisolid (0.25% w/v) marine broth agar plates supplemented with a final concentration of 2 µM rosmarinic acid were used to assess its effect on the motility of bacterial strains *S. pseudonitzschiae* F5, *Phaeobacter* sp. F10, and *A. macleodii* F12. Each strain was incubated in marine broth overnight then gently inoculated, using a sterilized toothpick, at the center of the agar surface. Triplicate plates were incubated at 26°C for 3 days, after which the proportion

of motility area was measured using the ImageJ software (http://rsb.info.nih.gov/ij/) by calculating the area of bacterial diffusion. Significant differences between control plates and rosmarinic acid-treated plates were determined by Student's *t*-test ($p < 0.05$).

Fluorescence microscopy. Axenic *A. glacialis* A3 cultures with an initial cell density of ~4,000 cells/mL in the mid-exponential phase were inoculated with cultures of strains *S. pseudonitzschiae* F5, *Phaeobacter* sp. F10, and *A. macleodii* F12 at a cell density of ~1×10^4 cells/mL grown overnight in marine broth at 26°C after centrifugation at 4000 rpm for 10 minutes followed by washing twice with sterile *f/2* medium. One mL co-cultures of *A. glacialis* A3 with strains *S. pseudonitzschiae* F5, *Phaeobacter* sp. F10, and *A. macleodii* F12 in mid-exponential phase were gently filtered onto 3-µm 25 mm polycarbonate membrane filters (Whatman). 8 µL Moviol-SYBR Green I (Thermo Fisher Scientific, MA) mixture was used to stain cells as described previously (5) and 1 mL Alcian blue in 0.06% glacial acetic acid (pH 2.5) was used to stain transparent exopolymeric particles for 10 minutes at room temperature. Samples were visualized on an epifluorescent microscope (Leica DMI6000 B, Germany) using L5 and Y5 fluorescence filter sets.

AzeR global distribution and homology analysis. The amino acid sequence of an azelaic acid transcriptional regulator, AzeR, from Bez *et al* (47) was used to search for potential homologs in the bacterial isolates *S. pseudonitzschiae* F5, *Phaeobacter* sp. F10, and *A. macleodii* F12 on BLASTX (e-value threshold of 1e-05). The resulting hits from the consortium isolates and the AzeR sequence were then used to generate a hidden Markov model (HMM) profile on HMMER v3.1b2 with hmmbuild. The hmm profile (Dataset S4) was queried against the *Tara* Oceans Microbiome Reference Gene Catalog version 1 on the Ocean Gene Atlas (http://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/) webserver (48) with an e-value threshold of 1e-50 and a bitscore threshold of 150. Geographical distributions and taxonomic abundances of homologs found in surface and deep chlorophyll maximum samples across all size fractions (0-3 µm) were visualized as donut plots across a world map. The same hmm profile was then queried against the Pfam database (49) with an e-value threshold of 1e-100, resulting in 1,621 hits. Duplicate hits, hits with <200 amino acids, and hits with no taxonomic classification were discarded. The remaining sequences were clustered on USEARCH (50) with an identity threshold of 90%. The resulting 1,043 sequences, in addition to the ones used to build the hmm profile, were aligned on MUSCLE v3.8.31 and the alignment trimmed using trimAl v1.267 (51) on "gappyout" mode. FastTree v2.1.10 (52) was used to infer phylogeny and the unrooted tree was visualized on the Interactive Tree of Life (iTOL) tool v5 (53).

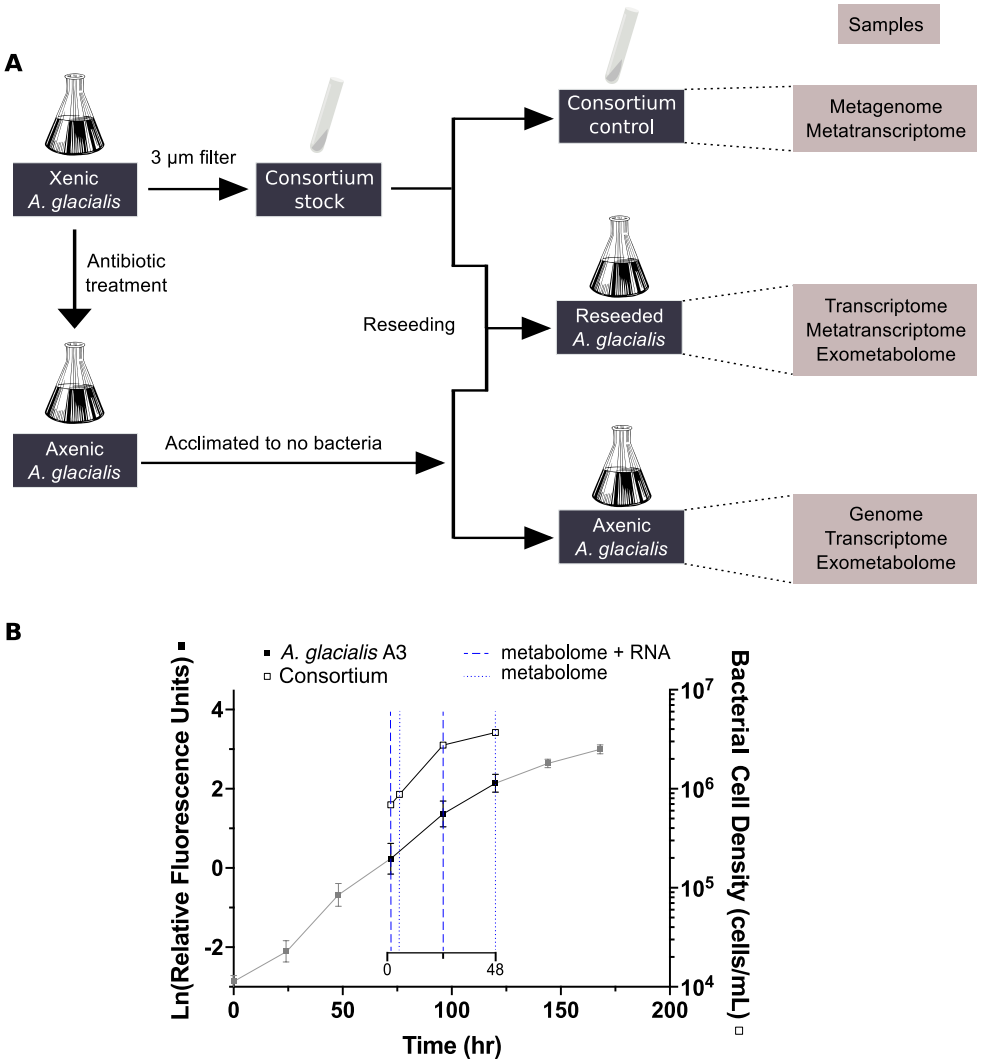

Figure S1. Experimental schematic and growth of the diatom and bacterial consortium during reseeding. (**A**) Scheme of the reseeding experiment described in the *Supplementary Methods*. Briefly, xenic *A. glacialis* A3 culture was made axenic using antibiotics and the resulting axenic culture was acclimated to absence of bacteria and subsequently used for genome sequencing. To reseed this axenic culture with bacteria, xenic *A. glacialis* A3 cultures were used to remove diatom cells and obtain a consortium stock, a portion of which was used to obtain a consortium metagenome. At the beginning of the reseeding experiment, the consortium stock was added to either axenic *A. glacialis* A3 or to sterile media (for the consortium RNA negative control) as described in the Methods. A second axenic *A. glacialis* A3 culture served as diatom control. RNA and exometabolomes were collected at different time points from each set of samples. (**B**) Growth of *A. glacialis* A3 and the microbial consortium. Because all cultures were grown in seawater-based *f/2* media that does not support significant heterotrophic growth, bacterial growth after reintroducing the microbial consortium to the diatom indicated uptake of diatom-excreted DOM. Closed squares represent diatom *in vivo* chlorophyll *a* fluorescence while open squares represent bacterial cell density. Grey points on the *A. glacialis* A3 growth curve denote time points before and after sampling. The secondary x-axis indicates the beginning of reseeding of the consortium (t=0). Dashed lines indicate time points at which RNA and metabolome samples were collected (0.5, 24 hours) while dotted lines indicate time points at which only metabolome samples were collected (4, 48 hours). Error bars represent standard deviation (SD) of triplicate cultures.

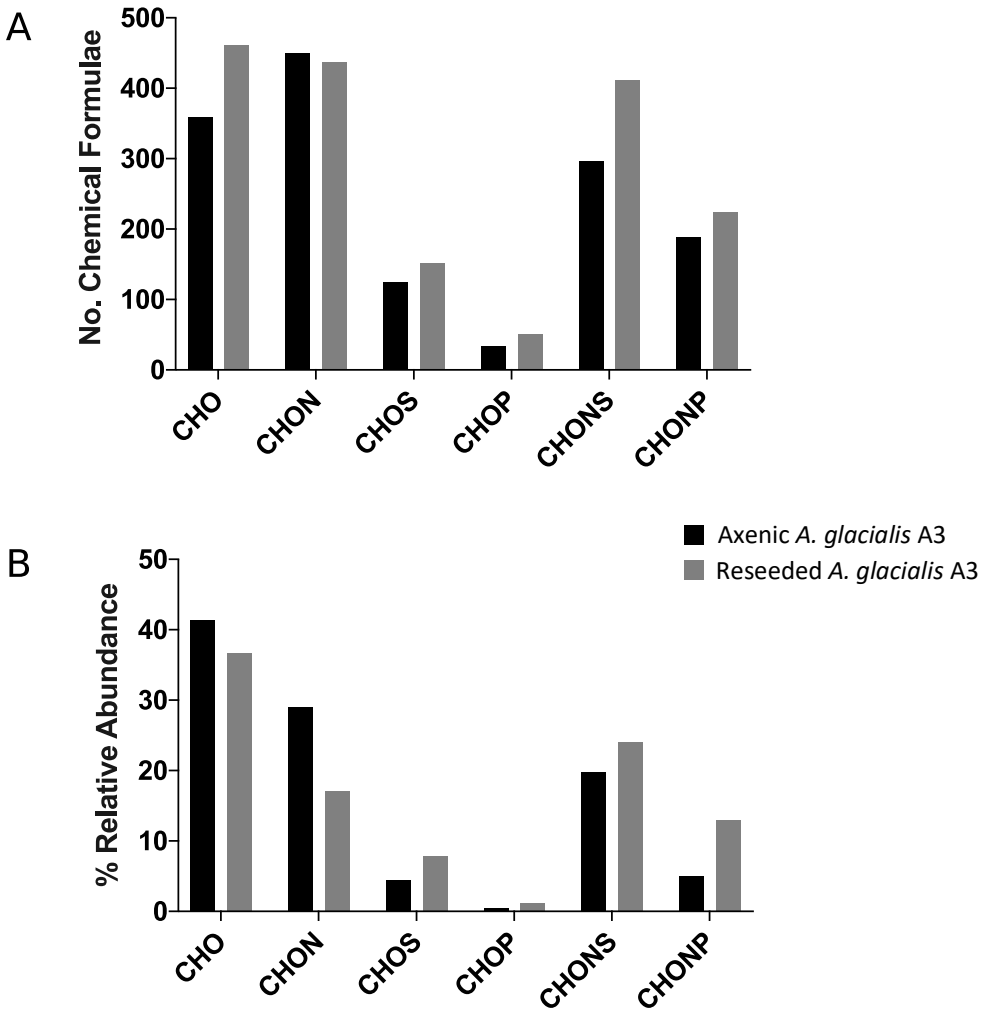

Figure S2. Diversity and abundance of chemical formulae of SPE-extracted DOM in the exometabolome. Exometabolomes were analyzed on a FT-ICR-MS as described in the *Supplementary Methods*. **(A)** Number of SPE-extracted chemical formulae from the axenic diatom and reseeded cultures 24 hours after reseeding. (**B)** Relative abundance of chemical formulae in the axenic diatom and reseeded cultures. Chemical formulae calculation was performed with an error threshold of 0.5 ppm from the exact mass for each chemical formula and isotopic fine structure. Chemical formulae were only included in the analysis if 100% of theoretical isotope peaks matched the isotopic fine structure of each formula. Biological samples were pooled to acquire sufficient signal for analysis.

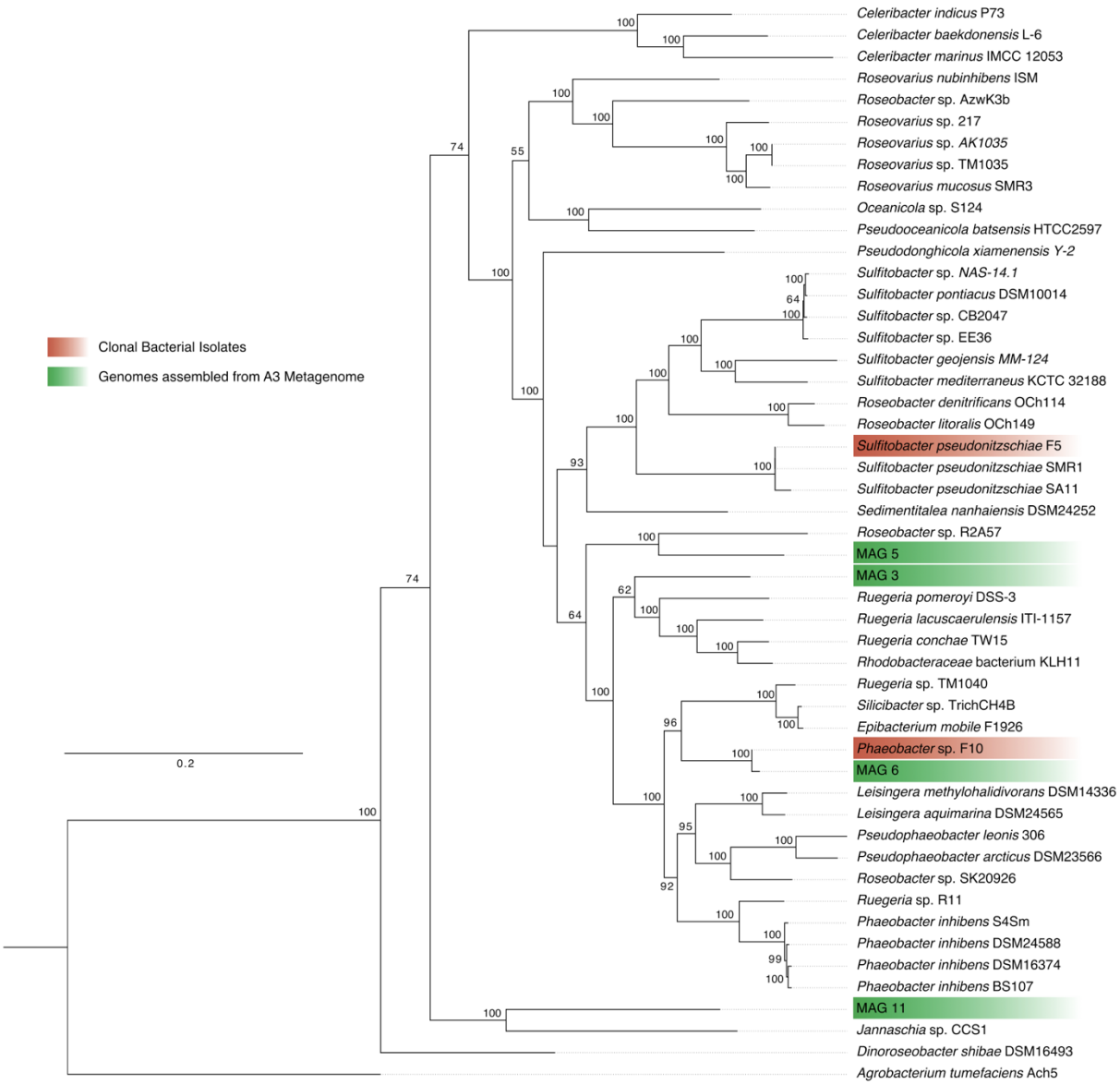

Figure S3. Metagenomically-assembled genomes (MAGs) and isolated bacterial strains from the consortium belonging to roseobacters. Maximum-likelihood phylogeny of whole-genome sequences from 50 bacterial strains comprising 43 species from the Rhodobacteraceae family, four MAGs, two isolates, and one outgroup species (*Agrobacterium tumefaciens* Ach5). Numbers adjacent to branches represent node support calculated with 1,000 bootstraps. Accession numbers of all strains are listed in Table S6.

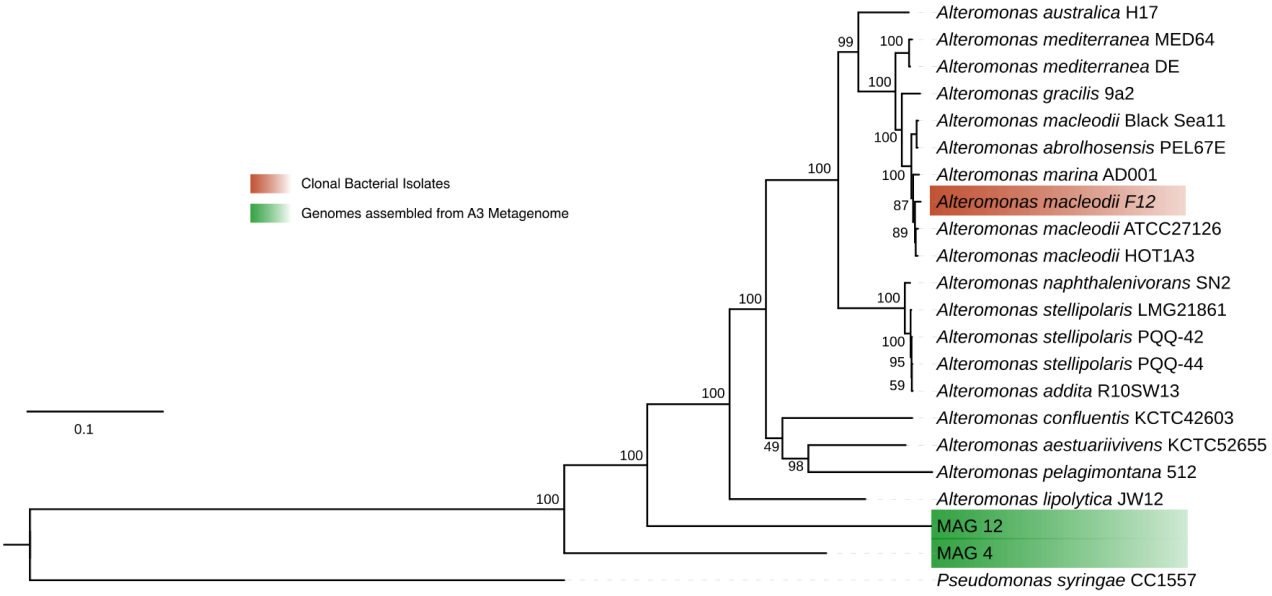

Figure S4. MAGs and isolated bacterial strains from the consortium belonging to Alteromonadaceae. Maximum-likelihood phylogeny of whole-genome sequences from 22 bacterial strains comprising 18 species from the Alteromonadaceae family, two MAGs, one isolate, and one outgroup species (*Pseudomonas syringae* CC1557). Numbers adjacent to branches represent node supports calculated with 1,000 bootstraps. Accession numbers of all strains are listed in Table S7.

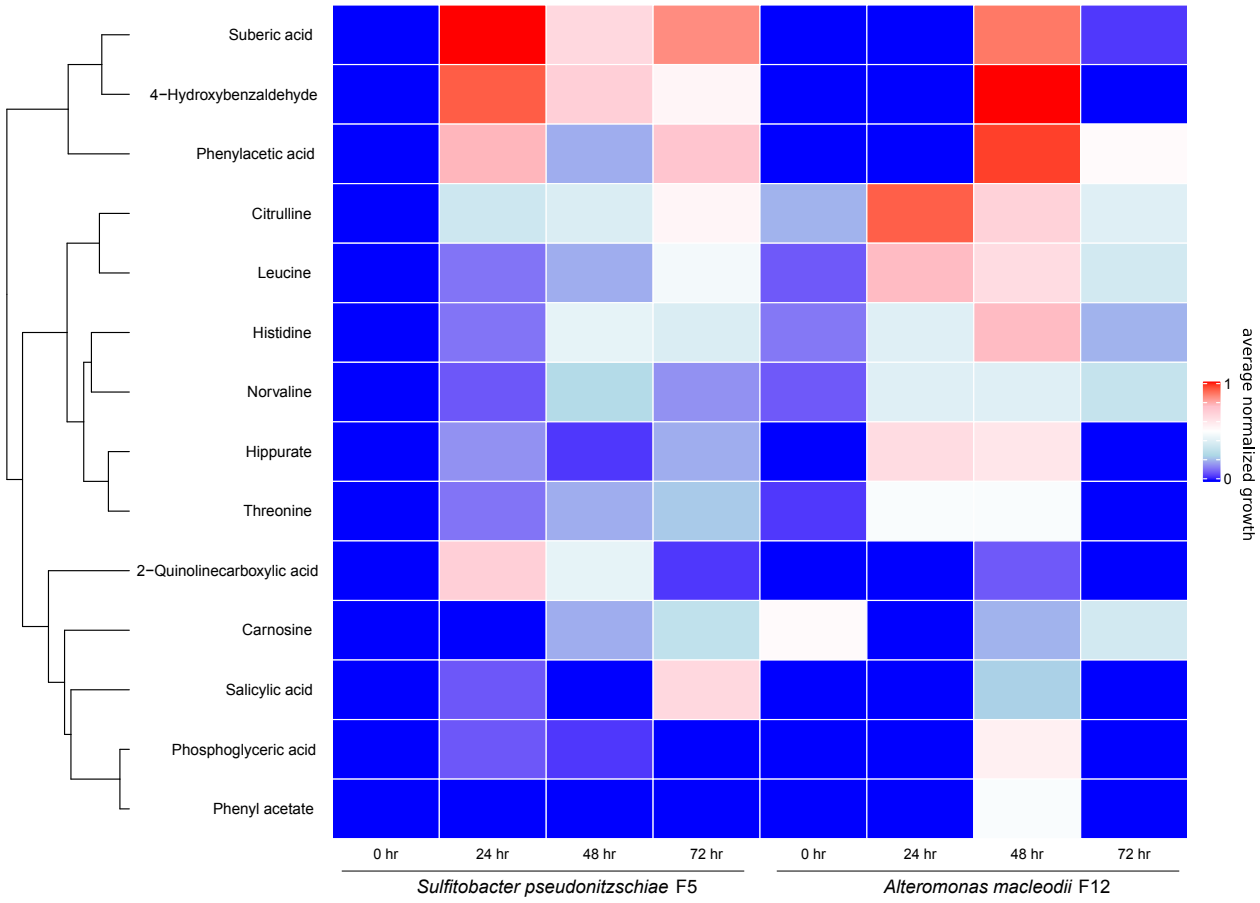

Figure S5. A subset of confirmed metabolites promotes the growth of strains *S. pseudonitzschiae* F5 and *A. macleodii* F12 isolated from the *A. glacialis* A3 consortium. Cell density (OD_{600}) of *S. pseudonitzschiae* F5 and *A. macleodii* F12 grown on a subset of confirmed metabolites from the exometabolome. Bacteria were grown in 10% marine broth supplemented with 100 µM final concentration of each molecule. Colors represent average growth from biological triplicates normalized to growth on 10% MB control.

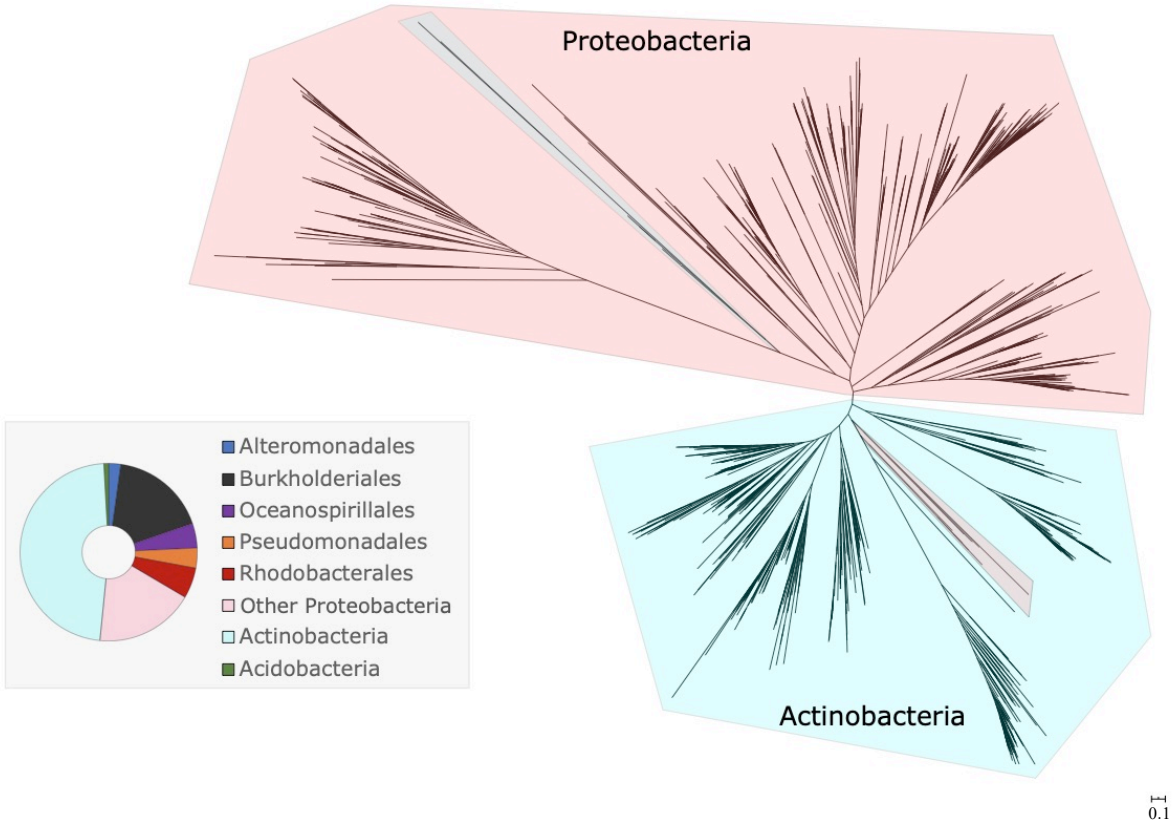

Figure S6. Bacterial response to azelaic acid is restricted to a handful of taxa. Maximum-likelihood tree of the azelaic acid transcriptional regulator, AzeR, from 1,043 protein sequences shows that response to azelaic acid in the Proteobacteria phylum is restricted to mostly five taxa. The donut plot depicts the taxonomy of all homologs.

Table S1. Summary of the assembly of metagenomically-assembled genomes (MAGs) recovered from the microbial consortium shotgun metagenome and their closest reference genome from the NCBI Reference Sequence Database according to amino acid identity (AAI).

MAG	Recruitment in bacterial consortium metagenome (%)	Total length (bp)	Contigs	Predicted genes	N50	GC content (%)	Completeness (%)	Closest genome in RefSeq	AAI (%)
3	0.26	4761801	26	4515	499224	58.70	99.49	*Phaeobacter gallaeciensis*	67.38
4	3.99	4662739	66	4310	109594	43.57	91.67	*Alteromonas australica*	57.7
5	0.44	3964870	16	3851	483383	57.71	99.47	*Phaeobacter gallaeciensis*	63.58
6	9.35	3933228	60	3850	150332	60.01	98.58	*Phaeobacter gallaeciensis*	86.2
8	2.07	3505326	20	3269	267358	44.08	99.14	*Thalassolituus oleivorans* MIL 1	51.93
9	34.18	3330214	59	3031	78673	34.53	98.49	*Polaribacter reichenbachii*	64.59
10	8.31	2639318	3	2602	1552570	57.02	99.2	*Altererythrobacter ishigakiensis*	96.99
11	1.46	2548972	12	2529	306111	56.15	98.79	*Roseibacterium elongatum* DSM 19469	62.11
12	1.03	2510135	16	2343	188996	43.72	96.03	*Alteromonas australica*	60.34
13	14.64	2198323	12	2271	368442	48.93	94.88	*Halomonas aestuarii*	48.42

Table S2. Average % mRNA reads mapped to the MAGs relative to total mRNA reads after quality control and % DE genes for MAGs relative to total number of genes in each MAG.

	Consortium only at 0.5 hours	Reseeded consortium at 0.5 hours		Reseeded consortium at 24 hours	
MAG	average % mRNA mapped	average % mRNA mapped	DE genes (%)	average % mRNA mapped	DE genes (%)
3	0.85	1.05	14.05	0.51	1.08
4	6.26	0.58	0.16	0.69	1.61
5	0.37	0.56	11.76	0.58	0
6	9.63	30.13	4.97	24.92	5.84
8	0.71	1.74	0.09	6.08	0
9	42.97	3.86	2.91	3.12	7.54
10	7.81	16.99	4.19	4.02	6.07
11	0.53	2.41	6.73	0.69	0
12	1.56	0.56	2.39	1.15	0.04
13	13.84	2.73	2.38	0.85	3.40

Table S3. List of selected expressed genes in *A. glacialis* A3, including genes depicted in Fig. 3. Genes with a false discovery rate (FDR) adjusted *p*-value < 0.1 were considered to be differentially expressed. The values correspond to the log_2-fold change at the two timepoints in response to reseeding relative to axenic controls. Blank cells indicate no differential expression.

Process/Pathway	Annotation	Gene ID	Log_2-Fold Change	
			0.5 hr	24 hr
Leucine Biosynthesis	2-isopropylmalate synthase	MSTRG.7187	5.2	5.2
		MSTRG.8032	1.9	-
	3-isopropylmalate dehydratase	MSTRG.13271	-	-
	3-isopropylmalate dehydrogenase	MSTRG.11273	-	-
	Leucine or branched chain amino acid transaminase	MSTRG.9291	-	-
Threonine Biosynthesis	Aspartate kinase/aspartokinase	MSTRG.14880	-	-
	Aspartate-semialdehyde dehydrogenase	MSTRG.7521	2.6	-
	Homoserine dehydrogenase	MSTRG.13994	2.7	-
	Homoserine kinase	MSTRG.13479	-	-
	Threonine synthase	MSTRG.15183	-	-
Methionine Biosynthesis	Cystathionine gamma-lyase	MSTRG.14985	-	5.0
		MSTRG.5022	-	-
	Cystathionine-β-lyase	MSTRG.6803	-	-
		MSTRG.630	-	-
	Methionine synthase	MSTRG.10935	-	-
Shikimate pathway for biosynthesis of aromatic amino acids	DAHP_synth_2	MSTRG.1214	2.0	-
	3-dehydroquinate synthase	MSTRG.12529	-	-
		MSTRG.12530	-	-
		MSTRG.8477	-	-
	Shikimate dehydrogenase	MSTRG.4610	4.1	-
	Shikimate kinase	MSTRG.14973	-	-
	3-phosphoshikimate 1-carboxyvinyltransferase	MSTRG.810	-	-
	Chorismate synthase	MSTRG.11056	-	-
	Chorismate mutase/ 3-deoxy-7-phosphoheptulonate synthase	MSTRG.1214	2.0	-
	Aspartate-prephenate aminotransferase	MSTRG.3079	-	-

Tyrosine and Phenylalanine Biosynthesis	Prephenate dehydrogenase/ arogenate/prephenate dehydratase	MSTRG.4156	2.0	-
Tryptophan Biosynthesis	Anthranilate synthase	MSTRG.2535	-	-
	Anthranilate phosphoribosyltransferase	MSTRG.10304	-	-
	Phosphoribosylanthranilate isomerase	MSTRG.12883	-	-
	Indole-3-glycerolphosphate synthase	MSTRG.406	-	-4.0
	Tryptophan synthase	MSTRG.6148	-	-
Calvin cycle	Ribulose bisphosphate carboxylase (RuBisCO)	MSTRG.8319 – large chain	-	-
		MSTRG.8223 – small chain	-	-
	Phosphoglycerate kinase (PGK)	MSTRG.1980	-	-
		MSTRG.2076	-5.3	-
		MSTRG.4100	-	-
		MSTRG.4704 – (Chloroplastic)	-1.6	-2.9
		MSTRG.4930	-	-
		MSTRG.4931	-	-
	Glyceraldehyde-3-phosphate dehydrogenase	MSTRG.11073	-	-
		MSTRG.803	-	-
		MSTRG.805	-	-
		MSTRG.9841	-	-
	Triose phosphate isomerase	MSTRG.12260	-	-
		MSTRG.3195	-	-
		MSTRG.5009	-1.6	-3.3
		MSTRG.6633 – (Cytosolic)	-	-
		MSTRG.6634 – (Cytosolic)	-	-
		MSTRG.804 – (Cytosolic)	-	-
	Fructose-bisphosphate aldolase	MSTRG.11774	-	-
		MSTRG.12540	-	-
		MSTRG.3156	-	-2.9
		MSTRG.6894	-	-

		MSTRG.9714	-	-
	Fructose-1,6-bisphosphatase	MSTRG.13012	-	-
		MSTRG.13662	-	-3.3
		MSTRG.3685	-	-1.9
		MSTRG.4408	-	5.2
		MSTRG.4695	-	-
	Transketolase	MSTRG.11092	6.3	-
		MSTRG.12402	3.0	1.9
		MSTRG.6202	-	-
		MSTRG.6203	-1.5	-3.7
	Sedoheptulose-1,7-bisphosphatase	MSTRG.2255 – (Chloroplastic)	-3.6	-3.8
	Phosphopentose isomerase	MSTRG.14854 – (Chloroplastic)	-	-2.0
		MSTRG.5171 – (Chloroplastic)	-	-
	Phosphoribulokinase	MSTRG.7086 – (Chloroplastic)	-2.2	-2.9
		MSTRG.7087 – (Chloroplastic)	-	-
Glycolysis	Enolase	MSTRG.11006	2.4	-
		MSTRG.11924	1.6	-
	Pyruvate kinase	MSTRG.2557	-	5.3
		MSTRG.6652	3.8	5.2
		MSTRG.1740	-3.7	-6.7
TCA Cycle	Glutamate dehydrogenase	MSTRG.6004	2.2	-
		MSTRG.115	-	5.4
	Citrate synthase	MSTRG.5756	1.8	-
	Aconitase	MSTRG.3015	2.2	-
	Isocitrate dehydrogenase	MSTRG.10398	-	-
		MSTRG.2824	-	-
	α-Ketoglutarate dehydrogenase	MSTRG.7999	-	-
	Succinyl-CoA synthetase	MSTRG.9255	-	-
	Succinate dehydrogenase	MSTRG.12076	-	-
		MSTRG.15148	-	-

		MSTRG.2893 (assembly factor 2)	-	-
		MSTRG.4574 (assembly factor 2)	-	-
		MSTRG.5051	-	-
		MSTRG.6152	-	-
	Fumarase	MSTRG.150	-	-
		MSTRG.13565	-	-
	Malate dehydrogenase	MSTRG.5650	-	-
		MSTRG.8836	-	-
Urea Cycle	Agmatinase (AgM)	MSTRG.13929	-	-
	Arginase (Arg)	MSTRG.12510	-	-
		MSTRG.9765	-	-
	Argininosuccinate lyase (AsL)	MSTRG.8783	2.4	-
	Argininosuccinate synthase (AsuS)	MSTRG.9347	-	-
		MSTRG.9372	-	-
	Carbamoyl-phosphate synthase (CPS)	MSTRG.8837	-	-
		MSTRG.8838	-	-
	Ornithine decarboxylase (OdC)	MSTRG.10176	-	-
		MSTRG.11618	-	-
		MSTRG. 7737	-	-
	Ornithine cyclodeaminase (OCD)	MSTRG.11212	-	-
		MSTRG.5004	-	-
		MSTRG. 8409	-	-
	Ornithine transcarboxylase/ ornithine carbamoyltransferase (OTC)	MSTRG. 10929	-	-
	Urease (Ure)	MSTRG.5242	-	-
		MSTRG.14527	-	4.1
		MSTRG.6749	-	-
Nitrate Assimilation	Nitrate transporter (NIT1)	MSTRG.7864	-	-
	Nitrate reductase (NR)	MSTRG.13053	-	-
		MSTRG.1608	-	-
	Plastid Nitrite transporter (NaR1)	MSTRG.8511	-	-

	Ferredoxin nitrite reductase (NiR)	MSTRG.10808	-2.7	-2.8
	NADPH nitrite reductase (NasB)	MSTRG.13868	-	-1.7
		MSTRG.717	-	-
Membrane Transporters	Transmembrane amino acid transporter protein	MSTRG.12321	-	-2.9
	Putative sodium-coupled neutral amino acid transporter	MSTRG.12635	4.8	-
	Putative sodium-coupled neutral amino acid transporter (K14997)	MSTRG.12323	-	-
	Amino acid/polyamine transporter	MSTRG.13113	-	-
	Amino acid/polyamine transporter	MSTRG.3541	-	-
	Sodium-coupled neutral amino acid transporter	MSTRG.4924	-	-
	Tryptophan/ tyrosine permease family	MSTRG.6793	-	-
	Xylulose 5-phosphate/phosphate translocator (Chloroplastic)	MSTRG.6981	-7.0	-5.7
	Phosphoenolpyruvate/phosphate translocator 2 (Chloroplastic)	MSTRG.8004	-4.0	-5.1
Polyamine Related	Spermidine synthase	MSTRG.8285	3.8	-
	Spermidine/putrescine-binding periplasmic protein	MSTRG.13158	2.7	-
	N-carbamoylputrescine amidase	MSTRG.7725	-	-
Fatty Acid biosynthesis	Acetyl-CoA carboxylase	MSTRG.2420	-2.7	-2.6
		MSTRG.8023	-	-
		MSTRG.8024	-	-1.8
	S-malonyltransferase (fabD)	MSTRG.179	-	-
		MSTRG.8641	-	-
	3-oxoacyl-[acyl-carrier-protein] synthase II (fabF)	MSTRG.3172	-	-
		MSTRG.7769	-	-
	3-oxoacyl-[acyl-carrier-protein] synthase III (fabH)	MSTRG.8709	-	-3.3
	3-oxoacyl-[acyl-carrier protein] reductase	MSTRG.12209	-5.4	-6.1
		MSTRG.13403	-	-
	Enoyl-[acyl-carrier protein] reductase I (fabI)	MSTRG.11059	-	-3.5
	Long-chain acyl-CoA synthetase	MSTRG.11424	-	-
		MSTRG.14797	-	-
	Acyl-[acyl-carrier-protein] desaturase	MSTRG.8478	-1.9	-1.9
	Long-chain acyl-CoA synthetase	MSTRG.11424	-	-

Fatty Acid Degradation		MSTRG.14797	-	-
	Acyl-CoA oxidase	MSTRG.8635	-	-
	Butyryl-CoA dehydrogenase	MSTRG.2258	-	-
	Glutaryl-CoA dehydrogenase	MSTRG.11308	-	-
	Enoyl-CoA hydratase	MSTRG.7723	-	-
		MSTRG.4057	-	5.3
	3-hydroxyacyl-CoA dehydrogenase	MSTRG.14157	-	-
		MSTRG.7069	-	-
	Acetyl-CoA acyltransferase/ 3-ketoacyl-CoA thiolase A	MSTRG.13287	-	-
		MSTRG.8254	-	-
	Acetyl-CoA C-acetyltransferase	MSTRG.9055	-	-
	Delta-3-Delta-2-enoyl-CoA isomerase	MSTRG.11980	2.9	3.0
	Long-chain-fatty-acid--[acyl-carrier-protein] ligase	MSTRG.10490	3.7	-
		MSTRG.10491	3.3	6.2
		MSTRG.9802	-	1.7

Table S4. List of confirmed metabolites in Fig. 2*D*. Metabolites were confirmed by comparing retention time, accurate mass, isotopic pattern and fragmentation pattern of each metabolite to a library of in-house chemicals (Mass Spectrometry Metabolite Library of Standards, IROA Technologies, US). Additional molecules were confirmed using the Bruker MetaboBASE Plant Library and MetaboBASE Personal Library 2.0 (BrukerDaltonik, Germany). We were not able to confirm the annotations of molecules found only in reseeded samples. Analysis was done using Metaboscape v4.0 (BrukerDaltonik, Germany). Primary Database refers to the in-house chemical library. M= monoisotopic mass, RT= retention time, ΔmDA= mass difference in milliDalton, IP= isotopic pattern, Δppm= mass difference in parts per million.

M (Da)	RT (min)	ΔmDa	IP	MS/MS Score	Chemical Formula	Adduct	Name	Δppm	Database
103.0622	19.94	0.76	80	980.1	C4H9NO2	[4M+H]+	4-Aminobutanoate	7.4	Primary Database
117.078	0.45	1	39	999.9	C5H11NO2	[M+H+H]2+	Norvaline	8.5	Primary Database
119.0578	12.42	0.22	46	760.9	C3H7NO3	[M+H]+	Threonine	1.8	Primary Database
122.0369	10.25	0.06	45	773.6	C7H6O2	[M+H]+	4-Hydroxybenzaldehyde	0.5	Primary Database
131.0946	13.51	0.39	10	761.6	C6H13NO2	[M+H]+	Leucine	3	Primary Database
136.0511	19.91	0.9	62	967.5	C8H8O2	[M+H]+	Phenylacetic acid	6.6	Primary Database
136.0512	11.34	0.8	52	735.3	C8H8O2	[M-H]-	Phenyl acetate	5.9	Primary Database
138.031	4.26	0.97	36	922.2	C7H7NO2	[M+H]+	Salicylic acid	7	Primary Database
145.0739	8.71	0.11	23	759.1	C6H11NO3	[M+H]+	4-Acetamidobutanoic acid	0.8	Primary Database
146.0573	19.95	0.68	51	993.1	C6H10O4	[M+H]+	3-Methyglutaric acid	4.7	Primary Database
152.0464	18.8	0.6	59	846.5	C8H8O3	[M-H]-	4-Hydroxy-phenylacetate	3.9	Primary Database
161.0695	18.7	-0.52	52	860.1	C6H11NO4	[M+H]+	α-Aminoadipate	-3.2	Primary Database
167.0577	0.38	0.33	41	716.4	C8H9NO3	[M-H]-	4-Hydroxy-phenylglycine	2	Primary Database
169.0843	4.47	0.73	46	949.1	C7H11N3O2	[M+H]+	1-Methylhistidine	4.3	Primary Database
173.0482	1.34	-0.16	41	845	C10H7NO2	[M+H]+	2-Quinolinecarboxylic acid	-0.9	Primary Database
174.088	0.38	1	50	612	C8H14O4	[M+H]+	Suberic acid	5.7	Primary Database
175.0958	6.8	0.21	70	897.9	C6H13N3O3	[M+H]+	Citrulline	1.2	Primary Database
179.0571	16.74	0.95	52	989.3	C9H9NO3	[M+H]+	Hippurate	5.3	Primary Database
185.9925	9.33	0.49	56	882.7	C3H7O7P	[M+H]+	3-Phosphoglyceric acid	2.6	Primary Database
188.1047	6.87	0.26	70	872.7	C9H16O4	[M+H]+	Azelaic acid	1.4	Primary Database
221.0898	8.04	0.18	5.7	763.4	C8H15NO6	[M+H]+	N-acetyl-galactosamine	0.8	Primary Database
226.1063	11.44	0.67	33	998	C9H14N4O3	[M-H]-	Carnosine	3	Primary Database
282.2556	17.01	0.4	38.7	978	C18H34O2	[M-H]-	Oleic acid	1.4	Bruker MetaboBASE Plant Library

284.2709	18.17	0.61	2.9	997.6	C18H36O2	[M-H]-	Stearic acid	2.2	Bruker MetaboBASE Plant Library
301.298	10.46	0	19	824.1	C18H39NO2	[M+H]+	Sphinganine	0	Primary Database
342.2278	15.5	0.26	40	730.5	C11H14O2	[M+H]+	N-Tetradecanoylaspartic acid	0.8	Bruker MetaboBASE Personal Library 2.0
360.0843	18.77	0.74	26	928.6	C18H16O8	[M+H]+	Rosmarinic acid	2.1	Primary Database
541.3372	16.6	0.79	27	992.3	C25H52NO9P	[M-H]-	C_{16}-hydroxy-glycerophosphocholine	1.5	Bruker MetaboBASE Plant Library

Table S5. List of selected expressed genes by roseobacter metagenomically-assembled genomes MAG3, MAG5 and MAG6, including genes depicted in Fig. 3. Genes with a false discovery rate (FDR) adjusted *p*-value < 0.1 were considered to be differentially expressed. The values correspond to the log_2-fold change at the two timepoints in response to reseeding relative to consortium control. Blank cells indicate no differential expression.

MAG	Gene ID	Annotation	Log_2-Fold Change	
			0.5 hr	24 hr
MAG3	k99_165394_20	2-oxoisovalerate_dehydrogenase_subunit_alpha	-	-
	k99_165394_19	2-oxoisovalerate_dehydrogenase_subunit_beta	-	-
	k99_56693_167	2-oxoisovalerate_dehydrogenase_subunit_beta	-	-
	k99_258460_5	Alpha-ketoglutarate-dependent_taurine_dioxygenase	-	-
	k99_157987_224	Argininosuccinate_lyase	5.2	5.4
	k99_165394_71	Bicyclomycin_resistance_protein	5	-
	k99_152108_202	Biofilm_growth-associated_repressor	6.2	-
	k99_111971_14	Branched-chain-amino-acid_aminotransferase	5.6	-
	k99_152108_131	C4-dicarboxylate_TRAP_transporter_large_permease_protein_DctM	6.6	-
	k99_87748_267	C4-dicarboxylate_TRAP_transporter_large_permease_protein_DctM	6.3	-
	k99_115381_320	C4-dicarboxylate_TRAP_transporter_large_permease_protein_DctM	6.0	-
	k99_226572_386	C4-dicarboxylate_TRAP_transporter_large_permease_protein_DctM	5.6	-
	k99_165394_264	Chloramphenicol_acetyltransferase	5.2	-
	k99_87748_232	Creatinase	6.6	-
	k99_87748_404	Dimethlysulfonioproprionate_lyase_DddQ	6.6	-
	k99_226572_438	Flagellar_biosynthetic_protein_FlhB	6.6	-
	k99_256486_107	Glutamate/aspartate_import_permease_protein_GltK	6.2	-
	k99_115381_744	Glutamine_transport_ATP-binding_protein_GlnQ	5.4	-
	k99_226572_256	Heme-binding_protein_A	6.6	-
	k99_226572_290	High-affinity_branched-chain_amino_acid_transport_ATP-binding_protein_LivF	5.6	-
	k99_226572_17	High-affinity_branched-chain_amino_acid_transport_system_permease_protein_LivH	-	-
	k99_226572_237	High-affinity_branched-chain_amino_acid_transport_system_permease_protein_LivH	6.1	-
	k99_152108_173	High-affinity_branched-chain_amino_acid_transport_system_permease_protein_LivH	6.0	-
	k99_94583_24	High-affinity_branched-chain_amino_acid_transport_system_permease_protein_LivH	5.7	-

	k99_166876_14	Homogentisate_1,2C2-dioxygenase	6.0	6.4
	k99_166876_81	Homoserine/homoserine_lactone_efflux_protein	5.1	-
	k99_157987_68	Leucine-responsive_regulatory_protein	6.0	-
	k99_157987_182	Multidrug_export_protein_AcrF	7.5	-
	k99_152108_327	Multidrug_resistance_protein_MdtA	6.2	-
	k99_152108_446	Multidrug_resistance_protein_MdtK	7.6	-
	k99_115381_25	Periplasmic_dipeptide_transport_protein	6.0	-
	k99_87748_346	Phosphate_acetyltransferase	-	-
	k99_157987_27	Putative_aliphatic_sulfonates_transport_permease_protein_SsuC	6.2	-
	k99_115381_555	putative_dipeptidase_PepE	6.6	-
	k99_165394_144	Putative_multidrug_export_ATP-binding/permease_protein	6.7	-
	k99_258460_64	Putrescine_transport_system_permease_protein_PotH	5.4	-
	k99_115381_678	Putrescine-binding_periplasmic_protein	5.2	-
	k99_226572_141	Sarcosine_oxidase_subunit_alpha	6.9	-
	k99_115381_515	Sarcosine_oxidase_subunit_alpha	6.8	-
	k99_115381_401	Sarcosine_oxidase_subunit_alpha	6.6	-
	k99_19927_12	Sarcosine_oxidase_subunit_alpha	5.5	-
	k99_115381_517	Sarcosine_oxidase_subunit_beta	6.0	-
	k99_115381_664	Sialic_acid_TRAP_transporter_permease_protein_SiaT	6.3	-
	k99_94583_171	Sialic_acid_TRAP_transporter_permease_protein_SiaT	5.1	-
	k99_115381_666	Sialic_acid-binding_periplasmic_protein_SiaP	6.1	-
	k99_56693_271	Sorbitol_dehydrogenase	-	-
	k99_226572_418	Spermidine/putrescine_import_ATP-binding_protein_PotA	-	-
	k99_115381_675	Spermidine/putrescine_import_ATP-binding_protein_PotA	5.8	-
	k99_226572_460	Spermidine/putrescine-binding_periplasmic_protein	5.5	-
	k99_87748_422	Sulfoacetaldehyde_acetyltransferase	6.1	-
	k99_94583_81	Sulfoacetaldehyde_acetyltransferase	5.2	-
	k99_152108_210	Sulfopropanediol_3-dehydrogenase	6.8	-
	k99_87748_269	Sulfopropanediol_3-dehydrogenase	5.1	-
	k99_157987_26	Taurine_import_ATP-binding_protein_TauB	-	-

	k99_157987_24	Taurine--pyruvate_aminotransferase	-	-
	k99_87748_391	Taurine--pyruvate_aminotransferase	6.9	-
	k99_157987_25	Taurine-binding_periplasmic_protein	-	-
	k99_94583_9	Urease_subunit_alpha_1	5.9	8.4
	k99_247855_81	Xaa-Pro_dipeptidase	6.8	-
MAG5	k99_236746_15	2-oxoisovalerate_dehydrogenase_subunit_alpha	-	
	k99_236746_16	2-oxoisovalerate_dehydrogenase_subunit_beta	-	-
	k99_55525_106	2-oxoisovalerate_dehydrogenase_subunit_beta	-	-
	k99_55525_477	4-hydroxyphenylpyruvate_dioxygenase	-	-
	k99_55525_659	4-hydroxyphenylpyruvate_dioxygenase	-	-
	k99_181994_636	6'''-hydroxyparomomycin_C_oxidase	6.2	-
	k99_45374_273	Acetylornithine_deacetylase	7.3	-
	k99_45374_228	Aclacinomycin_methylesterase_RdmC	-	-
	k99_187880_412	AI-2_transport_protein_TqsA	-	-
	k99_232985_80	Argininosuccinate_lyase	6.7	-
	k99_181994_54	Autoinducer_2_sensor_kinase/phosphatase_LuxQ	6.1	-
	k99_27662_156	Autoinducer_2_sensor_kinase/phosphatase_LuxQ	5.3	-
	k99_55525_216	Bicyclomycin_resistance_protein	5.5	-
	k99_187880_435	Bicyclomycin_resistance_protein	-	-
	k99_181994_278	Branched-chain-amino-acid_aminotransferase	5.9	-
	k99_55525_568	C4-dicarboxylate_transport_sensor_protein_DctB	6.9	-
	k99_209332_6	C4-dicarboxylate_transport_sensor_protein_DctB	5.7	-
	k99_55525_233	C4-dicarboxylate_transport_sensor_protein_DctB	5.4	-
	k99_230788_7	C4-dicarboxylate_transport_transcriptional_regulatory_protein_DctD	5.2	-
	k99_45374_16	C4-dicarboxylate_TRAP_transporter_large_permease_protein_DctM	6.8	-
	k99_46830_4	Cation/acetate_symporter_ActP	4.4	-
	k99_236746_25	Chemotaxis_protein_CheW	5.4	-
	k99_181994_493	Creatinase	8.3	-
	k99_174471_78	Cypemycin_N-terminal_methyltransferase	-	-
	k99_40035_14	Dipeptidyl-peptidase_5	5.3	-

	k99_181994_127	Flagellar_biosynthesis_protein_FlhA	7.6	-
	k99_181994_129	Flagellar_biosynthetic_protein_FlhB	6.9	-
	k99_181994_55	Flagellar_P-ring_protein	-	-
	k99_181994_140	Flagellum-specific_ATP_synthase	5.3	-
	k99_181994_413	Glutamine_transport_ATP-binding_protein_GlnQ	6.1	-
	k99_232985_237	Glutamine_transport_ATP-binding_protein_GlnQ	5.5	-
	k99_181994_174	Glutamine-binding_periplasmic_protein	5.3	-
	k99_232985_278	Hemin_transport_system_permease_protein_HmuU	5.4	-
	k99_174471_8	Hemin_transport_system_permease_protein_HmuU	-	-
	k99_46830_30	Hemin_transport_system_permease_protein_HmuU	-	-
	k99_45374_386	High-affinity_branched-chain_amino_acid_transport_ATP-binding_protein_LivF	6.1	-
	k99_181994_33	High-affinity_branched-chain_amino_acid_transport_system_permease_protein_LivH	5.8	-
	k99_45374_387	High-affinity_branched-chain_amino_acid_transport_system_permease_protein_LivH	5.2	-
	k99_232985_19	High-affinity_branched-chain_amino_acid_transport_system_permease_protein_LivH	5.2	-
	k99_45374_373	Homogentisate_1,2C2-dioxygenase	6.2	-
	k99_45374_384	Leucine-2C_isoleucine-2C_valine-2C_threonine-2C_and_alanine-binding_protein	5.2	-
	k99_181994_28	Leucine-2C_isoleucine-2C_valine-2C_threonine-2C_and_alanine-binding_protein	-	-
	k99_187880_15	Low-molecular_weight_cobalt-containing_nitrile_hydratase_subunit_alpha	5.2	-
	k99_45374_378	Metallo-beta-lactamase_type_2	4.9	-
	k99_181994_520	Methyl-accepting_chemotaxis_protein_II	6.0	-
	k99_181994_779	Methyl-accepting_chemotaxis_protein_III	5.6	-
	k99_55525_831	Monocarboxylate_2-oxoacid-binding_periplasmic_protein	4.7	-
	k99_55525_603	Multidrug_resistance_protein_MdtK	5.5	-
	k99_55525_274	N-acyl_homoserine_lactonase	6.8	-
	k99_228385_23	p-hydroxyphenylacetate_3-hydroxylase-2C_reductase_component	-	-
	k99_181994_654	Periplasmic_dipeptide_transport_protein	6.2	-
	k99_27662_31	Periplasmic_dipeptide_transport_protein	6.0	-
	k99_174471_139	Periplasmic_dipeptide_transport_protein	-	-
	k99_181994_507	Phosphate_acetyltransferase	-	-
	k99_55525_55	Phosphate_acetyltransferase	-	-

	k99_55525_111	Phosphoglycerate_kinase	5.7	-
	k99_55525_111	Phosphoglycerate_kinase	5.7	-
	k99_209332_7	Phosphoglycerate_transport_regulatory_protein_PgtC	5.3	-
	k99_181994_645	Phthiocerol/phenolphthiocerol_synthesis_polyketide_synthase_type_I_PpsE	6.6	-
	k99_187880_326	putative_amino_acid_permease_YhdG	6.2	-
	k99_232985_25	putative_amino-acid_permease_protein_YxeN	5.7	-
	k99_45374_23	putative_D-2CD-dipeptide_transport_system_permease_protein_DdpC	5.2	-
	k99_187880_309	putative_dipeptidase_PepE	-	-
	k99_230788_113	putative_dipeptidase_PepE	-	-
	k99_232985_296	putative_dipeptidase_PepE	-	-
	k99_55525_525	Sarcosine_oxidase_subunit_alpha	6.3	-
	k99_55525_721	Sarcosine_oxidase_subunit_alpha	-	-
	k99_55525_524	Sarcosine_oxidase_subunit_gamma	-	-
	k99_236746_5	Sialic_acid_TRAP_transporter_large_permease_protein_SiaM	-	-
	k99_228385_20	Sialic_acid-binding_periplasmic_protein_SiaP	5.4	-
	k99_55525_621	Sodium-dependent_dicarboxylate_transporter_SdcS	6.5	-
	k99_27662_108	Sodium/glutamate_symporter	7.0	-
	k99_181994_161	Spermidine/putrescine_transport_system_permease_protein_PotB	6.2	-
	k99_187880_473	Spermidine/putrescine_transport_system_permease_protein_PotB	5.7	-
	k99_181994_509	Sulfoacetaldehyde_acetyltransferase	7.7	-
	k99_40035_70	Sulfopropanediol_3-dehydrogenase	6.9	-
	k99_232985_260	Sulfopropanediol_3-dehydrogenase	-	-
	k99_232985_212	Taurine_import_ATP-binding_protein_TauB	-	-
	k99_181994_513	Taurine_import_ATP-binding_protein_TauB	-	-
	k99_181994_511	Taurine--pyruvate_aminotransferase	5.2	-
	k99_181994_529	Taurine--pyruvate_aminotransferase	-	-
	k99_181994_512	Taurine-binding_periplasmic_protein	-	-
	k99_174471_156	Urease_subunit_alpha_1	6.5	-
	k99_209332_39	Virginiamycin_B_lyase	-	-
	k99_187880_180	Xylose_transport_system_permease_protein_XylH	5.2	-

	k99_187880_231	Aspartate_aminotransferase	5.8	-
MAG6	k99_160923_63	2-oxoisovalerate_dehydrogenase_subunit_beta	3.2	-
	k99_156299_21	Arginine_transport_ATP-binding_protein_ArtM	4.7	-
	k99_28679_10	Branched-chain-amino-acid_aminotransferase	-	-
	k99_25232_178	Chloramphenicol_acetyltransferase	4.8	-
	k99_240218_90	Dipeptide_transport_system_permease_protein_DppB	4.1	-
	k99_65518_124	Hemin-binding_periplasmic_protein_HmuT	3.3	-
	k99_18978_161	High-affinity_branched-chain_amino_acid_transport_ATP-binding_protein_LivF	4.6	-
	k99_156299_22	High-affinity_branched-chain_amino_acid_transport_ATP-binding_protein_LivF	3.6	-
	k99_25232_136	High-affinity_branched-chain_amino_acid_transport_system_permease_protein_LivH	3.9	-
	k99_156299_23	High-affinity_branched-chain_amino_acid_transport_system_permease_protein_LivH	3.2	-
	k99_18978_165	Leu/Ile/Val (Thr)-binding_protein	3.0	-
	k99_65518_111	Urease_accessory_protein_UreD	3.4	-
	k99_48975_37	Monocarboxylate_2-oxoacid-binding_periplasmic_protein	-	-2.8
	k99_48975_101	High-affinity_branched-chain_amino_acid_transport_ATP-binding_protein_LivF	-	3.5
	k99_160923_53	Branched-chain-amino-acid_aminotransferase	-	-3.7
	k99_240218_12	Iron_uptake_protein_A1	-3.2	-3.6
	k99_114588_15	Succinate_dehydrogenase_cytochrome_b556_subunit	-	-3.6
	k99_28679_52	Multidrug_resistance_protein_MexB	-	3.0

Table S6. List of accession numbers for roseobacter genomes used in the phylogenomic analysis of the roseobacter MAGs and laboratory cultured consortium strains (Fig. S3). The *Agrobacterium tumefaciens* Ach5 genome was used as an outgroup.

Bacterial strain	NCBI accession number
Celeribacter baekdonensis L-6	NZ_FNBL00000000
Celeribacter indicus P73	NZ_CP004393
Celeribacter marinus IMCC 12053	NZ_CP012023
Dinoroseobacter shibae DSM16493	CP000830
Epibacterium mobile F1926	NZ_CP015230
Jannaschia sp. CCS1	CP000264
Leisingera aquimarina DSM 24565	AXBE00000000
Leisingera methylohalidivorans DSM14336	CP006773
Oceanicola sp. S124	NZ_AFPM00000000
Phaeobacter inhibens S4Sm	LOHU01000000
Phaeobacter inhibens DSM24588	CP002972
Phaeobacter inhibens DSM16374	AXBB00000000
Phaeobacter inhibens BS107	NZ_CP031948
Phaeobacter sp. F10	*This study*
Pseudodonghicola xiamenensis Y-2	NZ_AUBS00000000
Pseudooceanicola batsensis HTCC2597	NZ_AAMO00000000
Pseudophaeobacter arcticus DSM23566	NZ_AXBF00000000
Pseudophaeobacter leonis 306	NZ_MWVJ00000000
Rhodobacteraceae bacterium KLH11	NZ_ACCW00000000
Roseobacter sp. AzwK3b	ABCR00000000
Roseobacter denitrificans OCh114	CP000362
Roseobacter litoralis OCh149	CP002623
Roseobacter sp. R2A57	-*
Roseobacter sp. SK20926	AAYC00000000
Roseovarius nubinhibens ISM	AALY00000000

Roseovarius sp. 217	AAMV00000000
Roseovarius sp. AK1035	NZ_CP030099
Roseovarius sp. TM1035	ABCL00000000
Roseovarius mucosus SMR3	NZ_CP020474
Ruegeria pomeroyi DSS-3	CP000031
Ruegeria lacuscaerulensis ITI-1157	ACNX00000000
Ruegeria conchae TW15	AEYW00000000
Ruegeria sp. TM1040	CP000375
Ruegeria sp. R11	ABXM00000000
Sedimentitalea nanhaiensis DSM24252	AXBG01000000
Sulfitobacter sp. NAS-14.1	AALZ00000000
Sulfitobacter pontiacus DSM10014	NZ_FNNB00000000
Sulfitobacter sp. CB2047	JPOY01000000
Sulfitobacter geojensis MM-124	JASE01000000
Sulfitobacter sp. EE36	AALV00000000
Sulfitobacter mediterraneus KCTC 32188	JASH01000000
Sulfitobacter pseudonitzschiae F5	*This study*
Sulfitobacter pseudonitzschiae SA11	-[†]
Sulfitobacter pseudonitzschiae SMR1	NZ_CP022415
Agrobacterium tumefaciens Ach5	NZ_CP011246

[*]JGI Genome Portal accession: IMG_2521172554
[†]JGI Genome Portal accession: IMG_2519103045

Table S7. List of accession numbers for Alteromonadaceae genomes used in the phylogenomic analysis of the Alteromonadaceae MAGs and a laboratory cultured consortium strain (Fig. S4). *Pseudomonas syringae* CC1557 genome was used as an outgroup.

Bacterial strain	NCBI accession number
Alteromonas abrolhosensis PEL67E	NZ_MEJH00000000
Alteromonas addita R10SW13	NZ_CP014322
Alteromonas aestuariivivens KCTC52655	NZ_QRHA00000000
Alteromonas australica H17	CP008849
Alteromonas confluentis KCTC42603	NZ_MDHN00000000
Alteromonas gracilis 9a2	NZ_PVNO00000000
Alteromonas lipolytica JW12	NZ_MJIC00000000
Alteromonas macleodii ATCC27126	CP003841
Alteromonas macleodii Black Sea 11	CP003845
Alteromonas macleodii HOT1A3	CP012202
Alteromonas marina AD001	NZ_JWLW00000000
Alteromonas mediterranea DE	CP001103
Alteromonas mediterranea MED64	CP004848
Alteromonas naphthalenivorans SN2	CP002339
Alteromonas pelagimontana 5.12	NZ_NGFM00000000
Alteromonas stellipolaris LMG21861	CP013926
Alteromonas stellipolaris PQQ-42	CP01534
Alteromonas stellipolaris PQQ-44	CP01534
Alteromonas macleodii F12	*This study*
Pseudomonas syringae CC1557	NZ_CP007014

Table S8. High-throughput sequencing information and read counts for the *A. glacialis* A3 genome (A3Ax), *A. glacialis* A3 RNA-seq samples (D1-D12), bacterial consortium metagenome (A3Bact), and bacterial RNA-seq samples (B1-B9).

Sample name	Description	Number of reads (raw)	Number of quality trimmed reads	Retained (%)	Lost (%)
A3Ax	Diatom genome	165811606	153290706	92.4	7.6
D1	Diatom only at 0.5 hours - rep 1	55597046	45927566	82.6	17.4
D2	Diatom only at 0.5 hours - rep 2	32696296	25491292	78.0	22.0
D3	Diatom only at 0.5 hours - rep 3	27352918	21994556	80.4	19.6
D4	Diatom only at 24 hours - rep 1	41701950	33661898	80.7	19.3
D5	Diatom only at 24 hours - rep 2	31989840	24022442	75.1	24.9
D6	Diatom only at 24 hours - rep 3	27881510	23237124	83.3	16.7
D7	Reseeded diatom at 0.5 hours - rep 1	26184068	20373958	77.8	22.2
D8	Reseeded diatom at 0.5 hours - rep 2	32222538	24054500	74.7	25.3
D9	Reseeded diatom at 0.5 hours - rep 3	37628278	29615596	78.7	21.3
D10	Reseeded diatom at 24 hours - rep 1	29849476	23719096	79.5	20.5
D11	Reseeded diatom at 24 hours - rep 2	32760398	26860754	82.0	18.0
D12	Reseeded diatom at 24 hours - rep 3	32198030	25647272	79.7	20.3
A3Bact	Bacterial consortium metagenome	905113782	828297028	91.5	8.5
B1	Consortium only at 0.5 hours - rep 1	204707556	135964890	66.4	33.6
B2	Consortium only at 0.5 hours - rep 2	123255494	75388906	61.2	38.8
B3	Consortium only at 0.5 hours - rep 3	106506474	66325910	62.3	37.7
B4	Reseeded consortium at 0.5 hours - rep 1	129606370	56559286	43.6	56.4
B5	Reseeded consortium at 0.5 hours - rep 2	103087344	79271378	76.9	23.1
B6	Reseeded consortium at 0.5 hours - rep 3	50236330	32005540	63.7	36.3
B8	Reseeded consortium at 24 hours - rep 1	41267162	12780816	31.0	69.0
B9	Reseeded consortium at 24 hours - rep 2	205670530	146660984	71.3	28.7

Dataset S1. (Excel CSV format) List of retention times (RT) and mass-to-charge values (m/z) for axenic and reseeded samples at four timepoints (0.5, 4, 24, and 48 hours) analyzed on an ultrahigh-performance liquid chromatography quadrupole time-of-flight mass spectrometer (UHPLC-QToF-MS).

Dataset S2. (Excel CSV format) List of expected masses (m/z) (ExpMass), peak intensities, theoretical masses (ThMass) and chemical formulae for the DOM composition in axenic samples using a Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS).

Dataset S3. (Excel CSV format) List of expected masses (m/z) (ExpMass), peak intensities, theoretical masses (ThMass) and chemical formulae for the DOM composition in reseeded samples using a Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS).

Dataset S4. (Text file) Hidden Markov model (HMM) profile of the resulting hits from the consortium isolates and the AzeR sequence.

SI References

1. G. Behringer *et al.*, Bacterial communities of diatoms display strong conservation across strains and time. *Front Microbiol* **9**, 659 (2018).
2. J. Ryther, R. Guillard, Studies of marine planktonic diatoms: II. Use of Cyclotella nana Hustedt for assays of vitamin B12 in sea water. *Canadian Journal of Microbiology* **8**, 437-445 (1962).
3. L. E. Brand, R. R. Guillard, L. S. Murphy, A method for the rapid and precise determination of acclimated phytoplankton reproduction rates. *Journal of plankton research* **3**, 193-201 (1981).
4. C. E. ZoBell, Studies on marine bacteria. I. The cultural requirements of heterotrophic aerobes. *J Mar Res* **4**, 42-75 (1941).
5. M. Lunau, A. Lemke, K. Walther, W. Martens-Habbena, M. Simon, An improved method for counting bacteria from sediments and turbid environments by epifluorescence microscopy. *Environmental Microbiology* **7**, 961-968 (2005).
6. S. Andrews (2010) FastQC: a quality control tool for high throughput sequence data. (Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom).
7. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
8. R. Kajitani *et al.*, Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res* **24**, 1384-1395 (2014).
9. A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072-1075 (2013).
10. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
11. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12**, 357-360 (2015).
12. H. Li *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
13. M. Pertea *et al.*, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology* **33**, 290 (2015).
14. D. M. Bryant *et al.*, A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell reports* **18**, 762-776 (2017).
15. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
16. J. J. Almagro Armenteros, C. K. Sønderby, S. K. Sønderby, H. Nielsen, O. Winther, DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33**, 3387-3395 (2017).
17. M. Krzywinski *et al.*, Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-1645 (2009).
18. P. Menzel, K. L. Ng, A. Krogh, Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* **7**, 11257 (2016).

19. D. Li, C. M. Liu, R. Luo, K. Sadakane, T. W. Lam, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-1676 (2015).
20. D. D. Kang, J. Froula, R. Egan, Z. Wang, MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
21. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043-1055 (2015).
22. A. M. Eren *et al.*, Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).
23. R. L. Rodriguez *et al.*, The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res* **46**, W282-W288 (2018).
24. T. Seemann, Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069 (2014).
25. T. Magoc, S. L. Salzberg, FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957-2963 (2011).
26. E. Kopylova, L. Noe, H. Touzet, SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211-3217 (2012).
27. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
28. A. Roberts, L. Pachter, Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Methods* **10**, 71-73 (2013).
29. B. E. Suzek, H. Huang, P. McGarvey, R. Mazumder, C. H. Wu, UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**, 1282-1288 (2007).
30. E. A. Franzosa *et al.*, Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods* **15**, 962-968 (2018).
31. M. Ashburner *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).
32. R. Caspi *et al.*, The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **44**, D471-480 (2016).
33. R. C. Team, R: a language for statistical computing. *Computer software]. Vienna, Austria: R Foundation for Statistical Computing, http://www.R-project. org* (2015).
34. H. Wickham, *ggplot2: elegant graphics for data analysis* (Springer, 2016).
35. N. E. Hamilton, M. Ferry, ggtern: Ternary diagrams using ggplot2. *Journal of Statistical Software* **87**, 1-17 (2018).
36. J. Xia, I. V. Sinelnikov, B. Han, D. S. Wishart, MetaboAnalyst 3.0--making metabolomics more meaningful. *Nucleic Acids Res* **43**, W251-257 (2015).
37. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).
38. S. Koren *et al.*, Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722-736 (2017).
39. M. J. Chaisson, G. Tesler, Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC bioinformatics* **13**, 238 (2012).
40. B. J. Walker *et al.*, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).

41. M. J. Ankenbrand, A. Keller, bcgTree: automatized phylogenetic tree building from bacterial core genomes. *Genome* **59**, 783-791 (2016).
42. S. R. Eddy, A new generation of homology search tools based on probabilistic inference. *Genome Inform* **23**, 205-211 (2009).
43. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).
44. J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution* **17**, 540-552 (2000).
45. J. Huerta-Cepas, F. Serra, P. Bork, ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol Biol Evol* **33**, 1635-1638 (2016).
46. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
47. C. Bez *et al.*, AzeR, a transcriptional regulator that responds to azelaic acid in Pseudomonas nitroreducens. *Microbiology* **166**, 73-84 (2020).
48. E. Villar *et al.*, The Ocean Gene Atlas: exploring the biogeography of plankton genes online. *Nucleic Acids Res* **46**, W289-W295 (2018).
49. S. El-Gebali *et al.*, The Pfam protein families database in 2019. *Nucleic Acids Res* **47**, D427-D432 (2019).
50. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461 (2010).
51. S. Capella-Gutierrez, J. M. Silla-Martinez, T. Gabaldon, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).
52. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
53. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**, W256-W259 (2019).