Molecular mechanisms of microbiome modulation by the diatom secondary metabolite azelaic acid

Ahmed A. Shibl[1,a], Michael A. Ochsenkühn[1], Amin R. Mohamed[1], Lisa Coe[1], Yejie Yun[1,b], Shady A. Amin[1,*]

[1]Marine Microbial Ecology Laboratory, Biology Program, New York University Abu Dhabi, Abu Dhabi 129188, United Arab Emirates

[a]Current affiliations: Genetic Heritage Group, Biology Program, New York University Abu Dhabi, Abu Dhabi 129188, United Arab Emirates & Public Health Research Center, New York University Abu Dhabi, Abu Dhabi 129188, United Arab Emirates

[b]Current affiliation: Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA 15213

[*]Corresponding author: samin@nyu.edu

Keywords

azelaic acid | microbiomes | signaling | diatom | phytoplankton–bacteria interactions | secondary metabolites

Abstract

Photosynthetic eukaryotes, such as plants and microalgae, modulate their microbiome using the dicarboxylate metabolite azelaic acid (Aze), a molecule that is used as a carbon source for some heterotrophs but is toxic to others. Uptake and assimilation mechanisms of Aze into bacterial cells are mostly unknown, nor its ability to promote or inhibit growth. Here, we use transcriptomics, isotope labeling, and coexpression networks of master transcriptional factors in two marine bacteria to identify the first putative Aze transporter in bacteria, map Aze catabolism through fatty acid degradation and downstream pathways, infer Aze toxicity to the ribosome, and show that Aze catabolism is restricted to 13 bacterial families across terrestrial and marine ecosystems dominated by algal and plant symbionts. Seawater mesocosms amended with Aze enrich for bacterial families that are able to catabolise Aze. These findings shed light on the role of infochemicals in modulating eukaryote-microbiome interactions across diverse ecosystems.

Introduction

The C_9 dicarboxylic acid, azelaic acid (hereafter Aze) is a ubiquitous, yet enigmatic metabolite produced by photosynthetic organisms, such as plants in terrestrial ecosystems and phytoplankton in aquatic environments (Khakimov et al., 2014; Shibl et al., 2020). It is postulated to be a product of the peroxidation of galactolipids (Zoeller et al., 2012), yet its exact biosynthetic pathway is unknown. Aze plays a crucial role as an infochemical that primes systemic acquired resistance against phytopathogens (Jung et al., 2009; Spoel & Dong, 2012; Wittek et al., 2014) and influences microbial diversity in soil (Korenblum et al., 2020). In marine environments, diatoms secrete Aze to modulate bacterial populations by promoting the growth of symbionts while inhibiting opportunists simultaneously (Shibl et al., 2020). In pharmacology, Aze is widely used in human skin care products for its antimicrobial properties to treat inflammatory acne, rosacea, and other dermatological disorders since the 1970s (Nazzaro-Porro, 1987; Searle et al., 2020). Despite this undeniable importance, no Aze uptake proteins are known in bacteria and its assimilation into bacterial cells to promote or inhibit growth is poorly understood. A single gene has recently been identified in *Pseudomonas nitroreducens* that acts as a transcriptional regulator of Aze, *azeR* (Bez et al., 2020), yet its regulatory target(s) are unknown. Understanding the molecular mechanisms that enable the transport and assimilation of Aze into bacterial cells to induce positive or negative phenotypes will shed light on how eukaryotic hosts influence and control their microbiome.

To examine how bacteria transport and catabolise Aze, we used *Phaeobacter* sp. F10 (hereafter *Phaeobacter*) as a model to study how Aze promotes bacterial growth and *Alteromonas macleodii* F12 (hereafter *A. macleodii*) as a model to study the inhibitory effects of Aze. Both bacteria were isolated from the diatom *Asterionellopsis glacialis* that was shown to produce Aze to modulate its microbiome (Shibl et al., 2020). A combination of transcriptomics, transcriptional coexpression networks of master transcriptional regulators, isotope labeling, and metabolomics were used to elucidate how these bacteria process Aze. In addition, *in situ* mesocosms with seawater microbial communities were conducted to examine how Aze influences microbial diversity in natural environments. We discover the first putative Aze transport system in bacteria and map out how each bacterium responds to and processes Aze through a complex regulatory network. In addition, we show that the Aze transport system and transcriptional regulator are prevalent in 13 families of bacteria spanning the terrestrial and marine environments and that *in situ* experiments with seawater microbes enrich microbial taxa that contain these genes.

Materials and Methods

Growth of bacterial isolates with Aze

Bacteria were grown for 24-48 hours in Zobell marine broth (ZoBell, 1941), then centrifuged and resuspended in sterile 10% Zobell marine broth at an OD_{600nm} of 0.3. For *Phaeobacter*, aliquots of this stock were used to inoculate replicate flasks containing 1 L 10% Zobell marine broth each, for a final approximate OD_{600nm} of 0.0003 (~$7.5x10^5$ cells/mL). For *A. macleodii*, this stock was used to inoculate sterile 10% marine broth solution at a final OD_{600nm} of 0.2. This was needed to evade the inhibitory effect of Aze on *A. macleodii* cells to achieve enough viable mRNA. Filter-sterilized azelaic acid (Sigma-Aldrich) was added to a final concentration of 500 µM to all the treatments, and an equal volume of filter-sterilized Milli-Q water was added to controls. All cultures were shaken in a shaker-incubator at 26°C for 8 hours. Growth was estimated by measuring OD_{600nm}. Cells were either centrifuged at 13,000 x g and cell pellets frozen or were filtered onto 0.22 µm Sterivex cartridges (Merck Millipore, Burlington, MA, USA) using a peristaltic pump at a flow rate of 40 mL/min. All cells were flash frozen in liquid nitrogen and stored at -80°C until RNA extraction.

RNA extraction and sequencing

Cartridges containing *Phaeobacter* cells were thawed on ice before extraction, while cell pellets of *A. macleodii* were processed directly from the freezer. Filter membranes were removed with sterile tweezers from cartridges and placed directly in the RLT lysis buffer (Qiagen). RNA was extracted from cell pellets and filter membranes using the RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. RNA samples were treated with DNase I to eliminate genomic DNA contamination and sent to NovogeneAIT Genomics (Singapore) for library preparation using the NEBNext Ultra RNA Library Prep Kit and paired-end (2x150) sequencing on the Illumina NovaSeq 6000 (San Diego, CA) platform. All time points were carried out in three biological replicates except for *Phaeobacter* at 0.5 hours, which had six replicates. All samples passed RNA QC except for one control replicate from *Phaeobacter* at 0.5 hours.

RNAseq analysis

Raw reads of *Phaeobacter* and *A. macleodii* samples were processed using fastp v0.22 (Chen et al., 2018) for quality filtering, adapter removal, and trimming. The resulting sequences were then aligned with their respective genomes using Bowtie2 v2.3.5 (Langmead & Salzberg, 2012). Resulting BAM files were processed on SAMtools v1.12 (Li et al., 2009) and read counts were generated with featureCounts v2.0.3 (Liao et al., 2014). Differential expression analysis

between treatment and control samples was done using DESeq2 v3.14 (Love et al., 2014) on R v4.1 (R Core Team). Genes were considered differentially expressed (DE) if they had a *p*-adjusted value of <0.05 and $\log_2$ fold change of ≥± 0.5. Pearson coefficient calculations and correlation plots were done using the R gplots package (Warnes et al., 2016). DE genes were fed into the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database and the expression ratios for pathways of interest were calculated as follows:

$$Expression\ ratio\ = \frac{No.upregulated\ reactions\ +\ No.downregulated\ reactions}{No.reactions\ in\ pathway\ present\ in\ the\ genome}$$

Scatter plots of genes in enriched KEGG pathways and bar plots of their expression ratios were generated using the R ggplot2 package (Wickham, 2016).

Phylogenetic analysis

Amino acid sequences of the azelaic acid transcriptional regulators, *azeR*, and azelaic acid TRAP substrate-binding protein (*azeT*) of *Phaeobacter* and *Pseudomonas nitroreducens* DSM9128 were used as queries on BLASTp. The search was done against the *nr* database with an e-value cutoff of 1e-50. From the resulting hits, duplicates and sequences less than 250 amino acids were removed. In addition, any sequences with an unknown or unclassified taxonomy were removed. Only species containing both proteins were retained, and their homologs with the lowest e-value and highest bitscore were used for phylogenetic inference. SeqKit (Shen et al., 2016) was used to concatenate *azeR* and *azeT* sequences from the same taxa. The multiple sequence alignment was built by Kalign3 (Lassmann, 2020) and a maximum-likelihood tree was generated with FastTree v2.1.11 (Price et al., 2010) with 1,000 bootstraps. The resulting phylogenetic tree was visualized and annotated on iToL v5 (Letunic & Bork, 2021) and Inkscape v1.1.1 (https://inkscape.org/).

Isotope labeling and metabolomics

Phaeobacter cells were grown as described above. Filter-sterilized azelaic acid-$^{13}C_9$ (Sigma-Aldrich) was added to triplicate cultures with an OD_{600nm} of 0.3 at a final concentration of 500 µM and cultures were shaken in a shaker-incubator at 26°C. Two-mL samples were collected at 0, 15, 30, 60 and 120 minutes after addition of azelaic acid and centrifuged at 13,000 x g for 5 minutes at 4°C, resuspended and washed with 35 g/L NaCl in PBS, then centrifuged and resuspended in ice cold 100% methanol. Subsequently, samples were sonicated for 2 minutes on ice, then centrifuged at 4°C for 5 minutes. Supernatant was collected and dried under nitrogen flow, then extract was stored at -20°C until analysis using mass spectroscopy.

Samples for direct infusion measurements are prone to salt interference and thus solid-phase extraction with PPL-bond elut columns (Agilent Technologies, US) was applied. High-resolution mass spectra were acquired on a Bruker solariX XR Fourier-transform ion cyclotron resonance mass spectrometer (FT-MS) (Bruker Daltonics GmbH, Germany) equipped with a 7T superconducting magnet and operated in ESI (+) ionization mode with 0.5 bar nebulizer pressure, 4.0 L dry gas and a 220 °C capillary temperature. Source optics parameters were 200 V at capillary exit, 220 V deflector plate, 150V funnel 1, 15V skimmer 1, 150 Vpp funnel RF amplitude, octopole frequency of 5 MHz and an RF amplitude of 450 Vpp. Para Cell parameters were set to -20V transfer exit lens, -10V analyzer entrance, 0V side kick, 3V front and back trap plates, -30V back trap plate quench and 24% sweep excitation power. For sample injection, a Triversera Nanomate (Advion BioSciences Inc., US) with 1 psi gas pressure and ESI (+) voltage of 1.7kV was used. Three-hundred and twenty scans were accumulated for each sample with an accumulation time of 200 ms and acquired with a time domain of 4 mega words over a mass range of m/z 75 to 1200, at an optimal mass range from 200-600 m/z. Spectra were internally calibrated using primary metabolites (e.g., amino acids, organic acids) in Data analysis 5.0 Software (Bruker Daltonics GmbH, Bremen, Germany). The FT-MS mass spectra were exported to peak lists with a cut-off signal-to-noise ratio (S/N) of 4. The masses for relevant ^{13}C isotope-labeled metabolites were calculated using enviPat: isotope pattern calculator (Loos et al., 2015). MSMS was performed at 15V, 25V and 35V collision voltage and characteristic fragments identified manually.

Seawater mesocosm and 16S rRNA sequencing

Seawater samples were collected in July 2021 from surface waters off the coast of Saadiyat Island (24°38'28.6"N 54°27'09.4"E), United Arab Emirates, kept in the dark, and immediately brought back to the lab. For a mesocosm experiment, the seawater was filtered through a 1.2-µm filter (Whatman, UK) to remove most phytoplankton and divided into 12 vessels each containing 30 mL. Control replicates (n=4 at T=0 hours) were immediately filtered onto a 0.2-µm filter (Whatman, UK) to capture microbial cells. Treatment replicates (n=4) with 100 µM Aze and another set of control replicates (n=4) with an equal volume of Milli-Q water added in lieu of Aze were incubated at 24°C for 16 hours in the dark (T=16 hours). After incubation, they were immediately filtered onto a 0.2-µm filter (Whatman, UK). Genomic DNA was extracted immediately after filtration from all filters using the DNeasy PowerWater Kit (Qiagen) and sent to NovogeneAIT Genomics (Singapore) for 16S rRNA amplification using the 515F-Y (5′-GTGYCAGCMGCCGCGGTAA-3′) and 926R (5′-CCGYCAATTYMTTTRAGTTT-3′) universal

primers and paired-end (2x250bp) sequencing on the Illumina NovaSeq 6000 (San Diego, CA) platform.

Clean raw reads were processed with the rANOMALY R package (Theil & Rifa, 2021) using the dada2 R package (Callahan et al., 2016) to generate amplicon sequence variants (ASVs). Taxonomic classification was based on the SILVA v138 database. Alpha-diversity was assessed using the richness indices; observed ASVs and Chao1, and diversity indices; Shannon and Simpson. Beta-diversity was assessed and visualized by principal coordinate analysis (PCoA) based on pairwise Bray-Curtis distance and tested by the permutational analysis of variance (PERMANOVA). The differential abundance of taxa across the treatment and control samples was calculated with DESeq2 v3.14 (Love et al., 2014) with a *p*-adjusted value cutoff of < 0.05. All plots were generated using the ggplot2 and phylosmith (Smith, 2019) R packages and all statistical analyses were performed on R 4.1.2.

Transcriptional master regulator analysis

We deployed the regulatory impact factor (RIF) algorithm (Reverter et al., 2010) to detect transcription factors (TFs) with high regulatory potential contributing to the observed transcriptional remodeling upon Aze uptake. RIF is designed to identify key regulatory loci contributing to transcriptome divergence between two biological conditions. Predicted TF lists were obtained from the genome annotation (GFF file) of both *Phaeobacter* and *A. macleodii* and normalized data (variance stabilized, $\log_2$-transformed counts) of these genes were retrieved. The same normalization strategy was applied to the DE genes. An expression matrix containing normalized expression of the TF genes and DE genes were subjected to RIF analysis per species. RIF analysis identifies regulators that are consistently differentially co-expressed with highly abundant and highly DE genes (RIF1 metric) and regulators that have the most altered ability to act as predictors of the abundance of DE genes (RIF2 metric). TFs were considered significant when the RIF score deviates ±1.96 standard deviation from the mean (t-test, $P < 0.05$).

Transcriptional coexpression networks (TCNs)

The Partial Correlation and Information Theory (PCIT) (Reverter & Chan, 2008) has been extensively used for gene network analysis (Alexandre et al., 2021; Botwright et al., 2021). We utilized PCIT to detect significant connections (edges) between pairs of genes (nodes) while considering the influence of a third player (gene). PCIT calculates pairwise correlations between given gene pairs after considering all possible three-way combinations of genes (triads) present within a gene expression matrix. The algorithm calculates partial correlations after exploring all

194 triads before determining the significance threshold that depends on the average ratio of partial
195 and direct correlation. The set of key regulatory TFs (identified by RIF analysis) and DE genes
196 per species were used for construction of the networks. Normalized data (variance stabilized,
197 log_2-transformed counts) of these genes were used for network construction. The PCIT-inferred
198 networks were visualized using Cytoscape Version 3.9 (Shannon et al., 2003). From these initial
199 networks, we explored a series of subnetworks; first connections between nodes were considered
200 when the partial correlation *r* was greater than ± 0.95. From these networks, hub genes (potential
201 regulatory components within the network) and their connected genes (first neighbors) were
202 extracted based on: 1) key regulatory factors (with highest RIF scores), 2) differential expression
203 significance, and 3) degree centrality (the number of connections of a node with other nodes in
204 the network).

205 Data availability

206 RNAseq raw reads of *Phaeobacter* are deposited in NCBI under the BioProject number
207 PRJNA823575. RNAseq raw reads of *A. macleodii* are deposited in NCBI under the BioProject
208 number PRJNA823732. 16S rRNA amplicon sequencing raw reads are deposited in NCBI under
209 the BioProject number PRJNA823745. All software packages used are open source.

210

211 **Results**

212 Growth and Transcriptional Response to Aze

213 To examine the effects of Aze on *Phaeobacter* and *A. macleodii*, each bacterium was
214 grown in the presence of 500 µM Aze and growth was compared to controls without Aze. As
215 reported previously (Shibl et al., 2020), Aze induced a statistically significant increase in growth
216 and overall cell yield of *Phaeobacter* while temporarily inhibiting the growth of *A. macleodii*
217 (Extended Data Fig. 1). To capture the transcriptional responses of these bacteria to Aze addition,
218 transcriptomic analyses were conducted from RNA samples collected at 0.5 and 8 hours after
219 Aze addition in *Phaeobacter*, and at 0.5 hours from *A. macleodii*.

220 *Phaeobacter* differentially expressed 273 and 558 genes in response to Aze at 0.5 and 8
221 hours, respectively, corresponding to ~20% of all CDS in its genome (Extended Data Table 1).
222 Most differentially expressed genes were upregulated (652) and most occurred at 8 hours (494).
223 Only 70 genes were co-regulated at both 0.5 and 8 hours, suggesting the immediate, short-term
224 response to Aze is generally different from long-term growth on the substrate. Indeed, hierarchical
225 clustering of Pearson correlation coefficients for differentially expressed genes across all

conditions showed distinct clustering of growth on Aze at 0.5 hours relative to the other conditions (Extended Data Fig. 2a). In contrast, *A. macleodii* differentially expressed 274 genes at 0.5 hours, corresponding to ~7% of all CDS in its genome, half of which were upregulated (Extended Data Table 1, Extended Data Fig. 2b).

Global analysis of KEGG pathways of all three transcriptomes show significant contrasts between *Phaeobacter* and *A. macleodii* (Fig. 1). While fatty acid degradation, amino acid metabolism and degradation, ABC transporters, oxidative phosphorylation, propanoate metabolism, and glyoxylate and dicarboxylate metabolism pathways exhibited similar expression ratios (*see* Methods) across both bacteria in response to Aze, upregulated genes in these pathways were mostly dominated by *Phaeobacter* transcripts while downregulated genes were mostly enriched in *A. macleodii* transcripts. Interestingly, fatty acid degradation genes were among the most upregulated genes in *Phaeobacter* (Fig. 1).

Aze Catabolism by *Phaeobacter*

No uptake genes are known for Aze. A putative C_4-dicarboxylate TRAP transporter substrate-binding protein (INS80_RS11065) was the most and the third most upregulated gene in *Phaeobacter* grown on Aze at 0.5 and 8 hours, respectively. The small and large permease genes (INS80_RS11060 and INS80_RS11055) neighboring this gene were among the top 20 most upregulated genes in the transcriptome at 0.5 hours and continued to be upregulated at 8 hours, implicating these genes in transporting Aze (Fig. 2, Extended Data Table 2, Supplementary Data 1). All three genes are co-localized with the putative Aze transcriptional regulator, *azeR* (INS80_RS11050, belonging to the *lclR* family transcriptional factor) (Shibl et al., 2020), which is upregulated at 0.5 hours. *In silico* operon predictions show that *azeR*, the small and large permeases all fall within a single operon while the substrate-binding gene is transcribed independently (Extended Data Fig. 3). We assign this cluster of genes the designation *azeTSLR* (*T*=substrate-binding protein, *S*=small permease, *L*=large permease, *R*=regulator). Once inside the cell, Aze appeared to be initiated into the fatty acid degradation pathway via a PaaI family thioesterase and acyl-coA ligase (*fadD*) that were present directly downstream of *azeTSLR* and were putatively co-transcribed and upregulated at 0.5 hours (Fig. 2, Extended Data Fig. 3, Extended Data Table 2). Aze-coA was then putatively degraded through a series of steps catalyzed by genes in the fatty acid degradation pathway (*acd, paaF, and fadN*) that were among the most highly upregulated genes at 0.5 hours and continued to be upregulated at 8 hours (Fig. 2, Extended Data Table 2, Supplementary Data 1). These successive reactions putatively

liberated two acetyl-coA molecules to generate glutaryl-coA. Glutaryl-coA can either be further degraded into acetyl-coA via *gcdH*, *paaF*, *fadN* and *atoB*, all of which were upregulated at both time points, or can be converted to glutaconyl-coA (Fig. 2).

To confirm the patterns observed in the *Phaeobacter* transcriptome and resolve the fate of glutaryl-coA, *Phaeobacter* was supplemented with $^{13}C_9$-Aze, and intracellular metabolomics samples were extracted and analyzed using a Fourier-transform ion cyclotron resonance mass spectrometer. Within 15-30 mins of addition of labeled Aze to cells, $^{13}C_9$-Aze, $^{13}C_9$-Azeloyl-coA, $^{13}C_7$-pimeloyl-coA and $^{13}C_5$-glutaryl-coA were detected and increased in relative abundance over the course of 2 hours (Fig. 2, Supplementary Data 3). In addition, $^{13}C_5$-glutaconyl-coA was detected at 2 hours and $^{13}C_3$-propionate-coA at 1-2 hours after addition of $^{13}C_9$-Aze. Glutaconyl-coA is a side product of glutaryl-coA that is shuttled into butanoate metabolism (Djurdjevic et al., 2011), while propionyl-coA is an important intermediate in glyoxylate and propanoate metabolism, indicating these pathways are activated downstream of Aze catabolism. Indeed, while fatty acid degradation was more upregulated at 0.5 hours than at 8 hours, glyoxylate and dicarboxylate metabolism, butanoate metabolism and other downstream pathways were more upregulated at 8 hours than at 0.5 hours (Fig. 3a, Extended Data Table 2).

Toxicity of Aze in *A. macleodii*

The genome of *A. macleodii* completely lacks any TRAP transport systems, suggesting Aze crosses the cell membrane non-specifically. In contrast to the *Phaeobacter* transcriptome, the highest upregulated gene 0.5 hours after Aze addition to *A. macleodii* was an efflux RND transporter periplasmic subunit (GKZ85_RS10170). The permease and outer membrane subunits of this efflux pump (GKZ85_RS10160, GKZ85_RS10165) that belong to the same operon were among the top 10 most upregulated genes in the *A. macleodii* transcriptome (Fig. 3b, Extended Data Table 3, Supplementary Data 2). This efflux system presumably removes Aze from the cytoplasm. Several other efflux transporters and ABC multidrug transporters were among the most upregulated genes in the presence of Aze relative to controls. Among upregulated genes in the presence of Aze were secretion systems, protein export, spore formation, stress-response proteins, heat shock proteins, proteases, and ribosome protection genes (Extended Data Table 3, Supplementary Data 2). Most ribosomal genes were either downregulated or not differentially expressed while fatty acid degradation was not differentially expressed in the presence of Aze. In addition, nucleotide metabolism, pyruvate metabolism, oxidative phosphorylation, electron transfer, amino acid metabolism and biosynthesis, and fatty acid biosynthesis were

downregulated (Figs. 1, 3b, Extended Data Table 3). Collectively, these transcriptomic patterns indicate *A. macleodii* is mitigating the toxic effects of Aze by effluxing it out of the cell while arresting cellular metabolism. Interestingly, several genes involved in ribosome protection and response to protein degradation (*rmf*, *hpf*, *hslR*, *hslU*, *hslV*) were upregulated, suggesting Aze may be disrupting protein synthesis (Extended Data Table 3).

AzeR and *xre* Transcriptional Networks

Due to the stark transcriptional response to Aze in each bacterium, we hypothesized that transcriptional regulation plays a major role in activating essential pathways that catabolise or mitigate the toxicity of Aze. To identify transcriptional master regulators for *Phaeobacter* and *A. macleodii*, we performed regulatory impact factor (RIF) analyses (Reverter et al., 2010). Putative transcriptional factors (TF) in each species were compared to the unique corresponding differentially expressed gene lists to identify only 29 and 12 regulators with significant scores (deviating ± 1.96 standard deviations from the mean; $P < 0.05$) for *Phaeobacter* and *A. macleodii*, respectively (Supplementary Data 4). Interestingly, *azeR* was the top TF identified in *Phaeobacter* and was also differentially expressed. In contrast, only two TFs namely XRE family transcriptional regulator (*xre*), a bacterial TF associated with stress tolerance (Hu et al., 2019), and helix-turn-helix transcriptional regulators (GKZ85_RS06065 and GKZ85_RS16495) were identified while being differentially expressed in *A. macleodii*. To get insights into putative regulatory mechanisms during Aze uptake, we used transcriptional coexpression networks, which enabled characterization of transcriptional patterns during response to disease, stress, or development (Hartl et al., 2021; Rose et al., 2016; Yao et al., 2015).

Significant connections ($r \geq \pm 0.95$) within the initial network identified 286 genes with 7,208 connections in *Phaeobacter* and 274 genes with 7,102 connections in *A. macleodii* (Extended Data Figure 4). Subnetworks that contain *azeR* and *xre* show that they act as hub genes in putatively regulating specific categories of genes in *Phaeobacter* and *A. macleodii*, respectively. The subnetwork of *azeR* shows that it putatively regulates transcription of most of the pathways that are involved in assimilating Aze in *Phaeobacter*, including *azeTSL*, fatty acid degradation, butanoate metabolism, glyoxylate and dicarboxylate metabolism and oxidative phosphorylation, in addition to other genes that may be indirectly affected by Aze catabolism (Figs. 3a, 4a). In contrast, the subnetwork of *xre* shows that it putatively regulates transcription of the efflux system implicated in removal of Aze from the cytoplasm of *A. macleodii*, stress proteins and proteases, protein export, nucleotide metabolism, and amino acid metabolism and biosynthesis (Missiakas

et al., 1996) (Figs. 3b, 4b). These findings highlight the potential importance of transcriptional regulation in mediating the response to Aze and may explain the large differences in response between both bacteria.

AzeR-AzeT Phylogeny

Based on these results, we hypothesized that for a bacterium to efficiently catabolise Aze, it needs both *azeTSL* to control uptake of Aze and *azeR* to regulate Aze catabolism. Homologs of *azeR* and *azeT* from *Phaeobacter* and *P. nitroreducens* were used to mine bacterial genomes for the presence of both genes simultaneously. Genomes belonging to 13 bacterial families that spanned both marine and terrestrial environments possessed both gene homologs. Surprisingly, Rhodobacteraceae (to which *Phaeobacter* and many phytoplankton symbionts belong) constituted the most evolutionarily divergent family and the most abundant in genomes containing these genes (Fig. 5). Other families included phytopathogens (Pseudomonadaceae), nitrogen-fixing rhizobia (Rhizobiaceae, Nitrobacteraceae, Oxalobacteraceae) and anoxygenic phototrophs (Rhodospirillaceae).

Aze Addition to Seawater Mesocosms

To examine the influence of Aze on microbial populations, surface seawater was collected and pre-filtered to remove larger eukaryotes and enrich for prokaryotes. Seawater mesocosms were supplemented with 100 µM Aze or an equivalent volume of Milli-Q water and incubated in the dark for 16 hours, after which DNA was extracted and 16S rRNA amplicon sequencing was used to assess prokaryotic diversity. Diversity indices of Aze-treated seawater samples (Shannon & Simpson) were significantly lower (Wilcox, $p<0.05$) than control samples after 16 hours, while the richness of the microbial community was significantly higher (Wilcox, $p<0.05$) in Aze-treated seawater samples relative to control samples (Extended Data Fig. 5a). PCoA showed that the microbial community was significantly different (PERMANOVA, $p<0.001$) between each sample group (Extended Data Fig. 5b). Taxonomic classification at the family level showed that Rhodobacteraceae, SAR11 clade I, Cyanobiaceae, and Flavobacteriaceae constituted >60% of prokaryotic diversity across all samples. Notably, the abundance of Rhodobacteraceae increased in the Aze-treated samples (mean ~47%) relative to the T=16 hours control samples (mean ~37%) (Fig. 6a). Differential abundance analysis with DESeq2 (p-adjusted value < 0.05) identified 44 ASVs enriched after Aze addition relative to T=16 hours controls. At the genus level, these ASVs

357 belonged to unclassified Rhodobacteraceae (17), archaeal marine group II (19),
358 *Pseudoalteromonas* (6), *Vibrio* (1), and *Psychrosphaera* (1). Rhodobacteraceae had the highest
359 mean taxonomic proportion of ASVs ranging from ~1.5 to ~95, suggesting a significant reliance
360 on Aze, while the archaeal marine group II had the lowest mean ratios of ASVs ranging from
361 ~0.02 to ~0.04 (Fig. 6b).

362

363 **Discussion**

364 Aze is produced by plants to prime their immune system in response to phytopathogens
365 (Jung et al., 2009). Its toxic effects against some bacteria, including the acne-causing
366 *Propionibacterium acnes*, popularized its use in skin care products (Leeming et al., 1986). Aze
367 has also been shown to have antitumor effects on cancerous cells, as a competitive inhibitor of
368 mitochondrial oxidoreductases and steroid biosynthesis (Sieber & Hegel, 2014). Despite these
369 antagonistic effects, early reports show that some bacteria, mainly *Pseudomonads*, are able to
370 use it as a carbon source (Janota-Bassalik & Wright, 1964). Recently, Aze production by a
371 ubiquitous diatom has been shown to enable it to selectively promote growth of its bacterial
372 symbionts while simultaneously inhibiting growth of opportunists (Shibl et al., 2020). The ability of
373 a single metabolite to selectively inhibit or promote different populations of bacteria
374 simultaneously poses interesting questions about how this metabolite evolved in the eukaryotic
375 chemical repertoire to influence their microbiome.

376 The identification of Aze as the putative substrate of a tripartite ATP-independent
377 periplasmic (TRAP) transporter system is novel. Much less is known about TRAP transporters
378 and their substrates than ABC transporters, even though our knowledge of the substrates of the
379 latter is lacking. TRAP transporters are composed of a substrate-binding domain and two
380 transmembrane segments (large and small permeases), while lacking a nucleotide-binding
381 domain that is canonical in ABC transporters (Rosa et al., 2018). Recognized substrates for TRAP
382 transporters vary from sugars, mono- and di-carboxylates, organosulfur molecules, heterocyclic
383 carboxylates, and amino acids. For dicarboxylates, only C_4-substrates, such as succinate,
384 fumarate, and malate, have been identified as substrates for TRAP transporters, which explains
385 the common annotation of C_4-dicarboxylate transporter in many bacterial genomes, including
386 *Phaeobacter*. On the other hand, dicarboxylates with >4 carbons have no known uptake
387 mechanisms thus far (Mulligan et al., 2011), with the exception of adipate (C_6) that binds with high
388 affinity to a tripartite tricarboxylate transporter (TTT) (Rosa et al., 2017). This makes AzeTSL a
389 putative transporter of the longest chain dicarboxylate molecule known to date. The sheer

abundance of dicarboxylate transporters in bacterial genomes begs the question of whether they transport a variety of C_n-dicarboxylates, such as C_5-glutarate, C_6-adipate, C_7-pimelate, C_8-suberate, C_9-azelate and C_{10}-sebacate. Some of these metabolites are produced by primary producers and thus can potentially serve as growth factors for heterotrophic bacteria (Moran et al., 2022). Further work is needed to expand the substrate of dicarboxylate TRAP transporters.

Whereas TRAP transporters are widespread in bacteria (Mulligan et al., 2007), the IclR-family transcriptional regulators (to which AzeR belongs) are found in a limited number of taxa (Bez et al., 2020). The phylogenetic distribution of the AzeR-AzeT proteins shows that bacteria belonging to the marine Rhodobacteraceae family were evolutionarily highly divergent from other bacterial families, most of which are terrestrial (Fig. 5). Notably, four marine bacterial families (Stappiaceae, Rhizobiaceae, Sneathiellaceae, and Rhodospirillaceae) clustered closer to terrestrial taxa than Rhodobacteraceae, indicating that despite sharing similar ecological traits, *Phaeobacter* and other Rhodobacteraceae uptake and regulatory systems may have evolved as a result of adaptive selection.

We have shown that while *Phaeobacter* uses fatty acid degradation and subsequently glyoxylate and dicarboxylate metabolism and butanoate metabolism pathways to catabolise Aze, *A. macleodii* attempts to mitigate the toxicity of Aze by effluxing it out of the cytoplasm and by downregulating the ribosome and protein synthesis pathways (Fig. 3). However, the fact that *A. macleodii* possesses a complete fatty acid degradation pathway, similar to other bacteria that display inhibition by Aze, begs the question of why such bacteria cannot catabolise this metabolite. The answer may lie in the fact that *Phaeobacter*, and bacteria that catabolise Aze such as *P. nitroreducens*, possess the transcriptional factor AzeR (Fig. 5). Using regulatory impact factor analysis, we have shown that *azeR* in *Phaeobacter* putatively regulates its fatty acid degradation pathway, *azeTSL* along with other pathways essential for the catabolism of Aze (Fig. 4). By activating uptake, fatty acid degradation and other essential genes, AzeR enables bacteria to efficiently break down Aze. In contrast, bacteria deficient in AzeR are likely unable to activate these pathways rapidly and are thus prey to the toxic effects of Aze. Nevertheless, *A. macleodii* is able to recover from Aze, likely through the use of efflux pumps and activation of a variety of stress-related pathways, which appear to be mediated by the transcriptional factor XRE (Fig. 4b). While Aze exhibits a wide range of inhibitory activity, in *A. macleodii* ribosome and/or protein synthesis inhibition seem to be the causative effect of Aze, evidenced by the strong downregulation of most ribosomal protein coding genes, protein synthesis genes and proteases. This finding agrees with previous studies that demonstrated the bacteriostatic effects of Aze on

423 protein, DNA and RNA synthesis in *P. acnes* and *Staphylococcus epidermidis* (Bojar et al., 1988;
424 Bojar et al., 1991).

425 Aze addition to seawater mesocosms induced a statistically significant increase in the
426 relative abundance of Rhodobacteraceae and other taxa compared to controls. While
427 Rhodobacteraceae ASVs comprised the most abundant group to be influenced by Aze additions,
428 highlighting their adaptation to phytoplankton exudates, marine group II archaeal ASVs exhibited
429 the largest enrichment relative to controls (Fig. 6b). While we cannot discount competitive
430 interactions from giving rise to the significant enrichment of marine group II Euryarchaeota ASVs,
431 the possibility that Aze may be a growth substrate for this important group is exciting. Marine
432 group II Euryarchaeota remains uncultivated, hindering our understanding of its metabolic
433 capacity, though isotope probing and metagenomic analysis showed its assimilation of
434 phytoplankton-derived substrates (Orsi et al., 2016). Further work needs to be done to confirm
435 the link, if any, between Euryarchaeota and Aze.

436

437 **Conclusion**

438 Phytoplankton-derived secondary metabolites play a role in structuring algal microbiomes.
439 However, the perception and response of these metabolites on microbial heterotrophs remain
440 largely unknown. Metabolites like Aze may play an important role in ecology by enabling
441 eukaryotic hosts to selectively promote growth of their symbionts while simultaneously inhibiting
442 growth of parasitic bacteria. Bacteria that evolve mechanisms to assimilate Aze gain an
443 advantage over others that cannot in select environments, such as those surrounding
444 phytoplankton cells, known as the phycosphere, or near plant roots, known as the rhizosphere.
445 While Aze represents a single mechanism to structure the microbiome of primary producers, it is
446 likely one of many such metabolites that together work in concert to ensure the maintenance of
447 healthy microbial communities, which ultimately influence higher trophic levels and control global
448 biogeochemical cycling.

449

Acknowledgments

This work was supported by a grant from the Gordon and Betty Moore Foundation to S.A.A. (GBMF9335, https://doi.org/10.37807/GBMF9335) and by a grant from NYU Abu Dhabi to S.A.A. (AD179). We thank the NYU Abu Dhabi Core Technology Platforms for support related to mass spectrometry. We also thank Dain McParland for helping collect seawater samples.

The authors declare no competing interests.

References

Alexandre, P. A., Naval-Sánchez, M., Menzies, M., Nguyen, L. T., Porto-Neto, L. R., Fortes, M. R., & Reverter, A. (2021). Chromatin accessibility and regulatory vocabulary across indicine cattle tissues. *Genome biology*, *22*(1), 1-20.

Bez, C., Javvadi, S. G., Bertani, I., Devescovi, G., Guarnaccia, C., Studholme, D. J., Geller, A. M., Levy, A., & Venturi, V. (2020). AzeR, a transcriptional regulator that responds to azelaic acid in Pseudomonas nitroreducens. *Microbiology*, *166*(1), 73.

Bojar, R., Holland, K., Leeming, J., & Cunliffe, W. (1988). Azelaic acid: its uptake and mode of action in Staphylococcus epidermidis NCTC 11047. *Journal of applied bacteriology*, *64*(6), 497-504.

Bojar, R. A., Holland, K. T., & Cunliffe, W. J. (1991). The in-vitro antimicrobial effects of azelaic acid upon Propionibacterium acnes strain P37. *Journal of Antimicrobial Chemotherapy*, *28*(6), 843-853.

Botwright, N. A., Mohamed, A. R., Slinger, J., Lima, P. C., & Wynne, J. W. (2021). Host-parasite interaction of Atlantic salmon (Salmo salar) and the ectoparasite Neoparamoeba perurans in amoebic gill disease. *Frontiers in immunology*, *12*, 1900.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature methods*, *13*(7), 581-583.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884-i890.

Djurdjevic, I., Zelder, O., & Buckel, W. (2011). Production of glutaconic acid in a recombinant Escherichia coli strain. *Applied and environmental microbiology*, *77*(1), 320-322.

Hartl, C. L., Ramaswami, G., Pembroke, W. G., Muller, S., Pintacuda, G., Saha, A., Parsana, P., Battle, A., Lage, K., & Geschwind, D. H. (2021). Coexpression network architecture reveals the brain-wide and multiregional basis of disease susceptibility. *Nature neuroscience*, *24*(9), 1313-1323.

Hu, Y., Hu, Q., Wei, R., Li, R., Zhao, D., Ge, M., Yao, Q., & Yu, X. (2019). The XRE family transcriptional regulator SrtR in Streptococcus suis is involved in oxidant tolerance and virulence. *Frontiers in cellular and infection microbiology*, *8*, 452.

Janota-Bassalik, L., & Wright, L. (1964). Azelaic acid utilization by a Pseudomonas. *Microbiology*, *36*(3), 405-414.

489 Jung, H. W., Tschaplinski, T. J., Wang, L., Glazebrook, J., & Greenberg, J. T. (2009). Priming in
490 systemic plant immunity. *Science*, *324*(5923), 89-91.

491 Khakimov, B., Jespersen, B. M., & Engelsen, S. B. (2014). Comprehensive and comparative
492 metabolomic profiling of wheat, barley, oat and rye using gas chromatography-mass
493 spectrometry and advanced chemometrics. *Foods*, *3*(4), 569-585.

494 Korenblum, E., Dong, Y., Szymanski, J., Panda, S., Jozwiak, A., Massalha, H., Meir, S.,
495 Rogachev, I., & Aharoni, A. (2020). Rhizosphere microbiome mediates systemic root
496 metabolite exudation by root-to-root signaling. *Proceedings of the National Academy of*
497 *Sciences*, *117*(7), 3874-3883.

498 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*
499 *methods*, *9*(4), 357-359.

500 Lassmann, T. (2020). Kalign 3: multiple sequence alignment of large datasets. In: Oxford
501 University Press.

502 Leeming, J., Holland, K., & Bojar, R. (1986). The in vitro antimicrobial effect of azelaic acid. *British*
503 *Journal of Dermatology*, *115*(5), 551-556.

504 Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic
505 tree display and annotation. *Nucleic acids research*, *49*(W1), W293-W296.

506 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., &
507 Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*,
508 *25*(16), 2078-2079.

509 Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for
510 assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923-930.

511 Loos, M., Gerber, C., Corona, F., Hollender, J., & Singer, H. (2015). Accelerated isotope fine
512 structure calculation using pruned transition trees. *Analytical chemistry*, *87*(11), 5738-
513 5744.

514 Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion
515 for RNA-seq data with DESeq2. *Genome biology*, *15*(12), 1-21.

516 Missiakas, D., Betton, J. M., & Raina, S. (1996). New components of protein folding in
517 extracytoplasmic compartments of Escherichia coli SurA, FkpA and Skp/OmpH. *Molecular*
518 *microbiology*, *21*(4), 871-884.

519 Moran, M. A., Kujawinski, E. B., Schroer, W. F., Amin, S. A., Bates, N. R., Bertrand, E. M.,
520 Braakman, R., Brown, C. T., Covert, M. W., & Doney, S. C. (2022). Microbial metabolites
521 in the marine carbon cycle. *Nature Microbiology*, *7*(4), 508-523.

522 Mulligan, C., Fischer, M., & Thomas, G. H. (2011). Tripartite ATP-independent periplasmic
523 (TRAP) transporters in bacteria and archaea. *FEMS microbiology reviews*, *35*(1), 68-86.

524 Mulligan, C., Kelly, D. J., & Thomas, G. H. (2007). Tripartite ATP-independent periplasmic
525 transporters: application of a relational database for genome-wide analysis of transporter
526 gene frequency and organization. *Microbial Physiology*, *12*(3-4), 218-226.

527 Nazzaro-Porro, M. (1987). Azelaic acid. *Journal of the American Academy of Dermatology*, *17*(6),
528 1033-1041.

529 Orsi, W. D., Smith, J. M., Liu, S., Liu, Z., Sakamoto, C. M., Wilken, S., Poirier, C., Richards, T. A.,
530 Keeling, P. J., & Worden, A. Z. (2016). Diverse, uncultivated bacteria and archaea

underlying the cycling of dissolved protein in the ocean. *The ISME journal*, *10*(9), 2158-2173.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one*, *5*(3), e9490.

Reverter, A., & Chan, E. K. (2008). Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics*, *24*(21), 2491-2497.

Reverter, A., Hudson, N. J., Nagaraj, S. H., Pérez-Enciso, M., & Dalrymple, B. P. (2010). Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. *Bioinformatics*, *26*(7), 896-904.

Rosa, L. T., Bianconi, M. E., Thomas, G. H., & Kelly, D. J. (2018). Tripartite ATP-independent periplasmic (TRAP) transporters and tripartite tricarboxylate transporters (TTT): from uptake to pathogenicity. *Frontiers in cellular and infection microbiology*, *8*, 33.

Rosa, L. T., Dix, S. R., Rafferty, J. B., & Kelly, D. J. (2017). Structural basis for high-affinity adipate binding to AdpC (RPA 4515), an orphan periplasmic-binding protein from the tripartite tricarboxylate transporter (TTT) family in Rhodopseudomonas palustris. *The FEBS Journal*, *284*(24), 4262-4277.

Rose, N. H., Seneca, F. O., & Palumbi, S. R. (2016). Gene networks in the wild: identifying transcriptional modules that mediate coral resistance to experimental heat stress. *Genome Biology and Evolution*, *8*(1), 243-252.

Searle, T., Ali, F. R., & Al-Niaimi, F. (2020). The versatility of azelaic acid in dermatology. *Journal of Dermatological Treatment*, 1-11.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, *13*(11), 2498-2504.

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PloS one*, *11*(10), e0163962.

Shibl, A. A., Isaac, A., Ochsenkühn, M. A., Cárdenas, A., Fei, C., Behringer, G., Arnoux, M., Drou, N., Santos, M. P., & Gunsalus, K. C. (2020). Diatom modulation of select bacteria through use of two unique secondary metabolites. *Proceedings of the National Academy of Sciences*, *117*(44), 27445-27455.

Sieber, M., & Hegel, J. (2014). Azelaic acid: properties and mode of action. *Skin pharmacology and physiology*, *27*(Suppl. 1), 9-17.

Smith, S. (2019). phylosmith: an R-package for reproducible and efficient microbiome analysis with phyloseq-objects. *Journal of Open Source Software*, *4*(38).

Spoel, S. H., & Dong, X. (2012). How do plants achieve immunity? Defence without specialized immune cells. *Nature reviews immunology*, *12*(2), 89-100.

Taboada, B., Estrada, K., Ciria, R., & Merino, E. (2018). Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics*, *34*(23), 4118-4120.

Theil, S., & Rifa, E. (2021). rANOMALY: AmplicoN wOrkflow for Microbial community AnaLYsis. *F1000Research*, *10*.

573 Warnes, G., Bolker, B., & Bonebakker, L. (2016). gplots: various R programming tools for plotting
574 data. *R Package version 3.0*, *1*.

575 Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. springer.

576 Wittek, F., Hoffmann, T., Kanawati, B., Bichlmeier, M., Knappe, C., Wenig, M., Schmitt-Kopplin,
577 P., Parker, J. E., Schwab, W., & Vlot, A. C. (2014). Arabidopsis ENHANCED DISEASE
578 SUSCEPTIBILITY1 promotes systemic acquired resistance via azelaic acid and its
579 precursor 9-oxo nonanoic acid. *Journal of experimental botany*, *65*(20), 5919-5931.

580 Yao, P., Lin, P., Gokoolparsadh, A., Assareh, A., Thang, M. W., & Voineagu, I. (2015).
581 Coexpression networks identify brain region–specific enhancer RNAs in the human brain.
582 *Nature neuroscience*, *18*(8), 1168-1174.

583 ZoBell, C. E. (1941). Studies on marine bacteria. I. The cultural requirements of heterotrophic
584 aerobes. *J mar Res*, *4*, 41-75.

585 Zoeller, M., Stingl, N., Krischke, M., Fekete, A., Waller, F., Berger, S., & Mueller, M. J. (2012).
586 Lipid profiling of the Arabidopsis hypersensitive response reveals specific lipid
587 peroxidation and fragmentation processes: biogenesis of pimelic and azelaic acid. *Plant*
588 *Physiology*, *160*(1), 365-378.

589
590

Figures and Legends

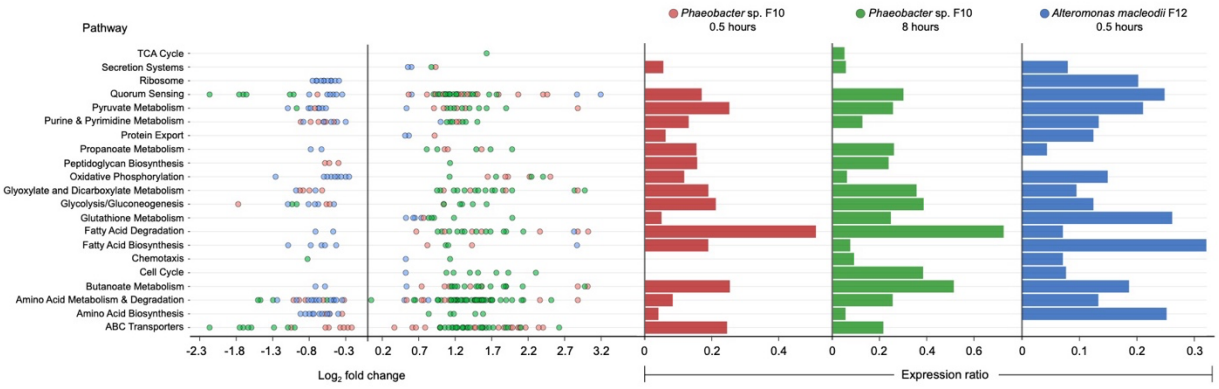


Figure 1. Enriched KEGG pathways in *Phaeobacter* and *A. macleodii*. Left: Log_2 fold-change of differentially expressed (DE) genes present in KEGG pathways listed on the y-axis. Each circle represents the differential expression of a gene in the presence of Aze relative to controls. Right: Expression ratios of KEGG pathways were calculated as indicated in the methods. Circle and bar colors indicate the strain and time point (*Phaeobacter* at 0.5 hours=red; *Phaeobacter* at 8 hours=green; *A. macleodii* at 0.5 hours=blue).

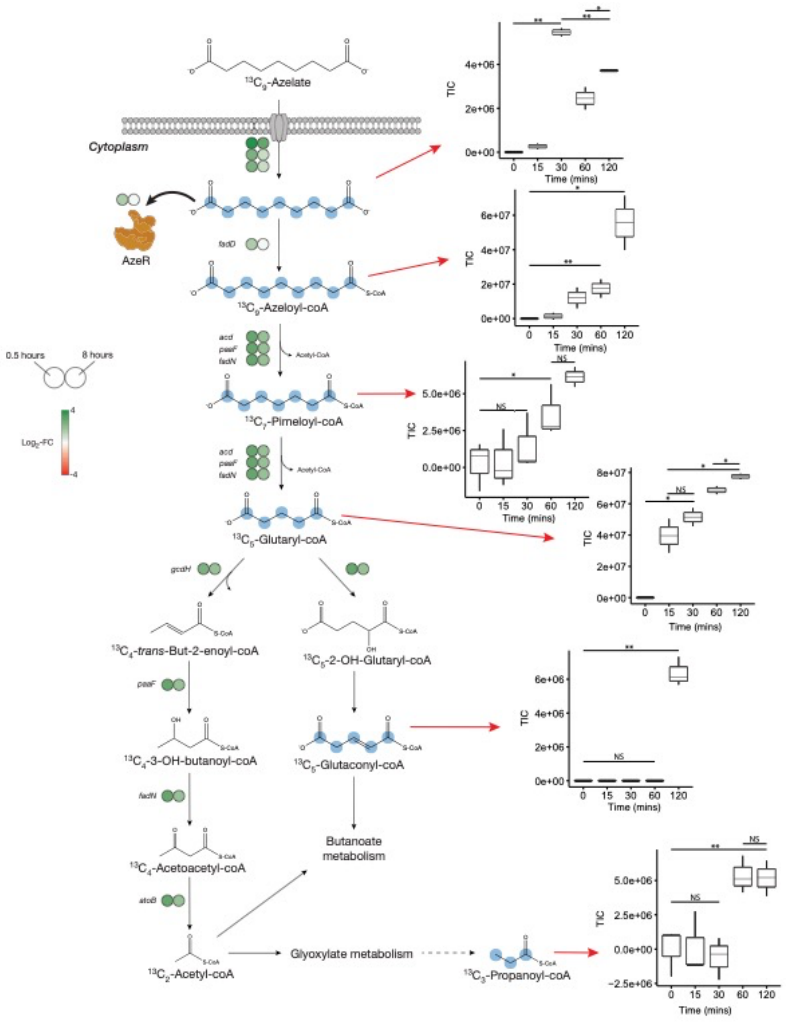


Figure 2. Proposed Azelaic acid catabolism in *Phaeobacter based on transcriptomics and isotope labeling*. Left: Log_2 fold-change of differentially expressed (DE) genes is shown as circles at 0.5 (left circle) and 8 hours (right circle) after the addition of Aze to *Phaeobacter* cells relative to controls. The Aze transport system (*azeTSL*) consists of 3 genes, the DE values of which are shown next to the transporter. Intermediate reactions for the successive liberation of acetyl-coA from azeloyl-coA and pimeloyl-coA are not shown; DE values of the genes involved are shown next to each overall reaction. ^{13}C-labeled metabolites detected in the intracellular metabolome of *Phaeobacter* cells are marked with cyan circles at each labeled carbon atom site. Right: Relative abundance of each detected labeled metabolite after addition of 100 µM ^{13}C-Aze to *Phaeobacter* cells, shown as total ion current (TIC). Box plot values are based on triplicates. NS indicates no significant relative abundance, * denotes $p<0.05$ and ** denotes $p<0.005$ based on a Student's t-test.

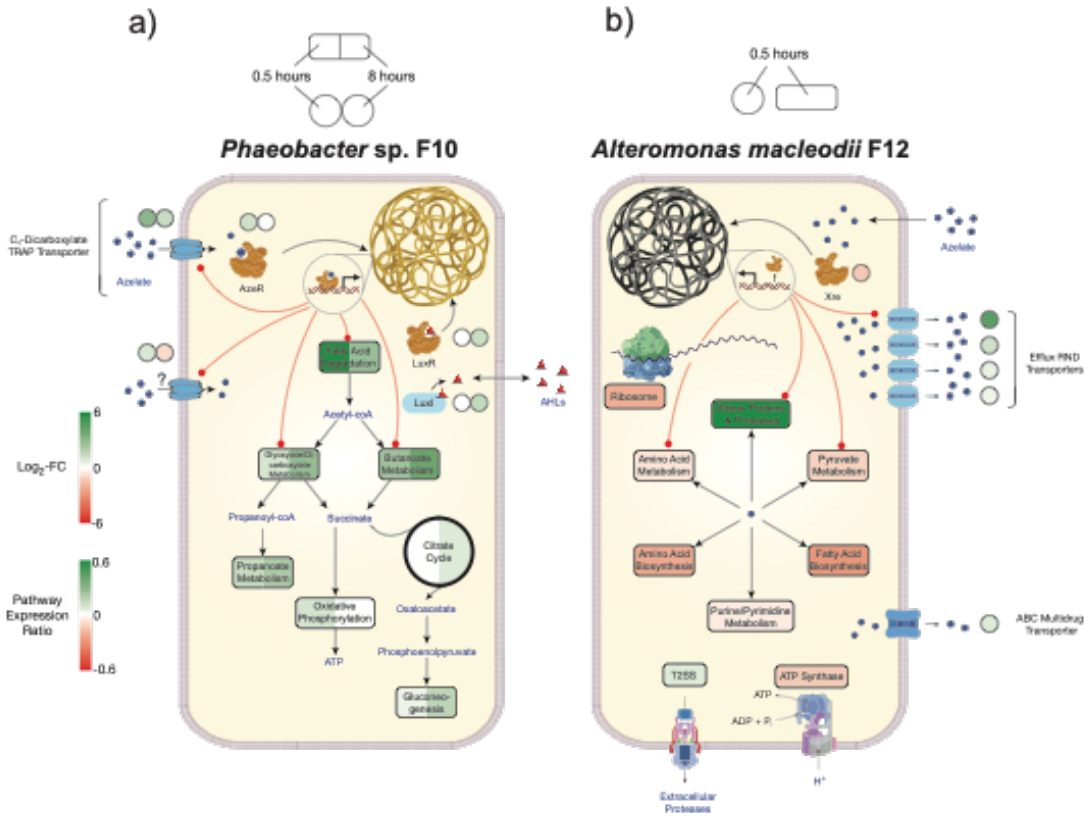


Figure 3. Model of Aze catabolism in (a) *Phaeobacter* and toxic effects in (b) *A. macleodii*. Colored boxes and circles represent metabolic pathway expression ratios and differential gene expression (log_2-FC), respectively, at 0.5 and 8 hours for *Phaeobacter* and at 0.5 hours for *A. macleodii.* White circles/boxes indicate no statistically significant DE genes/silent pathway. Putative transcriptional factor regulation of pathways/genes is shown by red lines. Log_2-fold change values shown for all transporters represent the mean log_2-FC value of all genes in the cluster. Pathway expression ratios were the same as in Fig. 1 except the sign of the ratio (indicating up- or downregulation) are indicated. Amino acid biosynthesis and metabolism ratios in (b) are outside the range of the pathway expression ratio scale (values ~1.5) but are colored in dark red to indicate significant downregulation. T2SS=Type II secretion system, AHLs=acyl-homoserine lactones.

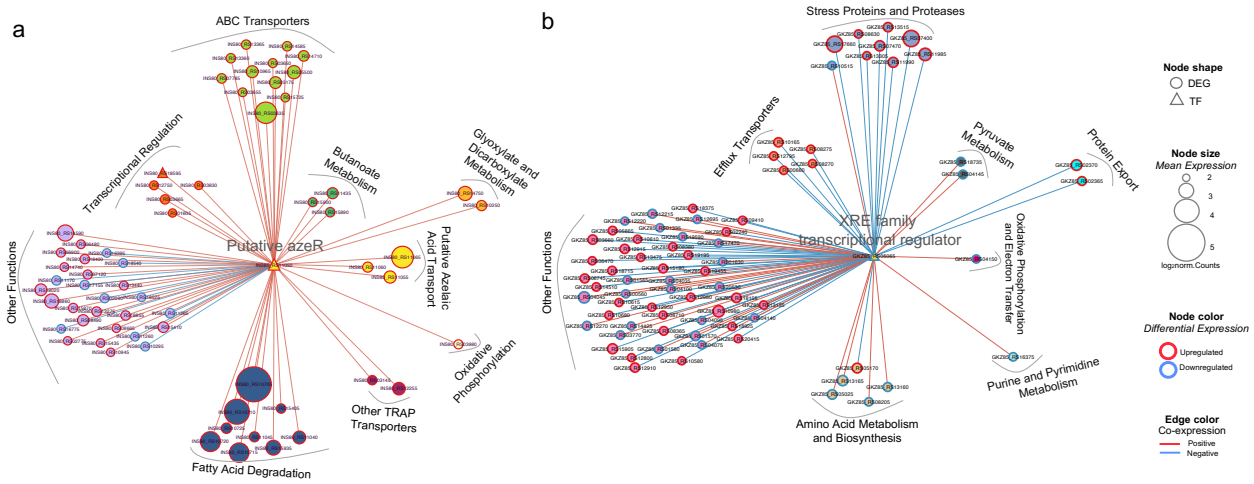


Figure 4. Transcriptional coexpression networks of hub transcriptional factors (TFs) *azeR* and *xre*, and their putative target genes in (a) *Phaeobacter* and (b) *A. macleodii.* Nodes represent genes (circles) and TFs (triangles) connected by edges based on significant co-expression correlation (PCIT; $r \geq \pm 0.95$). Nodes are grouped based on functions and represented by different colors. The size of the node corresponds to the normalized mean expression values in Aze-treated samples, whereas the color of the node border corresponds to differential expression. Edge colors indicate the direction of the correlation between each gene pair. DEG = differentially expressed gene, TF = transcriptional factor.

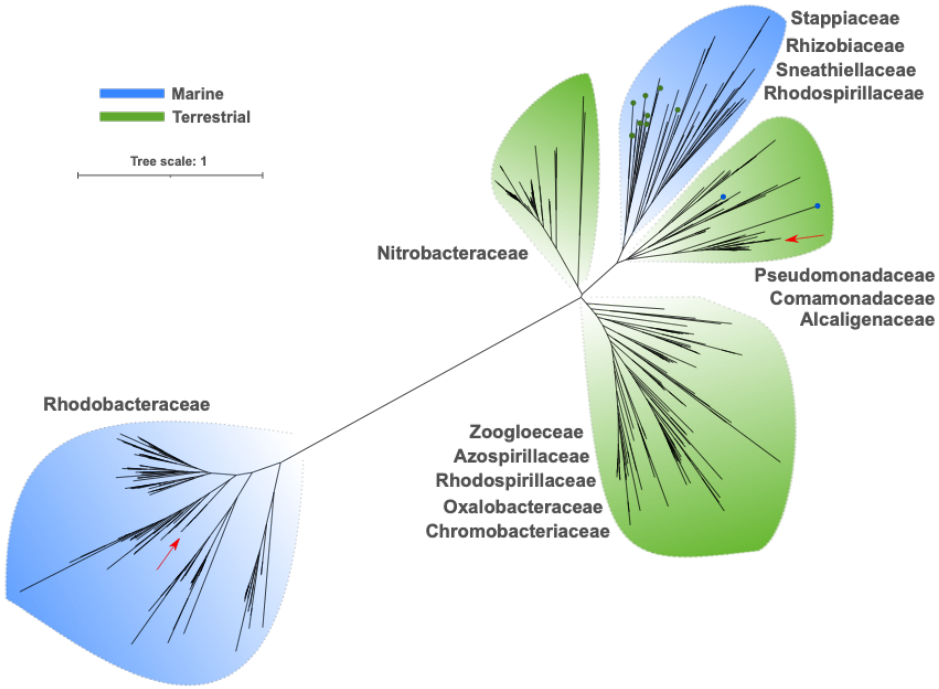
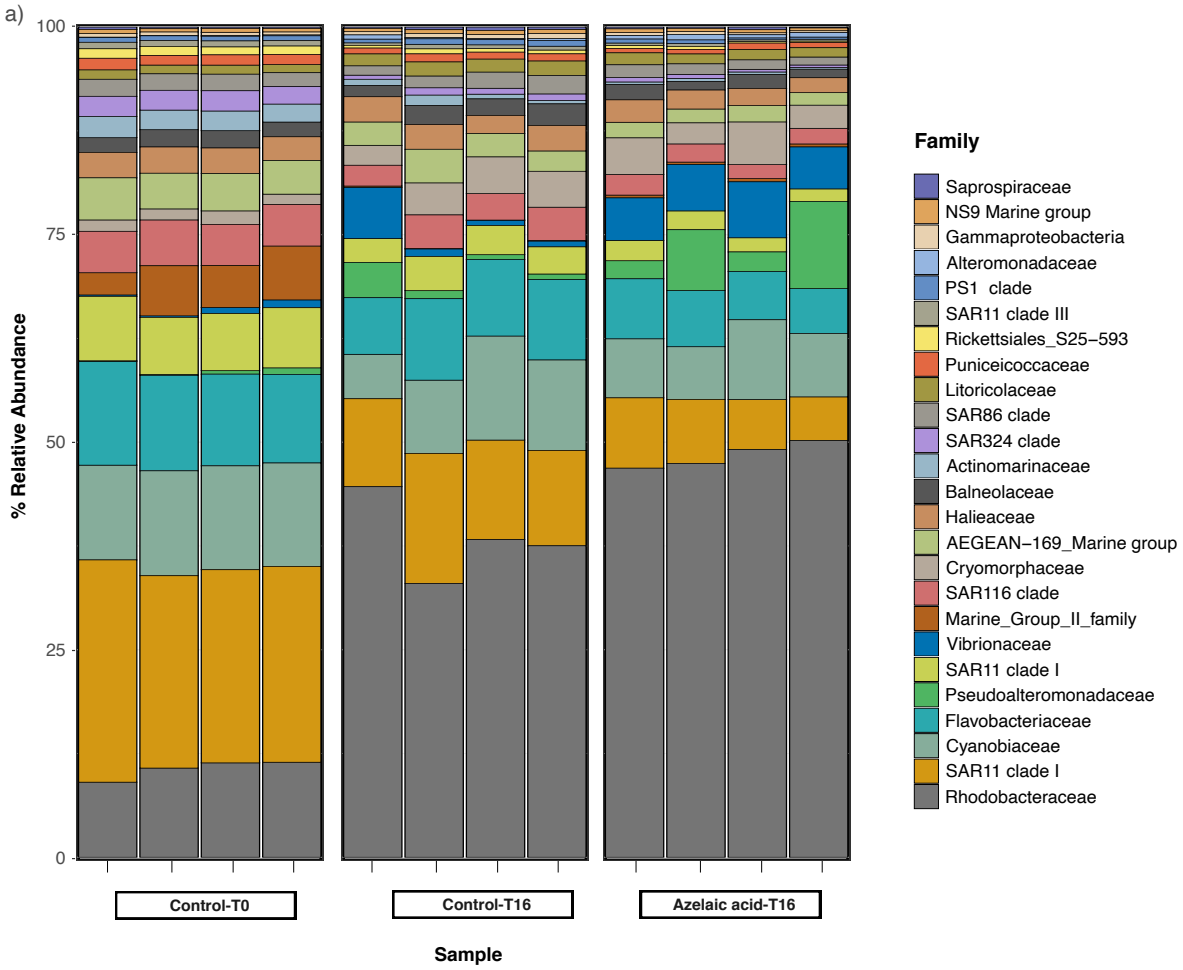


Figure 5. Phylogeny of *azeR* and *azeT* across marine and terrestrial bacterial genomes. The phylogenetic tree was inferred using concatenated amino acid sequences of AzeR and AzeT (C_4-dicarboxylate TRAP transporter substrate-binding protein) homologs extracted from *Phaeobacter* and *Pseudomonas nitroreducens*. Branches are colored according to the habitat from which the taxa were isolated and annotated according to the most dominant families. Green circles within the marine branch indicate terrestrial taxa, while blue circles within the terrestrial branches represent marine taxa. Red arrows in the marine and terrestrial branches indicate *Phaeobacter* and *Pseudomonas nitroreducens* sequences, respectively.



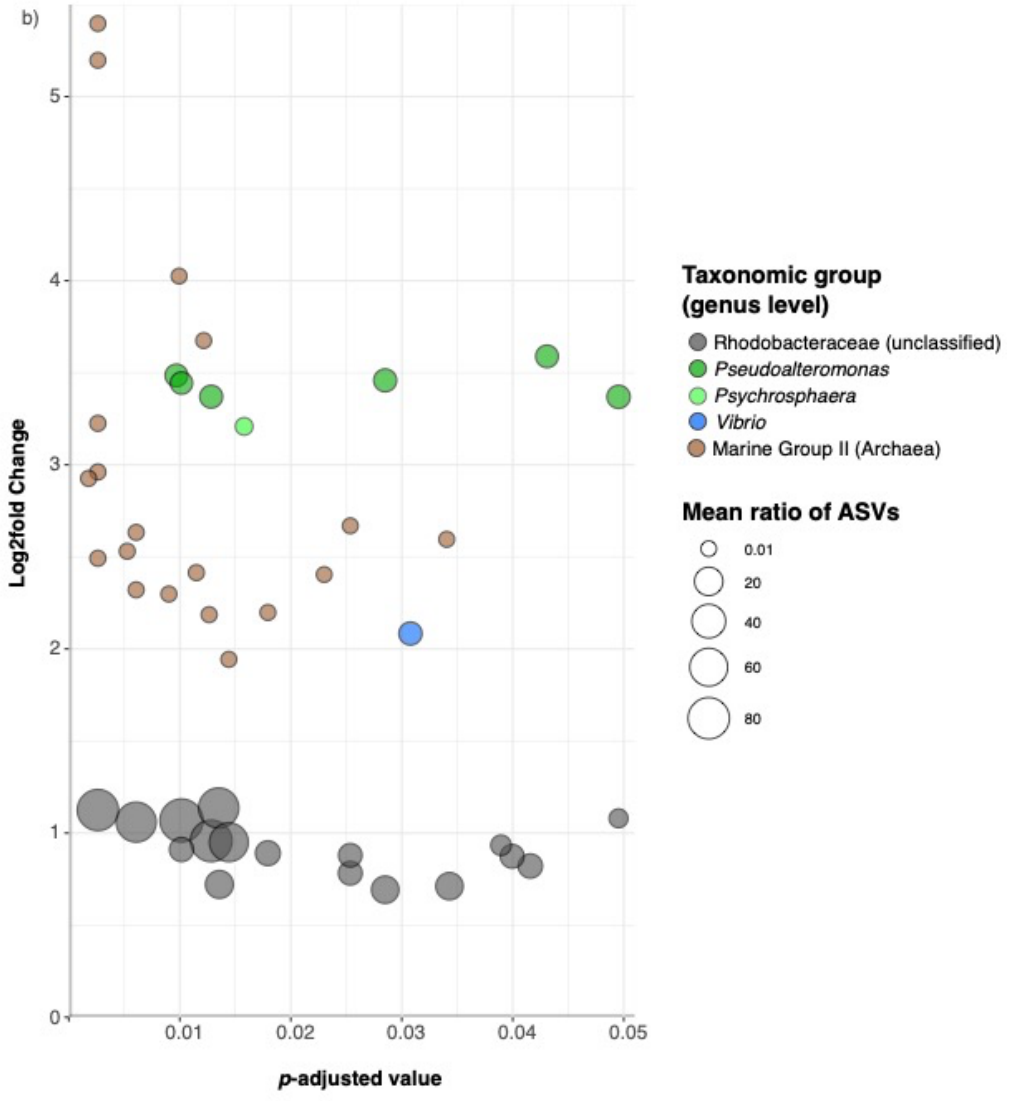
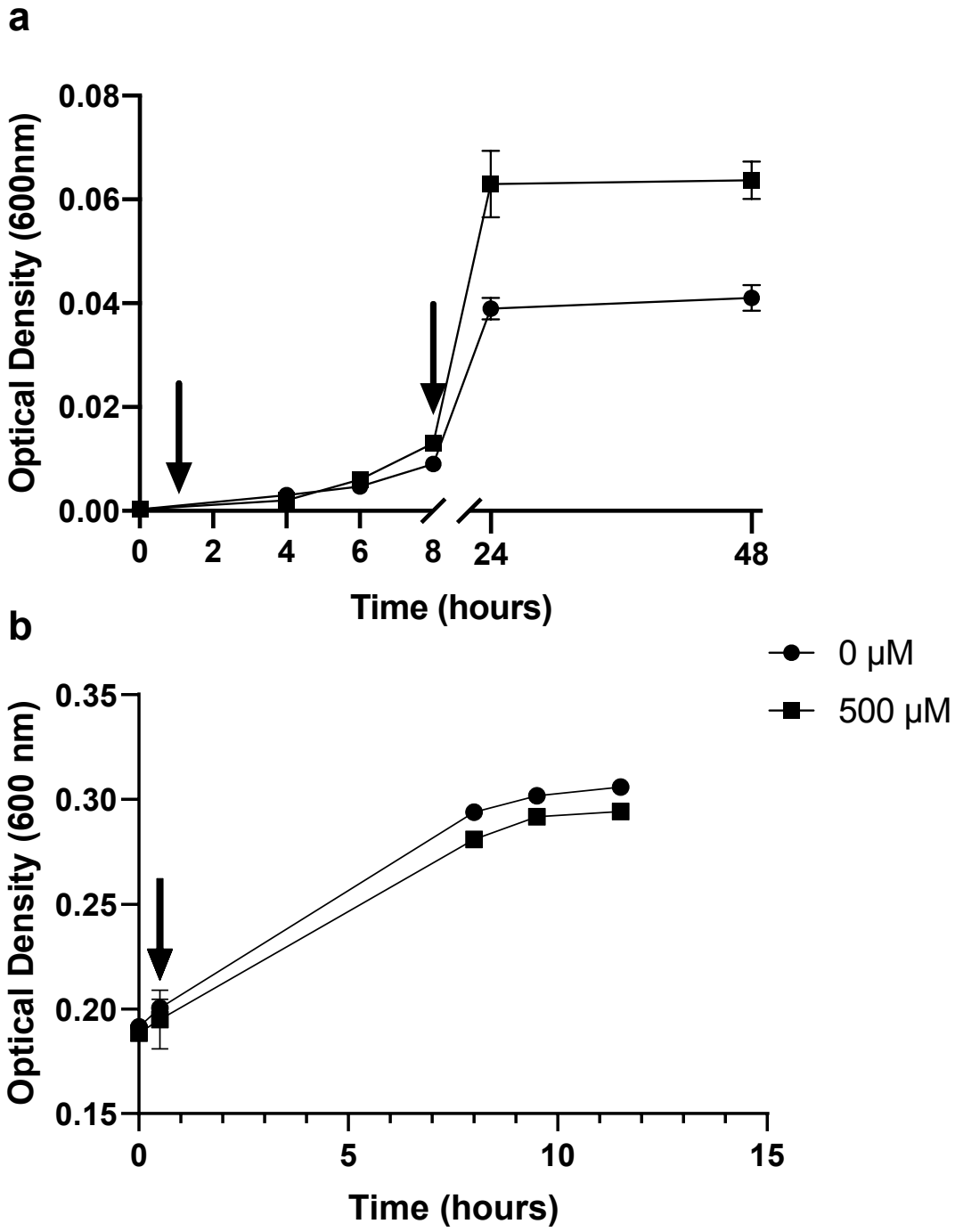


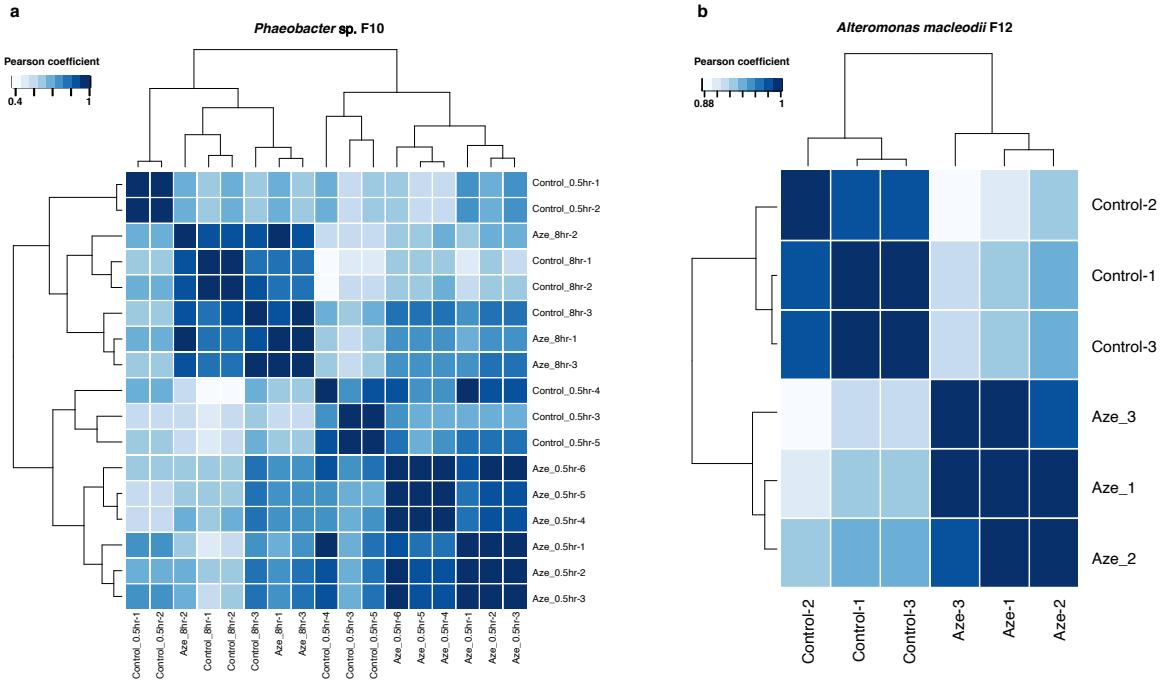
Figure 6. Bacterial and archaeal diversity in seawater samples treated with Aze. (a) Relative abundance of the top 25 microbial families based on 16S rRNA amplicon sequencing of Aze-treated samples at T=16 hours and control samples at T=0 and T=16 hours. **(b)** Distribution of the amplicon sequence variants (ASVs) belonging to significantly differentially abundant taxa between the Aze-treated and control samples at T=16 hours, according to their $\log_2$ fold change and *p*-adjusted values. The bubble size indicates the mean taxonomic proportion of each ASV, calculated as the mean number of reads in an ASV : mean number of reads present in all ASVs of the same taxonomic classification. The bubble color indicates the taxonomic classification of each ASV according to (a).

Extended Data Figures and Legends



Extended Data Figure 1. Effect of Aze on growth of (a) *Phaeobacter* and (b) *A. macleodii*. 500 µM Aze was added at T=0 hours to treatments while controls received an equivalent volume of Milli-Q water. RNA samples were collected at the times marked by arrows. Error bars represent standard deviation (s.d.) of triplicate cultures.

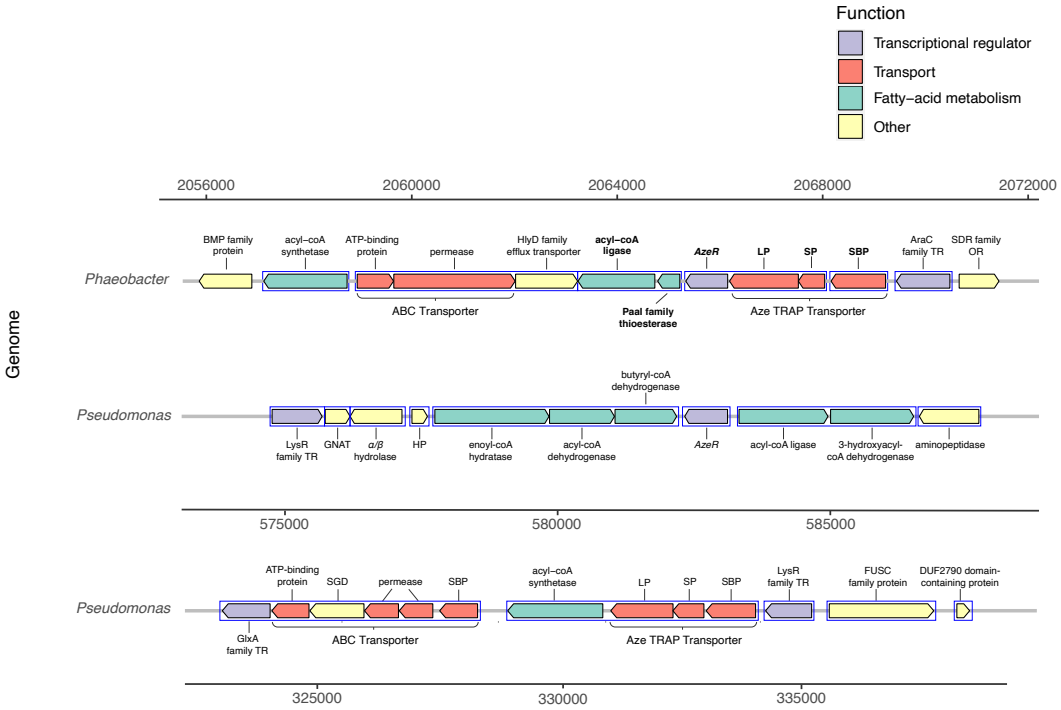
710



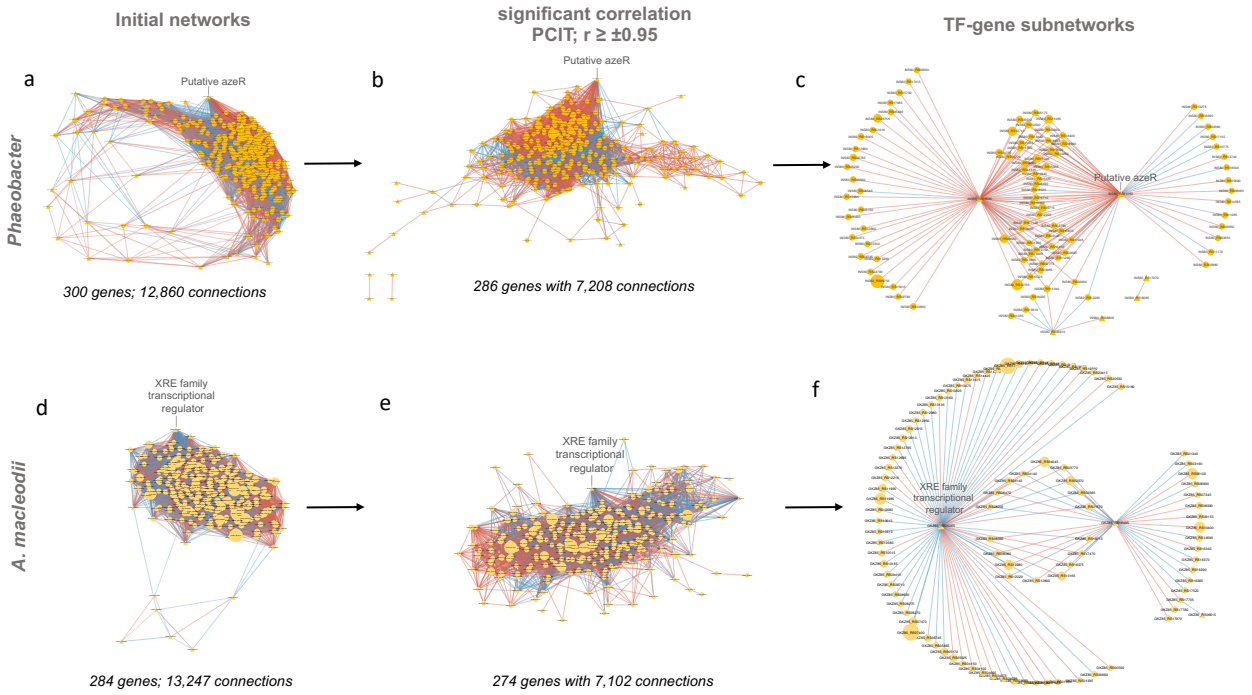
711

712 **Extended Data Figure 2. Comparison of RNASeq data after treatment with Aze relative to**
713 **controls.** Dendrograms indicate the hierarchical clustering of all (a) *Phaeobacter* and (b) *A.*
714 *macleodii* differentially expressed genes based on RNASeq samples. The color gradient refers to
715 the degree of correlation across the samples based on Pearson coefficient.

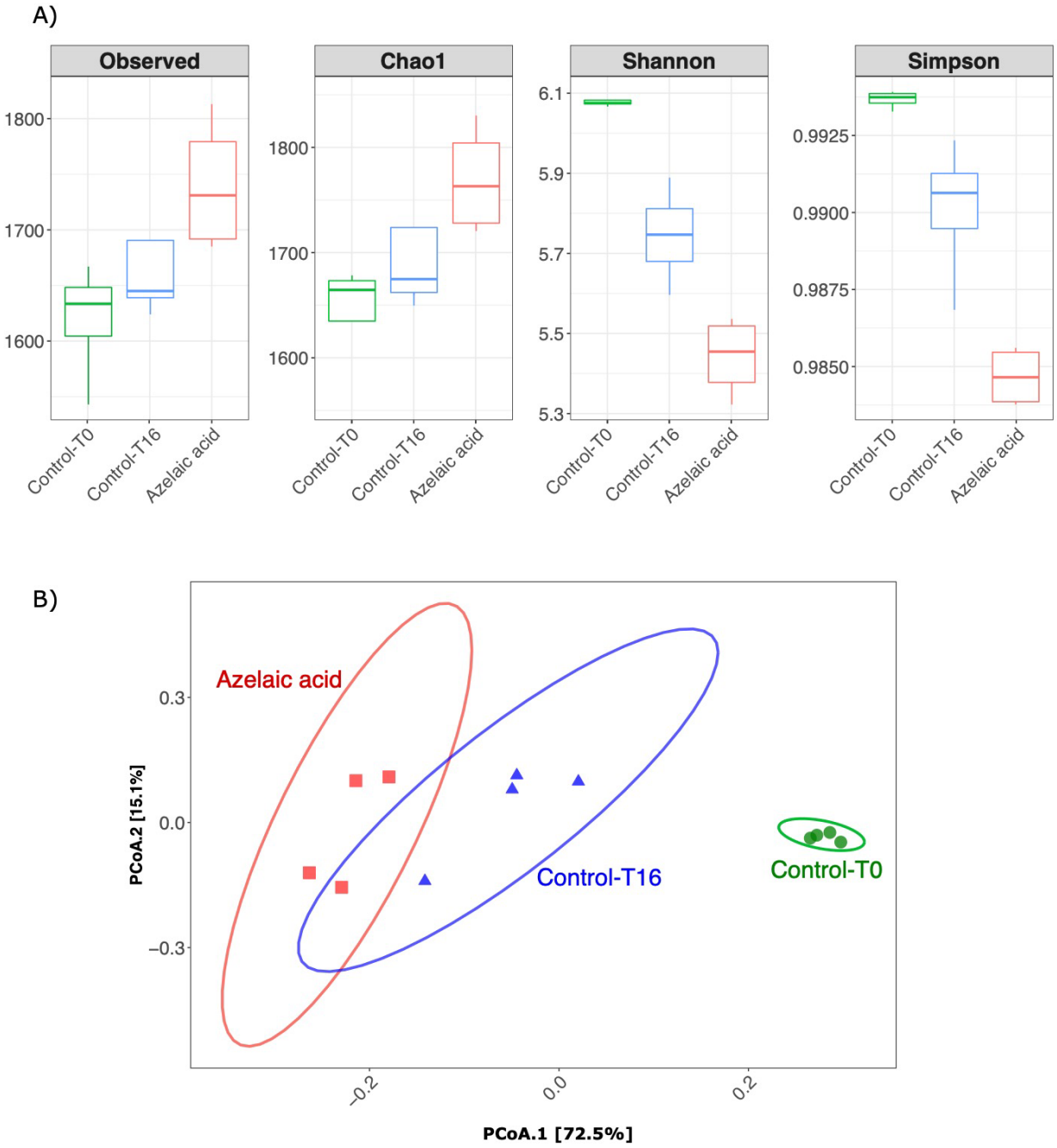
716



Extended Data Figure 3. Genomic neighborhood structure of *azeR* and *azeTSL* in *Phaeobacter* and *Pseudomonas nitroreducens* DSM 9128. Genes predicted *in silico* to belong to a single operon are marked by blue boxes. Bold-faced gene abbreviations in *Phaeobacter* denote upregulated genes in response to Aze. Operon prediction was carried out using OperonMapper (Taboada et al., 2018). *azeT* = substrate-binding protein, *azeS* = small permease, *azeL* = large permease, *azeR* = transcriptional regulator, TR = transcriptional regulator, OR = oxidoreductase, HP = hypothetical protein, GNAT = GNAT family N-acetyltransferase, SGD = succinylglutamate desuccinylase, SBP = substrate-binding protein.



726 **Extended Data Figure 4.** Transcriptional coexpression networks constructed using the Partial
727 Correlation coefficient with Information Theory algorithm in *Phaeobacter* and *A. macleodii*. Initial
728 networks (a, d) consisted of key transcriptional factors (TFs) (identified by regulatory impact factor
729 analysis) and differentially expressed genes. Nodes are depicted either as circles for genes or
730 triangles for TFs. Edges represent significant interactions between nodes. Edge color represents
731 directions of the interaction (red for positive correlation and blue for negative correlation). The
732 size of the node corresponds to the normalized mean expression values in Aze-treated samples.
733 Subnetworks were extracted from initial networks based on significant co-expression correlation
734 (PCIT; $r \geq \pm 0.95$) (b, e) and those containing only hub genes (identified based on RIF scores,
735 differential expression, and the degree centrality) and their connected genes (c, f).
736



737 **Extended Data Figure 5. Effect of Aze treatment on the alpha and beta-diversity of seawater**
738 **mesocosms. (a)** Alpha-diversity indices of observed OTUs, Chao1, Shannon and Simpson
739 across the Aze-treated and control samples. **(b)** PCoA of Bray-Curtis distances (PERMANOVA:
740 R^2 = 0.73; p < 0.001) between samples. The two principal components (PCoA1 and PCoA2)
741 explained 72.5% and 15.1% variance, respectively.

742