environmental microbiology

Environmental Microbiology (2022) **24**(7), 3195–3211 doi:10.1111/1462-2920.16074

Complex and flexible catabolism in *Aromatoleum aromaticum* pCyN1

Patrick Becker,[1] Annemieke Döhmann,[1] Lars Wöhlbrand,[1] Daniela Thies,[2] Christina Hinrichs,[1] Ramona Buschen,[1] Daniel Wünsch,[1] Meina Neumann-Schaal,[3,4] Dietmar Schomburg,[3,5] Michael Winklhofer,[6,7] Richard Reinhardt[8] and Ralf Rabus[1*]

[1]*General and Molecular Microbiology, Institute for Chemistry and Biology of the Marine Environment (ICBM), Carl von Ossietzky University of Oldenburg, Oldenburg, Germany.*

[2]*Department of Microbiology, Max Planck Institute for Marine Microbiology, Bremen, Germany.*

[3]*Research Group Bacterial Metabolism, Braunschweig Integrated Centre of Systems Biology (BRICS), Technische Universität Carolo-Wilhelmina zu Braunschweig, Braunschweig, Germany.*

[4]*Department of Analytics, Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany.*

[5]*Department of Bioinformatics and Biochemistry, Institute for Biochemistry, Biotechnology and Bioinformatics, Technische Universität Carolo-Wilhelmina zu Braunschweig, Braunschweig, Germany.*

[6]*Research Center Neurosensory Science, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany.*

[7]*Sensory Biology of Animals, Institute of Biology and Environmental Sciences (IBU), Carl von Ossietzky University of Oldenburg, Oldenburg, Germany.*

[8]*Max-Planck-Genome-Centre Cologne, Max Planck Institute for Plant Breeding Research, Cologne, Germany.*

Summary

Large quantities of organic matter are continuously deposited, and (a)biotic gradients intersect in the soil–rhizosphere, where biodegradation contributes to the global cycles of elements. The betaproteobacterial genus *Aromatoleum* comprises cosmopolitan, facultative denitrifying degradation specialists. *Aromatoleum aromaticum*. pCyN1 stands out for anaerobically decomposing plant-derived monoterpenes in addition to monoaromatic hydrocarbons, polar aromatics and aliphatics. The catabolic network's structure and flexibility in *A. aromaticum* pCyN1 were studied across 34 growth conditions by superimposing proteome profiles onto the manually annotated 4.37 Mbp genome. Strain pCyN1 employs three fundamentally different enzymes for C–H-bond cleavage at the methyl groups of *p*-cymene/4-ethyltoluene, toluene and *p*-cresol respectively. Regulation of degradation modules displayed substrate specificities ranging from narrow (toluene and cyclohexane carboxylate) via medium-wide (one module shared by *p*-cymene, 4-ethyltoluene, α-phellandrene, α-terpinene, γ-terpinene and limonene) to broad (central benzoyl-CoA pathway serving 16 aromatic substrates). Remarkably, three variants of ATP-dependent (class I) benzoyl-CoA reductase and four different β-oxidation routes establish a degradation hub that accommodates the substrate diversity. The respiratory system displayed several conspicuous profiles, e.g. the presence of nitrous oxide reductase under oxic and of low-affinity oxidase under anoxic conditions. Overall, nutritional versatility in conjunction with network regulation endow *A. aromaticum* pCyN1 with broad adaptability.

Introduction

The soil–rhizosphere is a complex ecosystem, harbouring approx. 1500 Gt of organic carbon deposit (Crowther *et al*., 2019) and 8 Gt of prokaryotic organisms (~90% bacteria) (Bar-On *et al*., 2018), with an estimate of 0.1–1 mg bacterial biomass per gram of soil (Fierer, 2017). An imbalance between mineralization of organic matter and biomass formation by soil microbiota increases the net flux of greenhouse gases (CO_2, CH_4, N_2O) (van Groenigen *et al*., 2011; Shcherbak *et al*., 2014; Bond-Lamberty *et al*., 2018); this effect is accelerated by a temperature increase, forming a positive feedback loop. However, the potential of the soil microbiota to mitigate adverse effects of climate change is also possible (Jansson and Hofmockel, 2020). Components of

Received 8 February, 2022; revised 14 May, 2022; accepted 16 May, 2022. *For correspondence. E-mail rabus@icbm.de; Tel. +494417983884; Fax +494417983404.

organic matter have highly diverse structures and differ greatly with respect to biodegradability, ranging from easily degradable biomacromolecules and fermentation end products to biochemically challenging aromatic compounds and (mono)terpenes. Another major factor controlling the decomposition of organic matter is the redox zonation, i.e. the availability of O_2 or alternative electron acceptors (in particular NO_3^-) for the conservation of respiratory energy.

Microorganisms from various phyla and kingdoms populate the soil–rhizosphere system and provide various services to the plants (for overview see, e.g. Philippot *et al*., 2013; Bulgarelli *et al*., 2013; Rolli *et al*., 2021; Herms *et al*., 2022). Recurrently encountered constituents of the soil microbiota are aerobic carbohydrate-utilizing members of the phylum Verrucomicrobia, e.g. '*Candidatus* Udaeobacter copiosus' (Brewer *et al*., 2016) and aerobic degradation specialists of the families *Burkholderiaceae*, e.g. *Burkholderia xenovorans* (Chain *et al*., 2006) and *Pseudomonadaceae*, e.g. *Pseudomonas putida* (Nelson *et al*., 2002). Members of the phylum Proteobacteria are often observed in anoxic zones (Fierer, 2017). These stand out through their ability to decompose wood, perform denitrification and fix nitrogen (Coyotzi *et al*., 2017; Crowther *et al*., 2019). They can also perform the anaerobic degradation of recalcitrant organic molecules (Rabus *et al*., 2016).

Aromatic compounds and particularly (mono)terpenes are abundant, structurally highly diverse, organic molecules in the soil–rhizosphere. Many form the building blocks of naturally occurring biomacromolecules (in particular lignin) (de Leeuw *et al*., 2006) and of insoluble macromolecular organic matter (kerogen) (Vandenbroucke and Largeau, 2007). Furthermore, monoaromatic compounds represent high-production-volume chemicals, used for the industrial production of a wide variety of commodities. Furthermore, these compounds, in particular, the alkylbenzenes and -phenols, are often of environmental concern due to their toxicology (Sikkema *et al*., 1995; ATSDR, 2008). Terpenoids are mainly produced by plants. They exhibit a large structural diversity, serving multiple roles, including as phytohormones or anti-oxidants (Pichersky and Raguso, 2018). Notably, terpenoids are also synthesized by prokaryotes and fungi (Boronat and Rodríguez-Concepcíon, 2015).

Aromatic and terpenoid hydrocarbons are highly reduced molecules and hence energy-rich. Consequently, they are attractive growth substrates for bacteria, albeit their biodegradation presents challenges due to their apolar character (Wilkes and Schwarzbauer, 2010). Activation of these inert compounds is required to enable their biodegradation. For this purpose, aerobic microorganisms use highly reactive oxygen species (O_2-derived) as co-substrate via oxygenase-based catalysis (Gibson and Harwood, 2002; Marmulla and Harder, 2014). In the absence of molecular oxygen (anoxic conditions), the thermodynamic challenge is considerable. Nevertheless, a broad variety of anaerobic bacteria employs a range of intriguing biochemical reactions for the O_2-independent activation and reductive dearomatization of these hydrocarbons (for overview, refer to, e.g. Heider and Fuchs, 1997; Carmona *et al*., 2009; Fuchs *et al*., 2011; Boll *et al*., 2014; Marmulla and Harder, 2014; Rabus *et al*., 2016).

Aromatoleum aromaticum pCyN1 anaerobically degrades a wide range of plant-derived monomeric compounds (monoterpenes, phenylpropanoids), aromatic compounds of proteinaceous- (e.g. phenylalanine) and potential anthropogenic origin (e.g. toluene, *p*-cresol). It can also switch between denitrification and O_2 respiration (Harms *et al*., 1999; Rabus *et al*., 2019). The remarkable ability to anaerobically degrade *p*-cymene requires an initial anaerobic hydroxylation reaction, yielding 4-isopropylbenzyl alcohol (Strijkstra *et al*., 2014), as well as a hybrid pathway of dearomatization and downstream β-oxidation (Küppers *et al*., 2019). The present study provides a broad survey of the architecture and substrate-/redox-dependent regulation of this strain's catabolic network. For this purpose, the genome of *A. aromaticum* pCyN1 was sequenced, manually annotated and combined with comprehensive proteomic profiling across 34 different substrate adaptation conditions. This enabled retracing the background of its catabolic adaptability to changing environmental conditions as encountered in dynamic habitats such as the soil–rhizosphere.

Experimental procedures

Biogeography, habitats and diversity

The global distribution of isolates and phylotypes belonging to the *Aromatoleum*/*Azoarcus*/*Thauera* cluster was assessed by means of literature searches (as of March 2020) and cartographically visualized on the basis of manually retrieved GPS coordinates of the sample locations and using the cartopy package for the python programming language (https://scitools.org.uk/cartopy). Sites were clustered according to the Density-Based Spatial Clustering of Applications with the Noise algorithm implemented in the python scikit-learn package, using great circle distances between sites as a metric and 490 km as the maximum distance between two sites for one to be considered as in the neighbourhood of the other. The plate carrée projection (equidistant cylindrical projection) was used for presenting the worldwide distribution of clusters. Detailed information on the compiled literature data is provided in Tables S1 and S2.

Bacterial strains, media, cultivation conditions and cell harvesting

Aromatoleum aromaticum pCyN1 was originally isolated with *p*-cymene under nitrate-reducing conditions from freshwater mud (Harms *et al*., 1999), then maintained in the laboratory and recently deposited to the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ; Braunschweig, Germany) under the culture collection number DSM 19016 (Rabus *et al*., 2019). *Aromatoleum aromaticum* pCyN1 was cultivated under nitrate-reducing conditions in a defined, ascorbate-reduced and bicarbonate-buffered mineral medium as well as under oxic conditions in the same mineral medium at 28°C, as previously described (Rabus and Widdel, 1995). Soluble substrates were added from aqueous stock solutions sterilized by filtration, while hydrophobic substrates were provided as dilutions in an inert carrier phase (2,2,4,4,6,8,8-heptamethylnonane). The following 30 growth substrates were used (order, numbering, and abbreviations as in Figs 2, 4 and 5; concentration is given in parenthesis for anaerobic growth if not indicated otherwise): 1, *p*-cymene, pCym (5%, vol./vol.); 2, 4-ethyltoluene, 4ETol (2%, vol./vol.); 3, α-phellandrene, αPhel (2%, vol./vol.); 4, α-terpinene, αTerp (2%, vol./vol.); 5, γ-terpinene, γTerp (2%, vol./vol.); 6, limonene, Lim (2%, vol./vol.); 7, toluene, Tol (2%, vol./vol.); 8, benzyl alcohol, BzOH (2 mM); 9, benzaldehyde, BzAl (6 mM); 10, benzoate, Bz (4 mM); 11, phenylacetate, PAc (4 mM, anaerobic; 2 mM, aerobic); 12, phenylalanine, Phe (4 mM, anaerobic; 2 mM, aerobic); 13, tyrosine, Tyr (2 mM); 14, *p*-cresol, pCr (2 mM); 15, 4-hydroxybenzoate, 4HBz (4 mM, anaerobic; 2 mM, aerobic); 16, 2-aminobenzoate, 2ABz (6 mM); 17, 3-hydroxybenzoate, 3HBz (2 mM); 18, hydrocinnamate, HCin (2 mM); 19, 3-(4-hydroxyphenyl)propanoate, 4HPP (2 mM); 20, cinnamate, Cin (2 mM); 21, *p*-coumarate, pCou (2 mM); 22, indoleacetate, IAA (2 mM); 23, cyclohexane carboxylate, CHC (2 mM); 24, acetone, Ace (5 mM); 25, ethanol, EtOH (8 mM); 26, acetate, Ac (8 mM); 27, malate, Mal (5 mM); 28, succinate, Suc (8 mM); 29, D/L-lactate, Lac (10 mM); and 30, pyruvate, Pyr (10 mM). All chemicals were of analytical grade.

Adaptation of *A. aromaticum* pCyN1 to each growth condition was conducted over at least five passages (80 ml culture volume). To provide sufficient cell material for proteomic profiling, cultivation was performed in 500-ml flat-bottomed glass bottles (400 ml culture volume) sealed with butyl rubber stoppers. At least six parallel cultures were run per substrate condition.

Harvesting of *A. aromaticum* pCyN1 cultures at half-maximal optical density was performed as previously described (Koßmehl *et al*., 2013). Essentially, the complete 400-ml cultures were centrifuged (14 334*g*, 30 min, 4°C), the pellets were washed in 250 ml washing buffer (100 mM Tris/HCl, 5 mM $MgCl_2 \times 6\ H_2O$, pH 7.5) and resuspended in 0.8 ml of the same washing buffer. Following further centrifugation (20 000*g*, 10 min, 4°C), pellets were shock frozen in liquid N_2 and stored at −80°C until further analysis.

DNA sequencing, assembly and annotation

Isolation of genomic DNA was carried out using the Genomic DNA kit (Qiagen, Hildesheim, Germany) according to the manufacturer's instructions. Recombinant plasmid and fosmid shotgun libraries were constructed. Plasmid libraries were generated from sonified DNA (Rabus *et al*., 2005). Additionally, a fosmid library was constructed (>40-fold physical coverage) for data finishing and assembly confirmation (Epicentre Technologies, Madison, WI, USA). Templates for sequencing were obtained by insert amplification via PCR or by plasmid isolation. Sequencing was carried out using ABI3730XL capillary systems (Thermo Fisher Scientific, Waltham, MA, USA). PHRAP [Phragment assembly program 1999 (http://www.phrap.org/phredphrapconsed.html)] and Consed (Gordon, 2003) were used to assess sequence quality and perform the assembly with a quality of <1 error in 100 000 bases.

Structural rRNAs and tRNAs were determined using RNAmmer and tRNAscan-SE. Protein-coding sequences (CDS) were predicted by the ORF-finding program Glimmer3 and manually revised and curated using Artemis (v.12.0) and InterPro. The ORF dataset generated was screened against non-redundant protein databases (SWISSPROT and TREMBL). Genomic islands and islets (<10 kbp) were predicted by applying IslandViewer 3. The genome was screened for phage-like regions by PHASTER, and a CRISPR recognition tool (CRT) served in detection of CRISPR sequences. The genome data of *A. aromaticum* $EbN1^T$ used for genomic comparison with *A. aromaticum* pCyN1 were obtained from our previous publication (Rabus *et al*., 2005). Deciphering the transporter complement rested on a previous study using *A. aromaticum* $EbN1^T$ (Tamang *et al*., 2009) and was performed by consulting the transporter classification database. References for the bioinformatics tools used are provided in Table S7.

Sequence accession numbers

The genome sequence of *A. aromaticum* pCyN1 has been submitted to GenBank under the BioProject PRJNA802808 with BioSample SAMN25597984 and accession number CP091977.

Profiling of soluble proteins by 2D DIGE and protein identification by MALDI-TOF/TOF-MS

Soluble proteins were extracted from *A. aromaticum* pCyN1 and 2D DIGE conducted essentially as previously described (Gade *et al*., 2003). Per substrate condition,

three cell pellets (~100 mg wet weight) representing biological replicates were suspended in lysis buffer (7 M urea, 2 M thiourea, 30 mM Tris/HCl, 4% CHAPS, pH 8.5) to account for biological variation (Zech *et al*., 2011). Following cell breakage using the PlusOne sample grinding kit (GE Healthcare, Munich, Germany), the protein concentration was determined according to the method of Bradford (1976). For minimal labelling, 200 pmol of Lightning SciDye DIGE fluors (SERVA, Heidelberg, Germany) were used per 50 μg of protein sample. Protein extracts of acetate-adapted cells cultivated under anoxic conditions served as the reference state and were labelled with Sci5. Protein extracts from the other 34 (30 anaerobic and four aerobic) substrate adaptation conditions represented the test states and were each labelled with Sci3. The internal standard contained equal amounts of all respective test (anaerobic or aerobic) and reference state(s) and was labelled with Sci2. Per gel, 50 μg each of the labelled reference state, test state and internal standard were applied, and three biological replicates per test state were performed. First dimension separation by isoelectric focusing (IEF) was conducted with 24-cm long IPG strips (pH 3–11 NL; GE Healthcare) run in a Protean i12 system (Bio-Rad, Munich, Germany). The IEF program used was as follows: 50 V for 13 h, 200 V for 1 h, 1000 V for 1 h, gradual gradient to 10 000 V within 2 h and 10 000 V for 6 h. Second dimension separation of proteins according to molecular size was achieved by SDS-PAGE (12.5% gels, vol./vol.) using an Ettan Dalt Twelve system (GE Healthcare).

2D DIGE gels were digitalised directly after completion of electrophoresis with a Typhoon 9400 scanner (GE Healthcare). Cropped gel images were analysed with the DeCyder software (version 5.0; GE Healthcare) allocated to five different work packages: (i) toluene, 4-ethyltoluene, *p*-cresol, benzaldehyde and benzyl alcohol (all anaerobic); (ii) phenylacetate, phenylalanine, tyrosine, 2-aminobenzoate, 3-hydroxybenzoate, 4-hydroxybenzoate, hydrocinnamate and (3-(4-hydroxyphenyl)propanoate (all anaerobic); (iii) α-phellandrene, α-terpinene, γ-terpinene, limonene, *p*-cymene and cyclohexane carboxylate (all anaerobic); (iv) lactate, pyruvate, malate, succinate, acetone, and ethanol (all anaerobic); and (v) benzoate, 3-hydroxybenzoate, phenylalanine, phenylacetate and indoleacetate (all aerobic except for indoleacetate). All of these five work packages included the anaerobic adaptation conditions benzoate and acetate as the anchor and reference states respectively. Adjustments at the scanner and parameters for spot detection were as described previously (Wöhlbrand *et al*., 2007). All three biological replicates were included for the reference and for each test state. Changes in the protein abundance of ≥|1.5|-fold were regarded as significant (Zech *et al*., 2011). Separate preparative colloidal Coomassie Brilliant Blue (cCBB)-stained gels were run (300 μg protein load) to obtain sufficient amounts of protein for reliable mass spectrometric identification (Wöhlbrand *et al*., 2007). Spots of interest were excised using the EXQuest spot cutter (Bio-Rad) from two cCBB-stained gels per analysed substrate state; these were subsequently washed and tryptically digested as previously described (Wöhlbrand *et al*., 2007).

Sample digests were spotted onto Anchorchip steel targets (Bruker Daltonik GmbH, Bremen, Germany) and analysed with an UltrafleXtreme MALDI-TOF/TOF mass spectrometer (Bruker Daltonik GmbH) as previously described (Zech *et al*., 2013). Peptide mass fingerprint (PMF) searches were performed with a Mascot server (version 2.3; Matrix Science, London, UK) against the translated genome of *A. aromaticum* pCyN1, with a mass tolerance of 25 ppm. Five lift spectra were collected to confirm PMF identification and three additional spectra were acquired of unassigned peaks applying feedback by the ProteinScape platform (version 3.1; Bruker Daltonik GmbH). In case of failed PMF identification, eight lift spectra of suitable precursors were acquired. MS/MS searches were performed with a mass tolerance of 100 ppm. For both, MS and MS/MS searches, Mascot scores not meeting the 95% certainty criterion were not considered significant. A single miscleavage was allowed (enzyme trypsin) and carbamidomethyl (C) and oxidation (M) were set as fixed and variable modifications respectively. Detailed information on protein identification is provided in Table S4.

To find similarities and differences between the 31 substrates in terms of their soluble protein expression profiles (across the same 148 protein species each), we computed the matrix of Pearson-product moment correlation coefficients, using the function numpy.corcoef in python, and dendrograms, using the function scipy.cluster.hierarchy.dendrogram in python, applied to standardized substrate-specific 2D DIGE profiles. Since dendrograms in general depend on the linkage criterion between clusters, we tested different linkages to assess the robustness of tree diagrams. The low levels in the cluster hierarchy, which define the actual substrate clusters such as monoterpenes, were found to be independent of the linkage used, as opposed to the high-level relations between groups of clusters. To convey a sense of the robustness of the clustering, we therefore chose to use the two clustering linkages that are most different to each other, i.e. which cluster substrates according to either the shortest or the farthest distance in the protein space.

Analysis of the membrane protein-enriched fractions

For the 34 growth conditions with two biological replicates each, the total membrane protein fractions were prepared

and analysed essentially as previously reported (Zech *et al*., 2013). Cell breakage was achieved using a French Press® (Sim-Aminco, Rochester, NY, USA). To solubilize the membrane proteins, the cell extracts were treated with ice-cold carbonate and hot SDS. Protein content was determined using the RC-DC assay (Bio-Rad), and protein separation was achieved using 12.5% SDS mini gels (10 × 7 cm; Bio-Rad). Each sample lane (10 μg protein load) was divided into four gel slices, and each slice was cut into smaller pieces (about 1 mm^3) prior to washing, reduction, alkylation and tryptic digest (Zech *et al*., 2013). Separation of peptides was performed with a nano-LC system (UltiMate 3000; Thermo Fisher Scientific, Germering, Bavaria, Germany) equipped with a 25-cm analytical column (C_{18}, 2 μm bead size, 75 μm inner diameter; Thermo Fisher Scientific) operated in a trap-column mode (C_{18}, 5 μm bead size, 2 cm length, 75 μm inner diameter; Thermo Fisher Scientific) using a 120-min linear gradient (Zech *et al*., 2013). The nano-LC eluent was continuously analysed by an online-coupled ion-trap mass spectrometer (amaZon speed ETD; Bruker Daltonik GmbH) using the captive spray electrospray ion source (Bruker Daltonik GmbH). The instrument was operated in positive mode with a capillary current of 1.3 kV and dry gas flow of 3 l min^{-1} at 150°C. Active precursor exclusion was set for 0.2 min. Per full scan MS, 20 MS/MS spectra of the most intense masses were acquired. Protein identification was performed with ProteinScape as described above, including a mass tolerance of 0.3 Da for MS and 0.4 Da for MS/MS searches and applying a target decoy strategy (false discovery rate <1%). The Mascot score of a given protein is defined as the sum of all corresponding peptide ion scores (Perkins *et al*., 1999). We used the Mascot score as a semi-quantitative measure for analyses of the membrane protein-enriched fractions. This is based on the fact that a higher abundance of a protein yields a higher number of detectable peptides (i.e. peptides above threshold) and, hence, increases the protein score. Considering the exponential nature of the Mascot score, marked increases in Mascot scores (due to higher numbers of protein identifying peptides) are, therefore, indicative of an increased protein abundance. Detailed information on relevant protein identification is provided in Table S4.

Results and discussion

Biogeography, habitats and diversity

Aromatoleum aromaticum pCyN1 belongs to the *Aromatoleum*/*Azoarcus*/*Thauera* cluster within the betaproteobacterial order *Rhodocyclales* (Reinhold-Hurek *et al*., 1993; Anders *et al*., 1995; Rabus *et al*., 2019). To assess the global occurrence of this cluster, a literature search was conducted; details are provided in the supplementary material, distinguishing isolated strains (Table S1) and molecular phylotypes (Table S2). Isolated strains (>100) are globally distributed, with the highest numbers of reports from sample locations in Asia, Europe and North America. Likewise, molecular phylotypes were detected on all continents, together with isolated strains, accounting for more than 200 records (Fig. 1A). Members of this phylogenetic cluster occur in a large variety of terrestrial and aquatic habitats, comprising aquatic/aquifer sediments, the rhizosphere, agricultural soils, polluted sites and wastewater treatment plants (Fig. 1B). In particular, the genera *Aromatoleum* and *Thauera*, which encompass the facultative anaerobic degradation specialists, display rather high diversity, i.e. 13 and 15 taxonomically described species respectively. This species diversity is also reflected in the availability of more than 30 genomes from these two genera (Fig. 1C; Rabus *et al*., 2019). Taken together, this synthetic study furnished proof for the suitability of *Aromatoleum* spp. as model organisms to study facultative anaerobic degradation specialists.

Genome

The general features of the 4 366 359 bp circular genome of *A. aromaticum* pCyN1 are presented in Table S3, including comparison to the closely related *A. aromaticum* EbN1^T (Rabus *et al*., 2005), as well as to both *Azoarcus* sp. strain CIB (Martín-Moldes *et al*., 2015) and the N_2-fixing, plant endophyte *Azoarcus* sp. BH72 (Krause *et al*., 2006). The chromosomal loci of genes encoding (i) the pathway-specific enzymatic modules for anaerobic and aerobic degradation and (ii) the components of the respiratory apparatus are illustrated in Fig. 2 (lower panel). The genomes of the *A. aromaticum* strains pCyN1 and EbN1^T are highly syntenic, as evident from the 2674 (~65%) chromosome-encoded proteins that share >98% sequence identity. Notably, the non-identical chromosomal loci encode strain-specific capacities, i.e. degradation of *p*-cymene for strain pCyN1, and of ethylbenzene for strain EbN1^T. The chromosomes of both strains are rich in genetic elements (~240 transposon-related genes, 2–3 phages, CRISPR class I type I-C) conferring high genome plasticity (Fig. 3).

Inspecting the 157.1 kbp chromosomal cluster of *A. aromaticum* pCyN1 that harbours the genes for anaerobic conversion of *p*-cymene to (*S*)-3-isopropylpimeloyl-CoA (Fig. 2, upper panel), several aspects become obvious: (i) the gene cluster is framed by several integrases and transposases; (ii) the genes for the peripheral oxidation of *p*-cymene to 4-isopropylbenzoyl-CoA (*cmdABC*, *iod*, *iad* and *ibl*) are not organized in a contiguous operon-like structure, but are located at rather distant

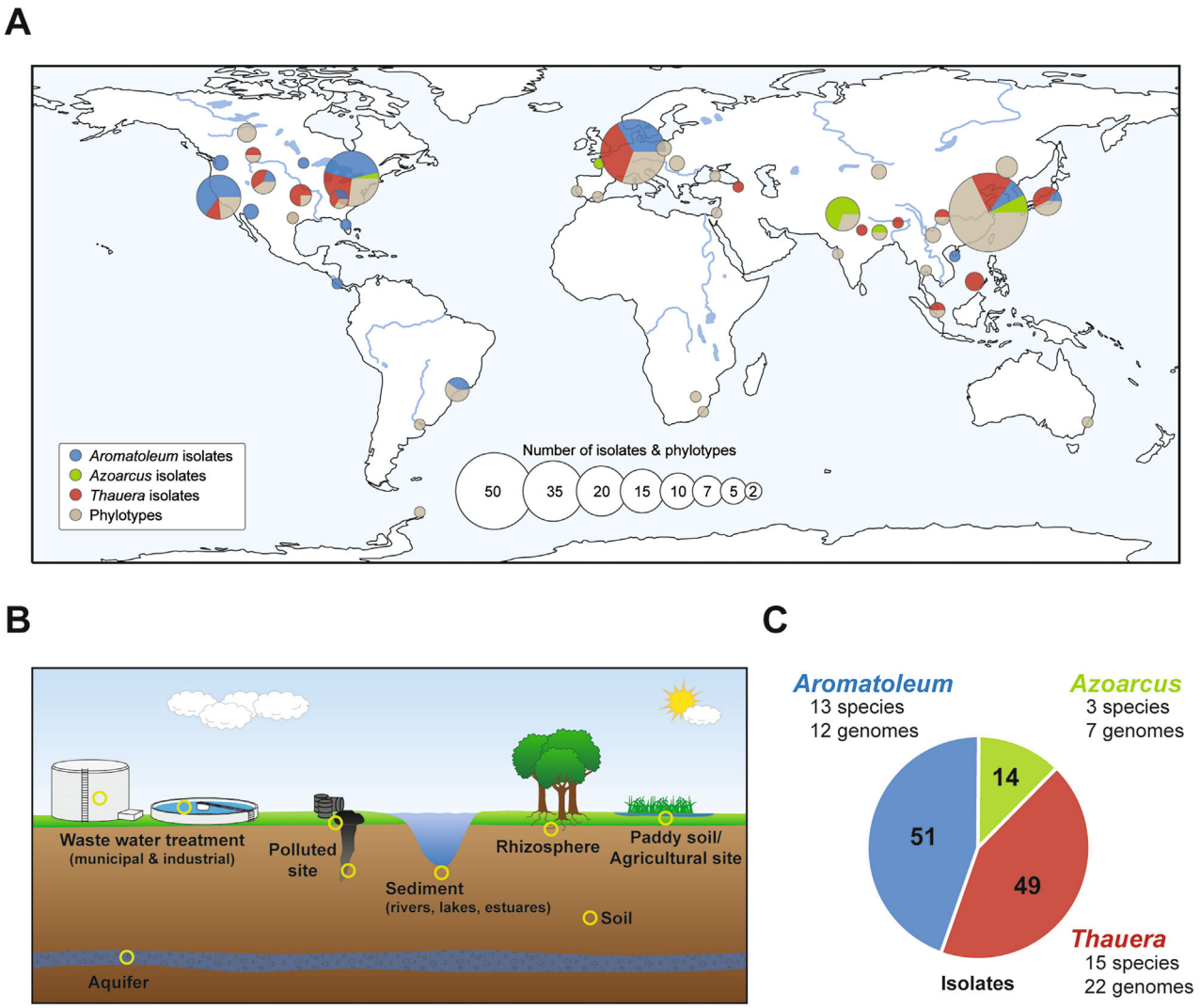

Fig. 1. Global occurrence of the *Aromatoleum*/*Azoarcus*/*Thauera* cluster (betaproteobacterial order *Rhodocyclales*).
A. Biogeography.
B. Habitats.
C. Diversity and genomes. Underlying data (as of March 2020) for isolated strains and molecular phylotypes were retrieved from published literature and are compiled in Tables S1 and S2.

positions, up to 74.6 kbp apart; (iii) likewise, the genes for reductive dearomatization to (*S*)-4-isopropyldienoyl-CoA (*bcrA1B1C1D1*) and subsequent β-oxidation yielding (*S*)-3-isopropylpimeloyl-CoA (*badH1I1K1*) are separated by 20.1 kbp; (iv) notably, this locus also harbours genes for the aerobic utilization of *p*-cymene, again flanked by genes for integrases and transposases, as well as a range of genes encoding proteins for diverse metabolic, cellular, or yet unknown functions; (v) five predicted genomic islands/islets and several transposases/integrases are scattered across the gene cluster. Taken together, one may speculate that the entire gene cluster was acquired by *A. aromaticum* pCyN1 via several consecutive steps of horizontal gene transfer (Koonin *et al*., 2001; Hall *et al*., 2017; Oliveira *et al*., 2017) enabling the strain to use plant-derived monoterpenes under anoxic as well as oxic conditions. Considering that the strain EbN1^T-typifying 'ethylbenzene' gene cluster is predicted to also represent a genomic island (Fig. 3), horizontal gene transfer appears to have been a major driver for shaping the genomes of these two strains, and probably of the pan-genome of the genus *Aromatoleum*.

Overall structure of the anaerobic degradation network

The anaerobic degradation network of *A. aromaticum* pCyN1 (Fig. 4; for compound numbering and names, see legend of Fig. 2) was constructed by superimposing proteomic profiles across 30 anaerobic single-substrate

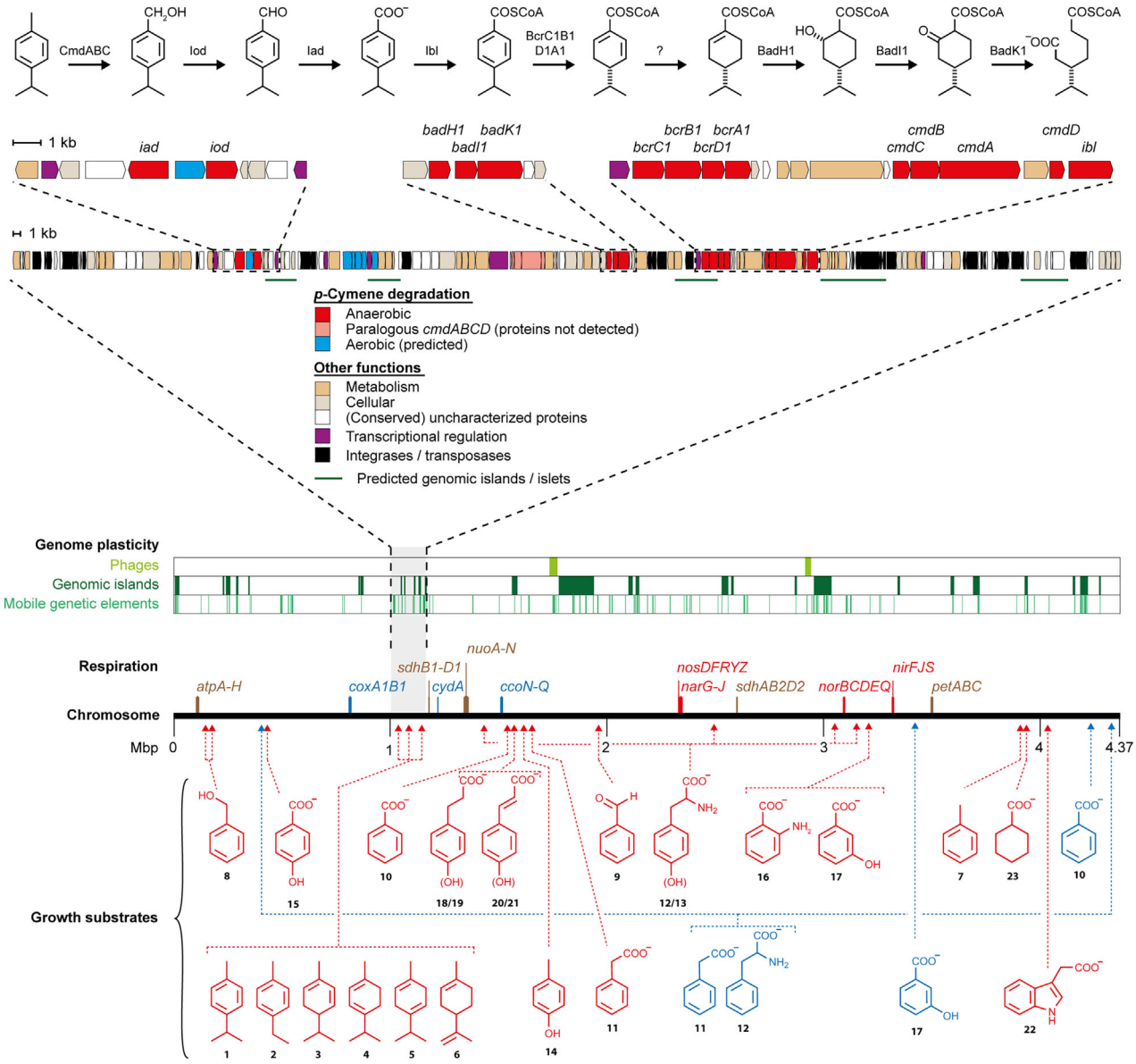

Fig. 2. Genome of *Aromatoleum aromaticum* pCyN1. Upper panel: Chromosomal arrangement of genes for the degradation of *p*-cymene; colour-coding is explained in the centre-positioned legend. The pathway for anaerobic degradation of *p*-cymene to (*S*)-3-isopropylpimeloyl-CoA was recently elucidated by Strijkstra *et al*. (2014) and Küppers *et al*. (2019). Middle panel: Genomic features that enable its plasticity. Lower panel: Scale of the chromosome indicating the positions for genes involved in respiratory energy conservation (complexes are defined in the legend of Fig. S3) and degradation of growth substrates. Compound names (abbreviations as used in Fig. 5 and Table S4): 1, *p*-cymene (pCym); 2, 4-ethyltoluene (4ETol); 3, α-phellandrene (αPhel); 4, α-terpinene (αTerp); 5, γ-terpinene (γTerp); 6, limonene (Lim); 7, toluene (Tol); 8, benzyl alcohol (BzOH); 9, benzaldehyde (BzAl); 10, benzoate (Bz); 11, phenylacetate (PAc); 12, phenylalanine (Phe); 13, tyrosine (Tyr); 14, *p*-cresol (pCr); 15, 4-hydroxybenzoate (4HBz); 16, 2-aminobenzoate (2ABz); 17, 3-hydroxybenzoate (3HBz); 18, hydrocinnamate (HCin); 19, 3-(4-hydroxyphenyl)propanoate (4HPP); 20, cinnamate (Cin); 21, *p*-coumarate (pCou); 22, indoleacetate (IAA); 23, cyclohexane carboxylate (CHC); 24, acetone (Ace); 25, ethanol (EtOH); 26, acetate (Ac); 27, malate (Mal); 28, succinate (Suc); 29, lactate (Lac); 30, pyruvate (Pyr).

adaptation conditions onto the genome-based predictions. This benefitted from previous studies of *A. aromaticum* EbN1^T (e.g. Rabus *et al*., 2014) and the slightly less-closely related, but enzymatically well-studied, *A. evansii* KB740^T (e.g. Hirsch *et al*., 1998) and *Thauera aromatica* K172^T (e.g. Heider *et al*., 1998). Proteogenomic details on the individual enzymes constituting the anaerobic degradation network of *A. aromaticum* pCyN1 are provided in Table S4. The network's structure integrates three consecutive building blocks: first, intriguing initial activation of chemically sluggish substrates, then substrate-specific peripheral

reaction sequences, which finally feed and converge into central degradation pathways. These elements are briefly addressed in the following examples.

Among the various initial activation reactions, the one concerned with a C–H-bond cleavage at phenyl-positioned methyl groups is particularly remarkable, as *A. aromaticum* pCyN1 deploys one of three fundamentally different enzymes, depending on the particular substrate. First, *p*-cymene (4-isopropyltoluene, compound no. 1) is oxidized to 4-isopropylbenzyl alcohol by the putative iron–sulfur molybdoenzyme *p*-cymene dehydrogenase (CmdABC) (Strijkstra *et al*., 2014). Second, toluene (no. 7) is initially added to its co-substrate fumarate by the glycyl-radical enzyme benzylsuccinate synthase (BssABC) forming benzylsuccinate (Heider *et al*., 2016). Third, *p*-cresol (4-methylphenol, no. 14) is hydroxylated to 4-hydroxybenzyl alcohol by FAD-dependent *p*-cresol methylhydroxylase (Cmh) (Vagts *et al*., 2020). Another reaction principle is the activation of mostly mon-oaromatic carboxylic acids to the respective CoA-esters, catalysed by a suite of CoA ligases, i.e. benzoate-CoA ligase (BclA1), hydroxybenzoate-CoA ligase (Hcl1), aminobenzoate-CoA ligase (Hcl2), 4-isopropylbenzoate-CoA ligase (Ibl), phenylacetate-CoA ligase (PadJ), indoleacetate-CoA ligase (IaaB), phenylpropanoid-CoA ligase (Cyn1_15670) and cyclohexane carboxylate-CoA ligase (AliA) (e.g. Egland *et al*., 1997; Schühle *et al*., 2003, 2016; Trautwein *et al*., 2012; Hirakawa *et al*., 2015).

Peripheral reaction sequences include (modified) β-oxidation to benzoyl-CoA in the case of toluene (no. 7) and phenylpropanoids (nos. 12, 13, 18–21), α-oxidation by phenylacetyl-CoA dehydrogenase (PadBCD) and decarboxylation by phenylglyoxylate:NAD^+ oxidoreductase (PadEFGHI) during the anaerobic degradation of phenylacetate (no. 12) (Hirsch *et al*., 1998; Rhee and Fuchs, 1999). In addition, there is a 9-step reaction sequence involving a decarboxylating benzylmalonyl-CoA dehydrogenase (IaaF) in the case of indoleacetate (IAA, no. 22) (e.g. Schühle *et al*., 2021).

The central benzoyl-CoA pathway of facultative anaerobes, as elucidated in *T. aromatica* $K172^T$, is initiated by the highly oxygen-sensitive (class I) benzoyl-CoA reductase (BcrCBAD), which reductively dearomatizes benzoyl-CoA to cyclic 1,5-dienoyl-CoA. This challenging reaction proceeds according to a Birch-like reaction, whereby two electrons are delivered by ferredoxin, radical intermediates are involved, and ATP is stoichiometrically hydrolysed (Boll, 2005, Boll, 2014; Buckel *et al*., 2014). The cyclic 1,5-dienoyl-CoA is then converted to 3-hydroxypimeloyl-CoA via modified β-oxidation reactions involving cyclohexa-1,5-dienecarbonyl-CoA hydratase (Dch), 6-hydroxycyclohex-1-ene-1-carbonyl-CoA dehydrogenase (Had), and lastly 6-oxocyclohex-1-ene-1-carbonyl-CoA hydratase (Oah) (Fuchs *et al*., 2011). Notably, in phototrophic alphaproteobacterial *Rhodopseudomonas palustris*, benzoyl-CoA is presumptively reduced to a

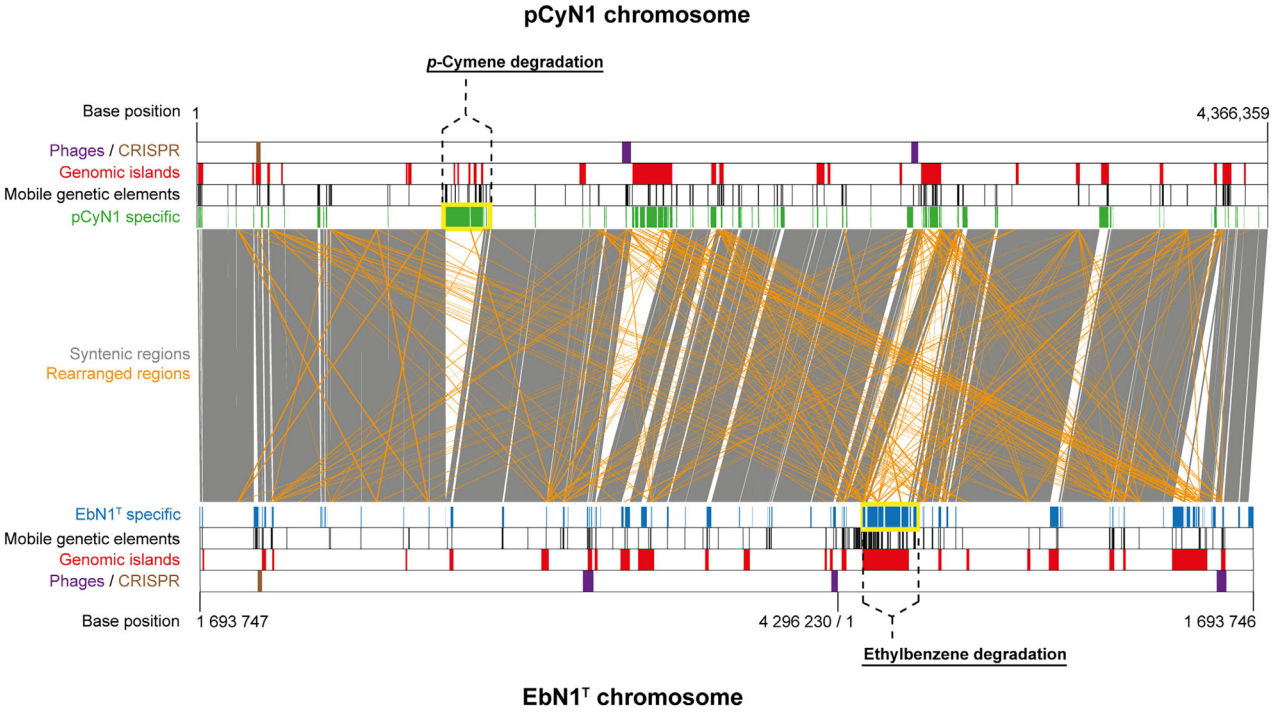

Fig. 3. Synteny plot of the chromosomes of *A. aromaticum* strains pCyN1 (top) and $EbN1^T$ (bottom). Syntenic regions are defined by >95% sequence identity, corresponding to 2930 (71%) chromosome-encoded proteins.

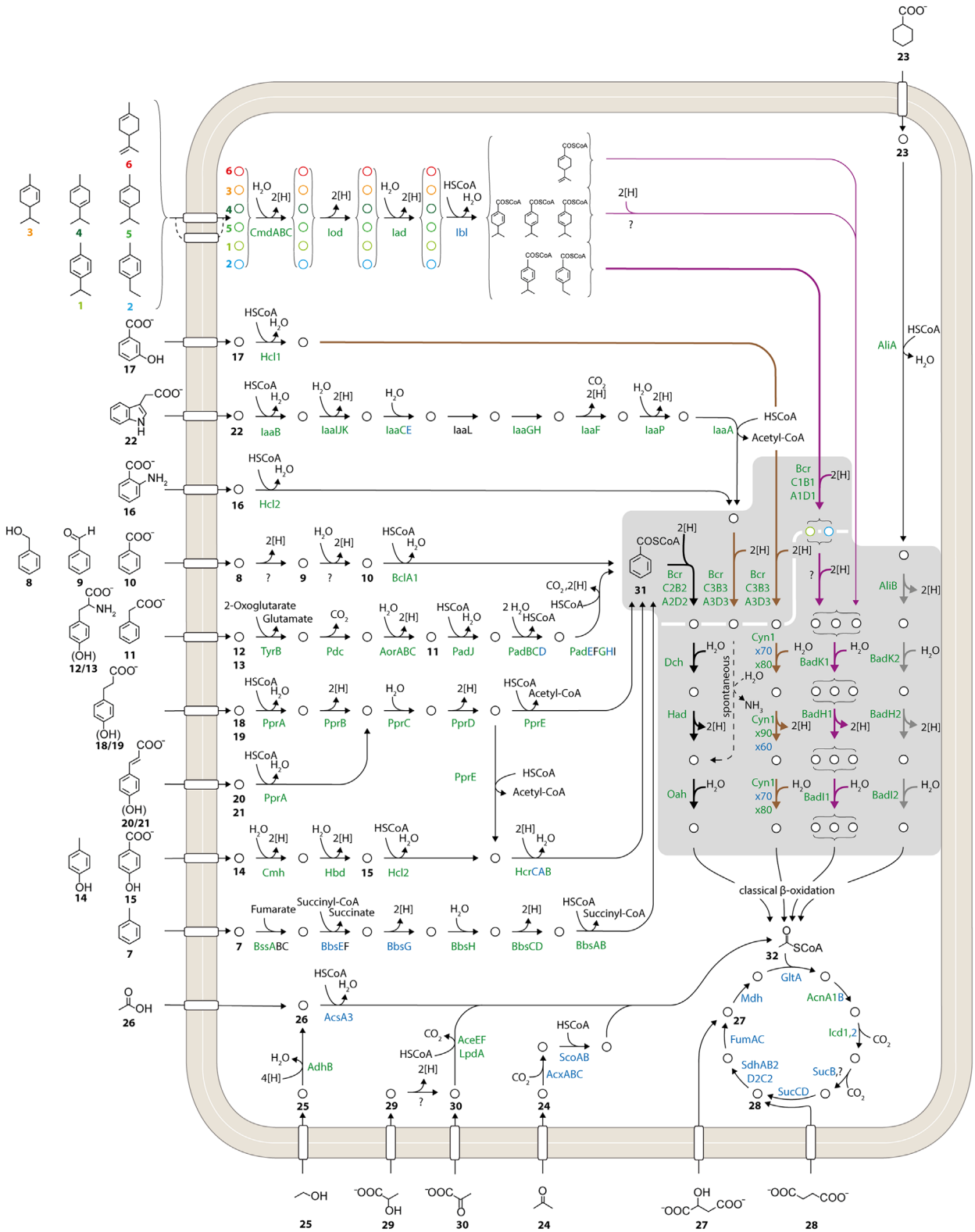

Fig. 4. Anaerobic degradation network of *A. aromaticum* pCyN1 for 30 different terpenoid, aromatic and aliphatic substrates supporting growth that is coupled to denitrification. Network construction was based on superimposing differential proteomic data onto genomic predictions. Specified proteins are colour coded as follows: black, genome predicted only; coloured, identified by proteomics (green, 2D DIGE of soluble proteins; blue, membrane protein-enriched fraction). In the case of multiple protein identifications, 2D DIGE is preferentially indicated. Enzyme names, their predicted functions and respective proteomic data are provided in Table S4. Compound names are as described in the legend of Fig. 2.

cyclic monoenoyl-CoA involving two consecutive reduction steps using two electrons each. Accordingly, subsequent conversion of the monoenoyl-CoA to pimeloyl-CoA involves a different set of β-oxidation enzymes: cyclohex-1-ene-1-carbonyl-CoA hydratase (BadK), 2-hydroxycyclohexane-carbonyl-CoA dehydrogenase (BadH) and 2-oxocyclohexane-carbonyl-CoA hydrolase (BadI) (Egland *et al*., 1997; Pelletier and Harwood, 1998, 2000). In the *A. aromaticum* pCyN1 studied here, we found several variants of this pathway (highlighted by a grey background in Fig. 4). (i) Thirteen aromatic compounds (nos. 7–15, 18–21), after conversion to benzoyl-CoA, are further degraded via the *Thauera*-type pathway involving BcrC2B2A2D2 as well as Dch, Had and Oah. (ii) 2-Aminobenzoate (no. 16), 3-hydroxybenzoate (no. 17) and indoleacetate (no. 22) require a paralogous benzoyl-CoA reductase (BcrC3B3A3D3). Downstream conversion of 2-aminobenzoate and indoleacetate proceeds via the same *Thauera*-type β-oxidation route (Dch, Had, Oah) as used for the 13 aforementioned aromatic compounds. In contrast, that of 3-hydroxybenzoate requires a different set of β-oxidation enzymes (Cyn1_30860/70/80/90). (iii) Subsequent to methyl group oxidation of the six monoterpenes (no. 1–6), degradation of the two aromatic ones (nos. 1 and 2) requires a further paralogous *Thauera*-type benzoyl-CoA reductase (BcrC1B1A1D1) in conjunction with a *Rhodopseudomonas*-type β-oxidation route (BadK1, BadH1, BadI1), which is also used for the four other, non-aromatic monoterpenes (nos. 3–6). More details on anaerobic degradation of these monoterpenes are provided below in the section 'Substrate-specific regulation of the anaerobic degradation network'. (iv) On activation to its CoA-ester (by AliA), the alicyclic substrate cyclohexane carboxylate (no. 23) is catabolized via yet another *Rhodopseudomonas*-type β-oxidation route (AliB, BadK2, BadH2, BadI2). Thus, *A. aromaticum* pCyN1 stands out by arranging three variants of class I benzoyl-CoA reductase and four different modified β-oxidation routes into a central hub, thus accommodating the downstream breakdown of 16 aromatic, six terpenoid and one alicyclic substrate(s) into the central intermediate acetyl-CoA, which is ultimately oxidized to CO_2 via the TCA cycle.

The seven aliphatic growth substrates (nos. 24–30) are channelled into the TCA cycle via mostly conventional reactions, as previously reported for *A. aromaticum* EbN1^T (Rabus *et al*., 2005; Wöhlbrand *et al*., 2007).

Substrate-specific regulation of the anaerobic degradation network

The heat map presented in Fig. 5 displays abundance profiles of all proteins constituting the anaerobic degradation network (Fig. 4) of *A. aromaticum* pCyN1 across 30 substrate adaptation conditions. The underlying data are provided in Table S4. Overall, the profiles of individual degradation modules show substrate specificities of varying degrees, contrasting the functionally required omnipresence of the TCA cycle enzymes. Further details on correlations and cluster relationships are provided in Fig. S1. Potential transcriptional sensory/regulatory systems controlling the degradation modules are compiled in Table S5. In the following, examples of degradation modules are presented.

Monoterpenes. Irrespective of which of the six terpenoic growth substrates (*p*-cymene, 4-ethyltoluene, α-phellandrene, α-terpinene, γ-terpinene and limonene) were used for adaptation, proteins previously assigned to anaerobic degradation of *p*-cymene via 4-isopropylbenzoyl-CoA to (*S*)-3-isopropylpimeloyl-CoA (Strijkstra *et al*., 2014; Küppers *et al*., 2019) were formed with high specificity compared to the other tested substrate conditions (Fig. 5, zoom-in A). This implies a congruous substrate range for this catabolic module, particularly for the proposed *p*-cymene dehydrogenase (CmdABC) (Strijkstra *et al*., 2014). Correspondingly, it stands to reason that CmdABC hydroxylates the methyl group of aromatic and alicyclic monoterpenes with varying types of *para*-positioned alkyl-side chains and degree of ring reduction. Notably, 4-ethyltoluene is also apparently channelled through the methyl-group targeting '*p*-cymene'-pathway, and not the 'ethylbenzene'-pathway. The latter was previously shown in *A. aromaticum* EbN1^T to involve hydroxylation of the benzylic methylene carbon catalysed by a novel Mo/FeS/heme-containing dehydrogenase (Kniemeyer and Heider, 2001). In accordance with this, the respective 'ethylbenzene'-gene cluster (Rabus *et al*., 2002) cannot be found in the genome of *A. aromaticum* pCyN1 (Fig. 3). In addition to this peripheral module, protein components of the downstream conversion, viz., *Thauera*-type class I benzoyl-CoA reductase followed by *Rhodopseudomonas*-type β-oxidation (Küppers *et al*., 2019), also revealed a rather high specificity for these monoterpenes. This, in turn, indicates a concerted transcriptional regulation, even though the respective genes are scattered across the '*p*-cymene' gene cluster (Fig. 2, upper panel). In close proximity to the 'monoterpene degradation' genes of *A. aromaticum* pCyN1, five transcriptional regulators are encoded. Among them, CmtR (TetR-family member) and Cyn1_10250 (MarR-family member) are most likely the relevant ones, since they were identified by proteomics (Table S5).

Cyclohexane carboxylate. Another example of pronouncedly substrate-specific protein formation is represented by the anaerobic degradation module for alicyclic cyclohexane carboxylate (Fig. 5, zoom-in C), encompassing five proteins (AliAB, BadH2I2K2). The MarR-family member BadR is encoded directly between the *aliB* and *badH2* genes and was detected by proteomics (Table S5). Furthermore, BadR from *Rhodopseudomonas palustris* was previously shown to repress

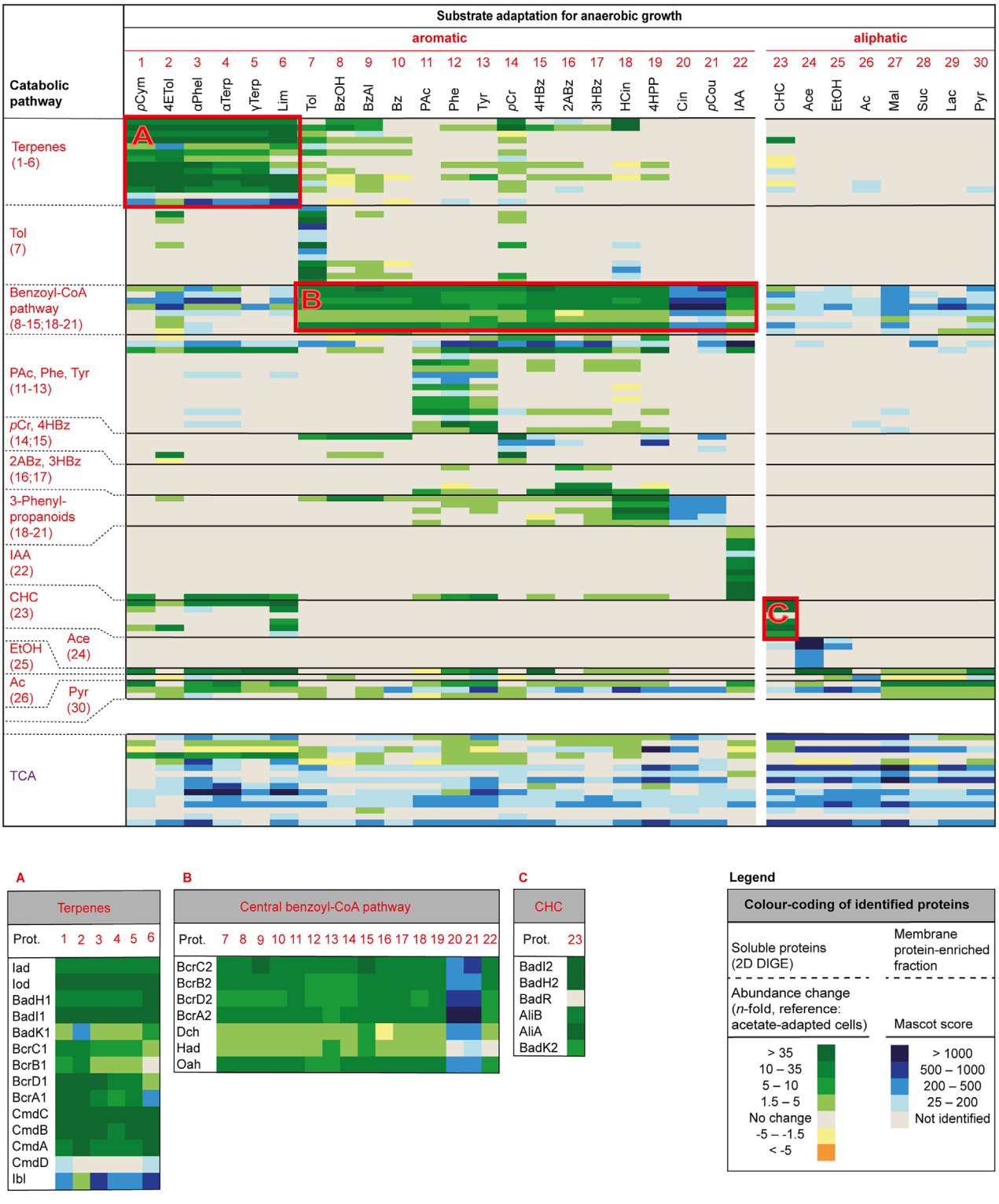

Fig. 5. Differential profiles of identified proteins constituting the degradation network of *A. aromaticum* pCyN1. Proteome profiles were established across 30 different substrates for anaerobic growth. Numbering and abbreviations of compounds are as described in the legend of Fig. 2. Soluble proteins were analysed by 2D DIGE in combination with MALDI-TOF/TOF-MS and the membrane protein-enriched fractions by GeLC-ESI-MS/MS; colour-coding of the profiles is indicated in the legend; detailed proteomic data are provided in Table S4. Zoom-ins (A–C) indicate the different degrees of substrate-specific protein formation. For comparison to aerobic growth with selected aromatic substrates refer to Fig. S2.

expression of the genes for cyclohexane carboxylate degradation (Hirakawa *et al*., 2015).

Benzoyl-CoA pathway. Agreeing with its central catabolic role, the canonical *Thauera*-type benzoyl-CoA pathway displays the highest protein abundances in cultures with 13 tested non-terpenoid monoaromatic substrates for anaerobic growth (Fig. 5, zoom-in B). Remarkably, during anaerobic growth with monoterpenes or aliphatic compounds, as well as during aerobic growth with aromatic compounds, several protein components of this pathway are also detectable, albeit at markedly lower abundances. Regulation of the central benzoyl-CoA pathway in the related *Azoarcus* sp. strain CIB was previously shown to be mediated by the transcriptional repressor BzdR, controlling the conserved P_N promoter and responding to benzoyl-CoA (Durante-Rodríguez *et al*., 2010, 2019). Correspondingly, the *bzdR* gene precedes those for benzoyl-CoA reductase in *A. aromaticum* pCyN1 (Table S5). The detection of proteins not related to the aromatic substrates appending to the central benzoyl-CoA pathway, as observed here, hints at a relaxed stringency of regulation in *A. aromaticum* pCyN1.

Transition from anaerobic to aerobic catabolism

Aerobic aromatic compound degradation. The capacity of *A. aromaticum* pCyN1 to also aerobically degrade aromatic compounds was studied using four selected representatives (details provided in Fig. S2 and Table S4). Protein constituents of the degradation modules were formed with high abundance under the respective substrate conditions, e.g. the key enzymes of aerobic degradation of benzoate via the 'box pathway' [benzoyl-CoA 2,3-epoxidase, BoxAB (Rather *et al*., 2011)] and of phenylacetate [phenylacetyl-CoA 1,2-epoxidase, PaaA1B1C1 D1E1 (Teufel *et al*., 2012)]. Notably, several protein subunits (e.g. BoxB and PaaC1) were also detected during anaerobic growth, indicating a relaxed degree of substrate/redox-specific regulation.

Respiration network. Aromatoleum aromaticum. pCyN1 performs oxidative phosphorylation via denitrification or O_2 respiration. The respective genetic equipment is scattered across the chromosome (Fig. 2, central panel) and resembles that of well-studied model proteobacteria (Zumft, 1997; Borisov and Verkhovsky, 2013; Vik, 2013). The substrate/redox-dependent dynamics of the respiration network of *A. aromaticum* pCyN1 were studied mainly on the basis of membrane proteome profiles, and are summarized in the following (further details provided in Fig. S3 and Table S4): (i) Shared between both respiratory modes are the NADH-dehydrogenase (complex I; NuoA−N) and the succinate dehydrogenase (complex II; SdhAB2C2D2), which represent the entry points into the membrane-embedded quinone pool, as well as the PetABC system (complex III), which connects the quinone pool with periplasmic cytochrome *c*. Corresponding with the general role of these three complexes, the Nuo proteins, SdhAD2 and PetABC were principally present at high abundances across all 34 growth conditions. (ii) Abundance profiles of denitrification enzymes were non-uniform. The nitrate/nitrite antiporter (NarK), nitrate (NO_3^-) reductase (NarGHI) and nitrite (NO_2^-) reductase (NirS) were detected with most of the substrates tested for anaerobic growth, but were essentially absent under oxic conditions. By contrast, nitric oxide (NO) reductase (NorBC) and nitrous oxide (N_2O) reductase (NosZ) could be detected under anoxic as well as oxic conditions. Noteworthy are the likewise high abundances of NosL and NosR under oxic conditions, since they are implicated in the maturation and activation of NosZ (Wunsch and Zumft, 2005; Zhang *et al*., 2019). Expression of denitrification genes requires activation by nitrate-responsive NarXL and NarQP (Stewart, 1993; Constantinidou *et al*., 2006), which are both encoded in the genome of *A. aromaticum* pCyN1. (iii) For O_2 respiration, *A. aromaticum* pCyN1 has a high-affinity (CoxA1B1, complex IV) and a low-affinity oxidase (CcoNOP, complex IV). While the former was only observed under a few of the tested conditions, the CcoNOP system was detected under all of them, albeit with markedly lower abundance during anaerobic growth using aromatic compounds. Apparently, CcoNOP formation is not inactivated on the transcriptional level in the absence of O_2, probably resulting from the absence of the O_2-responsive ArcAB regulator in *A. aromaticum* pCyN1 (Unden *et al*., 1994). However, genes potentially coding for an FNR-like regulator, controlling the metabolic response to oxia-to-anoxia transitions (Bauer *et al*., 1999; Sawer, 1999), are present. One may speculate that the observed pattern of O_2 respiration proteins may prepare *A. aromaticum* pCyN1 to immediately exploit a transient availability of O_2. (iv) As expected, the identified subunits of ATP-synthase of *A. aromaticum* pCyN1 were observed under basically all tested 34 growth conditions.

Genome occupancy and proteome coverage

The comprehensiveness of the current proteogenomic dataset makes it possible to establish the share of the genome's coding space that *A. aromaticum* pCyN1 devotes to its degradation and respiration networks (Fig. 6 and Table S6). The anaerobic degradation and respiration networks are composed of 153 (132 identified; ~90% of coding region) and 62 predicted proteins (54 identified; ~86% of coding region) respectively. For each network, one module was chosen in the following:

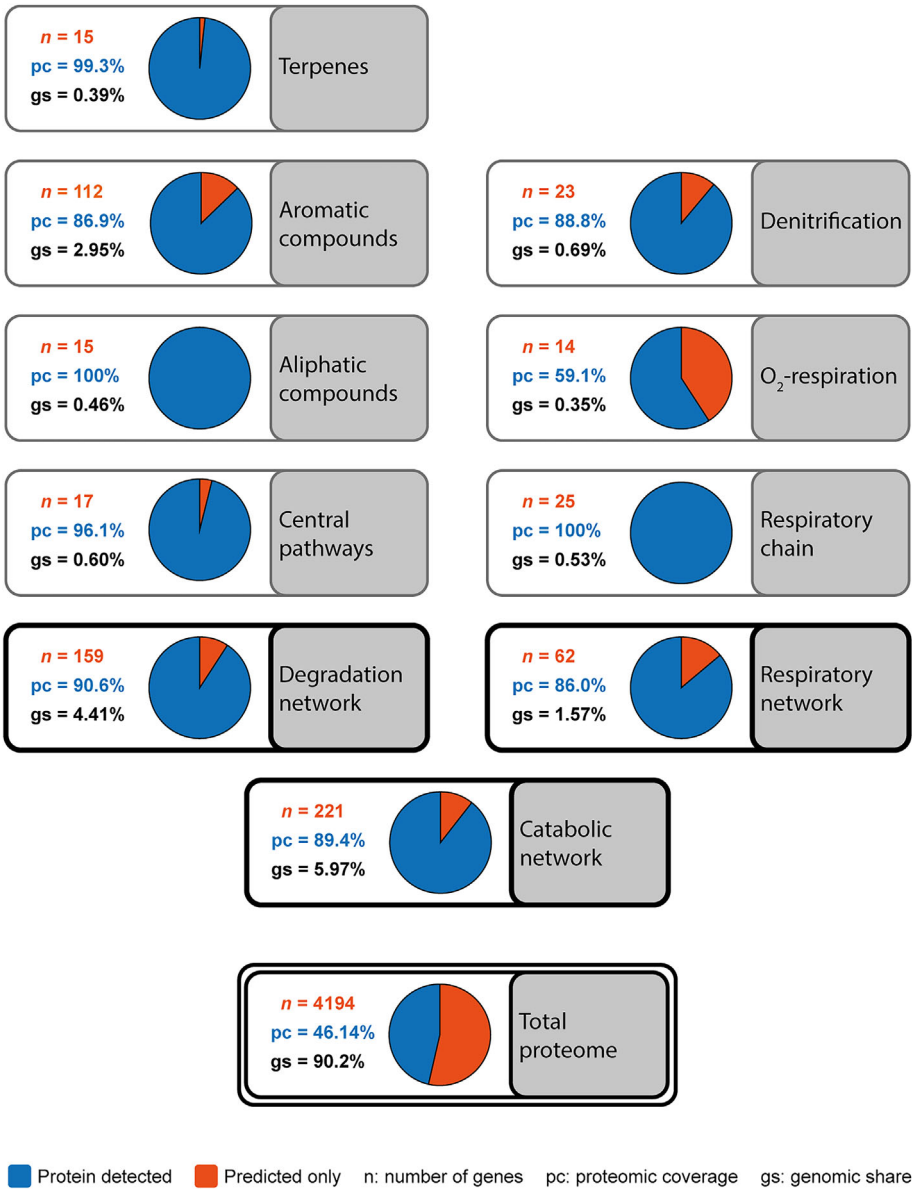

Fig. 6. Gene repertoire, proteome coverage and genome occupancy of the major modules constituting the catabolic network (see Fig. 4 and Fig. S3) of *A. aromaticum* pCyN1.

The module for anaerobic degradation of monoterpenes consists of 15 genes (in total ~17 kbp) and 14 of the predicted proteins were identified. Denitrification comprises 23 genes (~30 kbp), the products of 20 of which were identified. Taken together, the catabolism examined is realized by 215 predicted proteins (186 identified; ~89% of coding region) in *A. aromaticum* pCyN1, occupying 5.78% of the 4.37 Mbp genome's coding space. A recent proteogenomic study on the marine, versatile, sulfate-reducing deltaproteobacterium *Desulfonema magnum* revealed similar values for the catabolic network with 145 genes involved, close to 90% proteome coverage and 2.16% occupancy of the 8.03 Mbp genome (Schnaars *et al*., 2021).

Conclusions and outlook

This study of *A. aromaticum* pCyN1 sheds new light on the catabolic versatility and flexibility characterizing the globally widespread, multi-habitat inhabiting betaproteobacterial cluster *Aromatoleum*/*Azoarcus*/*Thauera*. The traits studied form the physiological basis for occupying a niche characterized by challenging microbe–environment interactions, i.e. altering substrate availability, biochemically challenging terpenoid and aromatic compounds, and fluctuating redox zonation. This soil–rhizosphere environment is biogeochemically relevant, since it represents an important link in the global C- and N-cycles, in particular by heterotrophic facultative denitrifiers (Canfield

et al., 2010). In this respect, nitrous oxide reductase (NosZ) is currently attracting particular attention, since its substrate N_2O, which is increasingly released due to N-fertilization (Shcherbak *et al*., 2014), represents a potent greenhouse gas (Montzka *et al*., 2011) and is an ozone-depleting substance (Ravishankara *et al*., 2009). Yet these facultative anaerobic degradation specialists appear under-investigated compared to well-studied aerobic degraders and to those thriving on easy/fast substrates such as polysaccharides, proteins and lipids. Beyond that, the recent detection of an increased abundance of *Thauera* spp. in the gut microbiota of humans afflicted by substance-use disorders (Xu *et al*., 2017) suggests that such facultative anaerobic degradation specialists also have relevant functions in environments other than those perceived to date (Fig. 1).

The wide-ranging proteome profiles resolved here will facilitate future research efforts. For example, the implied broad substrate spectrum of the predicted *p*-cymene dehydrogenase should be assessed by means of biochemical studies with the purified enzyme. Furthermore, the non-uniformity of protein profiles of the respiration network prompts both the study of, e.g. the effect of different O_2-concentrations, and also genetic analyses of the predicted O_2/N-oxide/redox-sensory/regulatory proteins.

Acknowledgment

Open Access funding enabled and organized by Projekt DEAL.

Author Contributions

R.Ra. conceived the study; R.Re. determined the genome sequence; D.T. performed the cultivation experiments; P.B., manually analysed the complete genome; P.B., R.Ra., D.W., M.N.-S. and D.S. refined the annotation; R.B. analysed the transcriptional regulators; P.B. integrated the differential proteomic data; A.D., C.H. and L.W. performed the proteomic analyses; R.Ra. carried out the literature study on biogeography; M.W. performed computational analyses; R. Ra. wrote the manuscript, with contributions from P.B. and L.W.

References

Anders, H.-J., Kaetzke, A., Kämpfer, P., Ludwig, W., and Fuchs, G. (1995) Taxonomic position of aromatic-degrading denitrifying pseudomonad strains K 172 and KB 740 and their description as new members of the genera *Thauera*, as *Thauera aromatica* sp. nov., and *Azoarcus*, as *Azoarcus evansii* sp. nov., respectively, members of the beta subclass of the *Proteobacteria*. *Int J Syst Microbiol* **45**: 327–333.

ATSDR. (2008) Toxicological profile for cresols. Agency for Toxic Substances and Disease Registry, U.S. Department of Health and Human Services.

Bar-On, Y.M., Phillips, R., and Milo, R. (2018) The biomass distribution on earth. *Proc Natl Acad Sci U S A* **115**: 6506–6511.

Bauer, C.E., Elsen, S., and Bird, T.H. (1999) Mechanisms for redox control of gene expression. *Annu Rev Microbiol* **53**: 495–523.

Boll, M. (2005) Key enzymes in the anaerobic aromatic metabolism catalysing Birch-like reductions. *Biochim Biophys Acta* **1707**: 34–50.

Boll, M., Löffler, C., Morris, B.E.L., and Kung, J.W. (2014) Anaerobic degradation of homocyclic aromatic compounds via arylcarboxyl-coenzyme A esters: organisms, strategies and key enzymes. *Environ Microbiol* **16**: 612–627.

Bond-Lamberty, B., Bailey, V.L., Chen, M., Gough, C.M., and Vargas, R. (2018) Globally rising soil heterotrophic respiration over recent decades. *Nature* **560**: 80–83.

Borisov, V.B., and Verkhovsky, M.I. (2013) Oxygen as acceptor. *EcoSal Plus* **6**: 1–32. https://doi.org/10.1128/ecosalplus.ESP-0012-2015.

Boronat, A., and Rodríguez-Concepcíon, M. (2015) Terpenoid biosynthesis in prokaryotes. *Adv Biochem Eng Biotechnol* **148**: 3–18.

Bradford, M.M. (1976) Rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248–254.

Brewer, T.E., Handley, K.M., Carini, P., Gilbert, J.A., and Fierer, N. (2016) Genome reduction in an abundant and ubiquitous soil bacterium '*Candidatus* Udaeobacter copiosus'. *Nat Microbiol* **2**: 16198.

Buckel, W., Kung, J.W., and Boll, M. (2014) The benzoyl-coenzyme A reductase and 2-hydroxyacyl-coenzyme A dehydratase radical enzyme family. *ChemBioChem* **15**: 2188–2194.

Bulgarelli, D., Schlaeppi, K., Spaepen, S., van Themaat, E. V.L., and Schulze-Lefert, P. (2013) Structure and function of the bacterial microbiota of plants. *Annu Rev Plant Biol* **64**: 807–838.

Canfield, D.E., Glazer, A.N., and Falkowski, P.G. (2010) The evolution and future of earth's nitrogen cycle. *Science* **330**: 192–196.

Carmona, M., Zamarro, M.T., Blázquez, B., Durante-Rodríguez, G., Juárez, J.F., Valderrama, J.A., *et al*. (2009) Anaerobic catabolism of aromatic compounds: a genetic and genomic view. *Microbiol Mol Biol Rev* **73**: 71–133.

Chain, P.S.G., Denef, V.J., Konstantinidis, K.T., Vergez, L. M., Agulló, L., Reyes, V.L., *et al*. (2006) *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *Proc Natl Acad Sci U S A* **103**: 15280–15287.

Constantinidou, C., Hobman, J.L., Griffiths, L., Patel, M.D., Penn, C.W., Cole, J.A., *et al*. (2006) A reassessment of the FNR regulon and transcriptomic analysis of the effects of nitrate, nitrite, NarXL, and NarQP as *Escherichia coli* K12 adapts from aerobic to anaerobic growth. *J Biol Chem* **281**: 4802–4815.

Coyotzi, S., Doxey, A.C., Clark, I.D., Lapen, D.R., van Cappellen, P., and Neufeld, J.D. (2017) Agricultural soil

denitrifiers possess extensive nitrite reductase gene diversity. *Environ Microbiol* **19**: 1189–1208.
Crowther, T.W., van den Hoogen, J., Wan, J., Mayes, M.A., Keiser, A.D., Mo, L., *et al*. (2019) The global soil community and its influence on biogeochemistry. *Science* **365**: eaav0550.
de Leeuw, J.W., Versteegh, G.J.M., and van Bergen, P.F. (2006) Biomacromolecules of algae and plants and their fossil analogues. *Plant Ecol* **182**: 209–233.
Durante-Rodríguez, G., Gutiérrez-del-Arroyo, P., Vélez, M., Díaz, E., and Carmona, M. (2019) Further insights into the architecture of the P_N promoter that controls the expression of the *bzd* genes in *Azoarcus*. *Genes* **10**: 489.
Durante-Rodríguez, G., Valderrama, J.A., Mancheño, J.M., Rivas, G., Alfonso, C., Arias-Palomo, E., *et al*. (2010) Biochemical characterization of the transcriptional regulator BzdR from *Azoarcus* sp. CIB. *J Biol Chem* **285**: 35694–35705.
Egland, P.G., Pelletier, D.A., Dispensa, M., Gibson, J., and Harwood, C.S. (1997) A cluster of bacterial genes for anaerobic benzene ring biodegradation. *Proc Natl Acad Sci U S A* **94**: 6484–6489.
Fierer, N. (2017) Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat Rev Microbiol* **15**: 579–590.
Fuchs, G., Boll, M., and Heider, J. (2011) Microbial degradation of aromatic compounds – from one strategy to four. *Nat Rev Microbiol* **9**: 803–816.
Gade, D., Thiermann, J., Markowsky, D., and Rabus, R. (2003) Evaluation of two-dimensional difference gel electrophoresis for protein profiling. Soluble proteins of the marine bacterium *Pirellula* sp. strain 1. *J Mol Microbiol Biotechnol* **5**: 240–251.
Gibson, J., and Harwood, C.S. (2002) Metabolic diversity in aromatic compound utilization by anaerobic microbes. *Annu Rev Microbiol* **56**: 345–369.
Gordon, D. (2003) Viewing and editing assembled sequences using Consed. *Curr Protoc Bioinformatics* **11**: 1–43.
Hall, J.P.J., Brockhurst, M.A., and Harrison, E. (2017) Sampling the mobile gene pool: innovation via horizontal gene transfer in bacteria. *Philos Trans R Soc B* **372**: 20160424.
Harms, G., Rabus, R., and Widdel, F. (1999) Anaerobic oxidation of the aromatic plant hydrocarbon *p*-cymene by newly isolated denitrifying bacteria. *Arch Microbiol* **172**: 303–312.
Heider, J., Boll, M., Breese, K., Breinig, S., Ebenau-Jehle, C., Feil, U., *et al*. (1998) Differential induction of enzymes involved in anaerobic metabolism of aromatic compounds in the denitrifying bacterium *Thauera aromatica*. *Arch Microbiol* **170**: 120–131.
Heider, J., and Fuchs, G. (1997) Anaerobic metabolism of aromatic compounds. *Eur J Biochem* **243**: 577–596.
Heider, J., Szaleniec, M., Martins, B.M., Seyhan, D., Buckel, W., and Golding, B.T. (2016) Structure and function of benzylsuccinate synthase and related fumarate-adding glycyl radical enzymes. *J Mol Microbiol Biotechnol* **26**: 29–44.
Herms, C.H., Hennessy, R.C., Bak, F., Dresbøll, D.B., and Nicolaisen, M.H. (2022) Back to the roots: exploring the role of root morphology as a mediator of beneficial plant-microbe interactions. *Environ Microbiol*. [Online]. https://doi.org/10.1111/1462-2920.15926.
Hirakawa, H., Hirakawa, Y., Greenberg, E.P., and Harwood, C.S. (2015) BadR and BadM proteins transcriptionally regulate two operons needed for anaerobic benzoate degradation by *Rhodopseudomonas palustris*. *Appl Environ Microbiol* **81**: 4253–4262.
Hirsch, W., Schägger, H., and Fuchs, G. (1998) Phenylglyoxylate:NAD^+ oxidoreductase (CoA benzoylating), a new enzyme of anaerobic phenylalanine metabolism in the denitrifying bacterium *Azoarcus evansii*. *Eur J Biochem* **251**: 907–915.
Jansson, J.K., and Hofmockel, K.S. (2020) Soil microbiomes and climate change. *Nat Rev Microbiol* **18**: 35–46.
Kniemeyer, O., and Heider, J. (2001) Ethylbenzene dehydrogenase, a novel hydrocarbon-oxidizing molybdenum/iron-sulfur/heme enzyme. *J Biol Chem* **276**: 21381–21386.
Koonin, E.V., Makarova, K.S., and Aravind, L. (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55**: 709–742.
Koßmehl, S., Wöhlbrand, L., Drüppel, K., Feenders, C., Blasius, B., and Rabus, R. (2013) Subcellular protein localization (cell envelope) in *Phaeobacter inhibens* DSM 17395. *Proteomics* **13**: 2743–2760.
Krause, A., Ramakumar, A., Bartels, D., Battistoni, F., Bekel, T., Boch, J., *et al*. (2006) Complete genome of the mutualistic, N_2-fixing grass endophyte *Azoarcus* sp. strain BH72. *Nat Biotechnol* **24**: 1385–1391.
Küppers, J., Becker, P., Jarling, R., Dörries, M., Cakić, N., Schmidtmann, M., *et al*. (2019) Stereochemical insights into the anaerobic degradation of 4-isopropylbenzoyl-CoA in the denitrifying bacterium strain pCyN1. *Chemistry* **25**: 4722–4731.
Marmulla, R., and Harder, J. (2014) Microbial monoterpene transformations – a review. *Front Microbiol* **5**: 346.
Martín-Moldes, Z., Zamarro, M.T., del Cerro, C., Valencia, A., Gómez, M.J., Arcas, A., *et al*. (2015) Whole-genome analysis of *Azoarcus* sp. strain CIB provides genetic insights to its different lifestyles and predicts novel metabolic features. *Syst Appl Microbiol* **38**: 462–471.
Montzka, S.A., Dlugokencky, E.J., and Butler, J.H. (2011) Non-CO_2 greenhouse gases and climate change. *Nature* **476**: 43–50.
Nelson, K.E., Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V.A.P., *et al*. (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol* **4**: 799–808.
Oliveira, P.H., Touchon, M., Cury, J., and Rocha, E.P.C. (2017) The chromosomal organization of horizontal gene transfer in bacteria. *Nat Commun* **8**: 841.
Pelletier, D.A., and Harwood, C.S. (1998) 2-Ketocyclohexanecarboxyl coenzyme A hydrolase, the ring cleavage enzyme required for anaerobic benzoate degradation by *Rhodopseudomonas palustris*. *J Bacteriol* **180**: 2330–2336.
Pelletier, D.A., and Harwood, C.S. (2000) 2-Hydroxycyclohexanecarboxyl coenzyme A dehydrogenase, an enzyme characteristic of the anaerobic benzoate degradation pathway used by *Rhodopseudomonas palustris*. *J Bacteriol* **182**: 2753–2760.

Perkins, D.N., Pappin, D.J.C., Creasy, D.M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**: 3551–3567.

Philippot, L., Raaijmakers, J.M., Lemanceau, P., and van der Putten, W.H. (2013) Going back to the roots: the microbial ecology of the rhizosphere. *Nat Rev Microbiol* **11**: 789–799.

Pichersky, E., and Raguso, R.A. (2018) Why do plants produce so many terpenoid compounds? *New Phytol* **220**: 692–702.

Rabus, R., Boll, M., Heider, J., Meckenstock, R.U., Buckel, W., Einsle, O., *et al*. (2016) Anaerobic microbial degradation of hydrocarbons: from enzymatic reactions to the environment. *J Mol Microbiol Biotechnol* **26**: 5–28.

Rabus, R., Kube, M., Beck, A., Widdel, F., and Reinhardt, R. (2002) Genes involved in the anaerobic degradation of ethylbenzene in a denitrifying bacterium, strain EbN1. *Arch Microbiol* **178**: 506–516.

Rabus, R., Kube, M., Heider, J., Beck, A., Heitmann, K., Widdel, F., *et al*. (2005) The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1. *Arch Microbiol* **183**: 27–36.

Rabus, R., Trautwein, K., and Wöhlbrand, L. (2014) Towards habitat-oriented systems biology of "*Aromatoleum aromaticum*" EbN1. Chemical sensing, catabolic network modulation and growth control in anaerobic aromatic compound degradation. *Appl Microbiol Biotechnol* **98**: 3371–3388.

Rabus, R., and Widdel, F. (1995) Anaerobic degradation of ethylbenzene and other aromatic hydrocarbons by new denitrifying bacteria. *Arch Microbiol* **163**: 96–103.

Rabus, R., Wöhlbrand, L., Thies, D., Meyer, M., Reinhold-Hurek, B., and Kämpfer, P. (2019) *Aromatoleum* gen. nov., a novel genus accommodating the phylogenetic lineage including *Azoarcus evansii* and related species, and proposal of *Aromatoleum aromaticum* sp. nov., *Aromatoleum petrolei* sp. nov., *Aromatoleum bremense* sp. nov., *Aromatoleum toluolicum* sp. nov. and *Aromatoleum diolicum* sp. nov. *Int J Syst Evol Microbiol* **69**: 982–997.

Rather, L.J., Weinert, T., Demmer, U., Bill, E., Ismail, W., Fuchs, G., *et al*. (2011) Structure and mechanism of the diiron benzoyl-coenzyme A epoxidase BoxB. *J Biol Chem* **286**: 29241–29248.

Ravishankara, A.R., Daniel, J.S., and Portmann, R.W. (2009) Nitrous oxide (N_2O): the dominant ozone-depleting substance emitted in the 21st century. *Science* **326**: 123–125.

Reinhold-Hurek, B., Hurek, T., Gillis, M., Hoste, B., Vancanneyt, M., Kersters, K., *et al*. (1993) *Azoarcus* gen. nov., nitrogen-fixing proteobacteria associated with roots of Kallar grass (*Leptochloa fusca* (L.) Kunth), and description of two species, *Azoarcus indigens* sp. nov. and *Azoarcus communis* sp. nov. *Int J Syst Bacteriol* **43**: 574–584.

Rhee, S.-K., and Fuchs, G. (1999) Phenylacetyl-CoA:acceptor oxidoreductase, a membrane-bound molybdenum-iron-sulfur enzyme involved in anaerobic metabolism of phenylalanine in the denitrifying bacterium *Thauera aromatica*. *Eur J Biochem* **262**: 507–515.

Rolli, E., Vergani, L., Ghitti, E., Patania, G., Mapelli, F., and Borin, S. (2021) 'Cry-for-help' in contaminated soil: a dialogue among plants and soil microbiome to survive in hostile conditions. *Environ Microbiol* **23**: 5690–5703.

Sawer, G. (1999) The aerobic/anaerobic interface. *Curr Opin Biotechnol* **2**: 181–187.

Schnaars, V., Wöhlbrand, L., Scheve, S., Hinrichs, C., Reinhardt, R., and Rabus, R. (2021) Proteogenomic insights into the physiology of marine, sulfate-reducing, filamentous *Desulfonema limicola* and *Desulfonema magnum*. *Microb Physiol* **31**: 36–55.

Schühle, K., Gescher, J., Feil, U., Paul, M., Jahn, M., Schägger, H., *et al*. (2003) Benzoate-coenzyme A ligase from *Thauera aromatica*: an enzyme acting in anaerobic and aerobic pathways. *J Bacteriol* **185**: 4920–4929.

Schühle, K., Nies, J., and Heider, J. (2016) An indoleacetate-CoA ligase and a phenylsuccinyl-CoA transferase in anaerobic metabolism of auxin. *Environ Microbiol* **18**: 3120–3132.

Schühle, K., Saft, M., Vögeli, B., Erb, T.J., and Heider, J. (2021) Benzylmalonyl-CoA dehydrogenase, an enzyme involved in bacterial auxin degradation. *Arch Microbiol* **203**: 4149–4159.

Shcherbak, I., Millar, N., and Robertson, G.P. (2014) Global metaanalysis of the nonlinear response of soil nitrous oxide (N_2O) emissions to fertilizer nitrogen. *Proc Natl Acad Sci U S A* **111**: 9199–9204.

Sikkema, J., DeBont, J.A.M., and Poolman, B. (1995) Mechanisms of membrane toxicity of hydrocarbons. *Microbiol Rev* **59**: 201–222.

Stewart, V. (1993) Nitrate regulation of anaerobic respiratory gene expression in *Escherichia coli*. *Mol Microbiol* **9**: 425–434.

Strijkstra, A., Trautwein, K., Jarling, R., Wöhlbrand, L., Dörries, M., Reinhardt, R., *et al*. (2014) Anaerobic activation of *p*-cymene in denitrifying betaproteobacteria: methyl group hydroxylation versus addition to fumarate. *Appl Environ Microbiol* **80**: 7592–7603.

Tamang, D.G., Rabus, R., Barabote, R.D., and Saier, M.H., Jr. (2009) Comprehensive analyses of transport proteins encoded within the genome of "*Aromatoleum aromaticum*" strain EbN1. *J Membr Biol* **229**: 53–90.

Teufel, R., Friedrich, T., and Fuchs, G. (2012) An oxygenase that forms and deoxygenates toxic epoxide. *Nature* **483**: 359–362.

Trautwein, K., Wilkes, H., and Rabus, R. (2012) Proteogenomic evidence for β-oxidation of plant-derived 3-phenylpropanoids in "*Aromatoleum aromaticum*" EbN1. *Proteomics* **12**: 1402–1413.

Unden, G., Becker, S., Bongaerts, J., Schirawski, J., and Six, S. (1994) Oxygen regulated gene expression in facultatively anaerobic bacteria. *Antonie Van Leeuwenhoek* **66**: 3–23.

Vagts, J., Weiten, A., Scheve, S., Kalvelage, K., Swirski, S., Wöhlbrand, L., *et al*. (2020) Nanomolar responsiveness of an anaerobic degradation specialist to alkylphenol pollutants. *J Bacteriol* **202**: e00595-19.

van Groenigen, K.J., Osenberg, C.W., and Hungate, B.A. (2011) Increased soil emissions of potent greenhouse gases under increased atmospheric CO_2. *Nature* **475**: 214–216.

Vandenbroucke, M., and Largeau, C. (2007) Kerogen origin, evolution and structure. *Org Geochem* **38**: 719–833.

Vik, S.B. (2013) ATP synthesis by oxidative phosphorylation. *EcoSal Plus* **2**: 1–46. https://doi.org/10.1128/ecosalplus.3.2.3.

Wilkes, H., and Schwarzbauer, J. (2010) Hydrocarbons: an introduction to structure, physico-chemical properties and natural occurrence. In *Handbook of Hydrocarbon and Lipid Microbiology*, Timmes, K.N. (ed). Berlin, Heidelberg: Springer-Verlag, pp. 3–48.

Wöhlbrand, L., Kallerhoff, B., Lange, D., Hufnagel, P., Thiermann, J., Reinhardt, R., *et al*. (2007) Functional proteomic view of metabolic regulation in "*Aromatoleum aromaticum*" strain EbN1. *Proteomics* **7**: 2222–2239.

Wunsch, P., and Zumft, W.G. (2005) Functional domains of NosR, a novel transmembrane iron-sulfur flavoprotein necessary for nitrous oxide respiration. *J Bacteriol* **187**: 1992–2001.

Xu, Y., Xie, Z., Wang, H., Shen, Z., Guo, Y., Gao, Y., *et al*. (2017) Bacterial diversity of intestinal microbiota in patients with substance use disorders revealed by 16S rRNA gene deep sequencing. *Sci Rep* **7**: 3628.

Zech, H., Echtermeyer, C., Wöhlbrand, L., Blasius, B., and Rabus, R. (2011) Biological versus technical variability in 2D-DIGE experiments with environmental bacteria. *Proteomics* **11**: 3380–3389.

Zech, H., Hensler, M., Koßmehl, S., Drüppel, K., Wöhlbrand, L., Trautwein, K., *et al*. (2013) Adaptation of *Phaeobacter inhibens* DSM 17395 to growth with complex nutrients. *Proteomics* **13**: 2851–2868.

Zhang, L., Wüst, A., Prasser, B., Müller, C., and Einsle, O. (2019) Functional assembly of nitrous oxide reductase provides insights into copper site maturation. *Proc Natl Acad Sci U S A* **116**: 12822–12827.

Zumft, W.G. (1997) Cell biology and molecular basis of denitrification. *Microbiol Mol Biol Rev* **61**: 533–616.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. Supporting Information.