Complete Genomes of the Anaerobic Degradation Specialists Aromatoleum petrolei ToN1T and Aromatoleum bremense PbN1T

Arne Weiten\(^a\) Kristin Kalvelage\(^a\) Patrick Becker\(^a\) Richard Reinhardt\(^b\)
Thomas Hurek\(^c\) Barbara Reinhold-Hurek\(^c\) Ralf Rabus\(^a\)

\(^a\)General and Molecular Microbiology, Institute for Chemistry and Biology of the Marine Environment (ICBM), Carl von Ossietzky University of Oldenburg, Oldenburg, Germany; \(^b\)Max-Planck-Genome-Centre Cologne, Max Planck Institute for Plant Breeding Research, Cologne, Germany; \(^c\)Department of Microbe-Plant Interactions, Faculty of Biology and Chemistry, University of Bremen, Bremen, Germany

Keywords
Complete genome · Comparative genomics · Mobilome · Metabolism · Physiology · Anaerobic degradation · Aromatic compounds · Denitrification · Aromatoleum petrolei ToN1\(^T\) · Aromatoleum bremense PbN1\(^T\)

Abstract
The betaproteobacterial genus Aromatoleum comprises facultative denitrifiers specialized in the anaerobic degradation of recalcitrant organic compounds (aromatic and terpenoid). This study reports on the complete and manually annotated genomes of Ar. petrolei ToN1\(^T\) (5.41 Mbp) and Ar. bremense PbN1\(^T\) (4.38 Mbp), which cover the phylogenetic breadth of the genus Aromatoleum together with previously genome sequenced Ar. aromaticum EbN1\(^T\) [Rabus et al., Arch Microbiol. 2005 Jan;183(1):27–36]. The gene clusters for the anaerobic degradation of aromatic and terpenoid (strain ToN1\(^T\) only) compounds are scattered across the genomes of strains ToN1\(^T\) and PbN1\(^T\). The richness in mobile genetic elements is shared with other Aromatoleum spp., substantiating that horizontal gene transfer should have been a major driver in shaping the genomes of this genus. The composite catabolic network of strains ToN1\(^T\) and PbN1\(^T\) comprises 88 proteins, the coding genes of which occupy 86.1 and 76.4 kbp (1.59 and 1.75%) of the respective genome. The strain-specific gene clusters for anaerobic degradation of ethyl-/propylbenzene (strain PbN1\(^T\)) and toluene/monoterpenes (strain ToN1\(^T\)) share high similarity with their counterparts in Ar. aromaticum strains EbN1\(^T\) and pCyN1, respectively. Glucose is degraded via the ED-pathway in strain ToN1\(^T\), while gluconeogenesis proceeds via the reverse EMP-pathway in strains ToN1\(^T\), PbN1\(^T\), and EbN1\(^T\). The diazotrophic, endophytic lifestyle of closest related genus Azoarcus is known to be associated with nitrogenase and type-6 secretion system (T6SS). By contrast, strains ToN1\(^T\), PbN1\(^T\), and EbN1\(^T\) lack nif genes for nitrogenase (including cofactor synthesis and enzyme maturation). Moreover, strains PbN1\(^T\) and EbN1\(^T\) do not possess tss genes for T6SS, while strain ToN1\(^T\) does and facultative endophytic “Aromatoleum” sp. CIB is known to even have both. These findings underpin the functional heterogeneity among Aromatoleum members, correlating with the high plasticity of their genomes.

© 2021 The Author(s)
Published by S. Karger AG, Basel

Arne Weiten and Kristin Kalvelage contributed equally to this study.
Introduction

Aromatic compounds and (mono)terpenes are structurally diverse and have a global abundance that is second only to carbohydrates. These organic molecules occur naturally as constituents of biomacromolecules (lignin, proteins) [de Leeuw et al., 2006], as phytohormones [Pichersky and Raguso, 2018] or as conversion products in anoxic sediments [Fischer-Romero et al., 1996]. Moreover, many aromatic compounds are widely used in industrial synthesis and are often of environmental concern due to their toxic effects [ATSDR, 2008; Sikkema et al., 2005]. Aromatic and terpenoid compounds have two-sided properties as substrates for microorganisms: they are attractive due to their energy-richness, but also demanding due to their chemical stability [Wilkes and Schwarzbauer, 2010]. Aerobic microorganisms meet the catabolic challenge by means of O₂-dependent oxygenase-based catalysis [Gibson and Harwood, 2002]. In anoxic environments (devoid of O₂), various anaerobic microorganisms have been recognized that employ special reactions for the O₂-independent activation of these inert molecules and their reductive dearomatization [e.g., Rabus et al., 2016].

The betaproteobacterial genus Aromatoleum comprises a suite of 12 taxonomically described denitrifying species that anaerobically degrade a large variety of aromatic and terpenoid compounds [Rabus et al., 2019]. This genus of degradation specialists is closely related to the genus Azoarcus harbouring N₂-fixing endophytes of grasses and rice plants [Reinhold-Hurek et al., 1993]. Notably, the Aromatoleum member strain CIB resumes an in-between position, as it anaerobically degrades aromatic compounds and is also capable of an endophytic lifestyle [Fernández et al., 2014]. Aromatoleum aromaticum EbN1T was isolated with ethylbenzene from freshwater sediment [Rabus and Widdel, 1995] and emerged as a model system of the genus Aromatoleum and anaerobic degradation specialists in general [for an overview: Rabus et al., 2014; Heider et al., 2016]. Ar. petrolei ToN1T and Ar. bremense PbN1T were isolated with toluene and propylbenzene, respectively, from the same habitat as Ar. aromaticum EbN1T [Rabus and Widdel, 1995]. Taken together, these three strains cover the phylogenetic breadth of the genus Aromatoleum [Rabus et al., 2019], with strains ToN1T and EbN1T as the currently most distantly related representatives and strain PbN1T residing in an in-between position.

The aims of the present study were (i) to provide manually annotated, high-quality genomes of two further representatives of the genus Aromatoleum, (ii) to compare their predicted metabolic potentials, and (iii) to assess their specialization on anaerobic degradation and delimitation from the Azoarcus-like diazotrophic, endophytic lifestyle.

Results and Discussion

General Genome Features

With sizes of 5,407,396 bp and 4,376,837 bp, respectively, the genomes of Ar. petrolei ToN1T and Ar. bremense PbN1T are at the upper and lower side of the size range previously reported for the genomes of related Ar. aromaticum EbN1T [Rabus et al., 2005], “Aromatoleum” sp. CIB (originally termed Azoarcus sp. CIB) [Martin-Moldes et al., 2015], and the diazotrophic endophyte Az. olearius BH72 [Krause et al., 2006]. A comparison of the general genome features between these five strains is presented in Table 1. Contrasting the presence of two plasmids in the genome of Ar. aromaticum EbN1T, Ar. petrolei ToN1T and Ar. bremense PbN1T possess only a single chromosome.

An overarching property of the four Aromatoleum genomes mentioned above is the high number of mobile genetic elements (Table 1), which was further investigated here by assessing the occurrence of such elements across genome sequenced members of the genera Aromatoleum (9), Azoarcus (7), and Thauera (20) (Fig. 1; online suppl. Fig. S1 and online suppl. Table S1; for all online suppl. material, see www.karger.com/doi/10.1159/000513167). Within the genus Aromatoleum, closely related Ar. bremense PbN1T and Ar. aromaticum EbN1T show highest diversity by harbouring phages (7 vs. 3), CRISPRs (one each) as well as transposase-related genes (170 vs. 237). “Aromatoleum” sp. CIB possesses 5 phages and 101 transposase-related genes, while phylogenetically most distantly branching Ar. petrolei ToN1T has only one incomplete phage and 54 transposase-related genes. Notably, this high share of mobile genetic elements viz. genome plasticity can be observed across 5 further members of the genus Aromatoleum. This applies also to 20 genome-sequenced members of the closely related genus Thauera, likewise comprising anaerobic degradation specialists for aromatic and terpenoid compounds [Anders et al., 1995]. Archeotypical Th. aromatica K172T harbours 3 phages, 1 CRISP, and 91 transposase-related genes, while Th. aminoaeromatica S2 stands out among the 36 investigated genomes by harbouring 8 CRISPRs. Thus, horizontal gene transfer (HGT) [Oliveira et al., 2017] apparently has been
a key driver in shaping the genomes of *Aromatoleum*/*Thauera* members, providing them with a plethora of biochemical tools to exploit recalcitrant small organic molecules as growth substrates and to thereby niche in their dynamic habitats (aquatic sediments, soils, rhizosphere). In accord with the high genome plasticity, the genomes of *Ar. petrolei* ToN1T and *Ar. bremense* PbN1T show virtually no synteny with that of *Ar. aromaticum* EbN1T, whereas the extent of genomic divergence correlates with the phylogenetic distance between the strains (online suppl. Fig. S2). In contrast to these observations with the genera *Aromatoleum* and *Thauera*, the occurrence of mobile genetic elements is less uniform within the genus *Azoarcus*. While the model organism *Az. olearius* BH72 is essentially devoid of mobile genetic elements, the genomes of *Az. communis* strains harbour for example a considerable number of transposases (Fig. 1; online suppl. Table S1) [Krause et al., 2006]. Nevertheless, the apparently more stable genomes of these endophytes most likely reflect the constant environmental conditions provided by the host plant tissue and indicate that the evolutionary trajectory to establish N2-fixing root associations is considerably less complex [Reinhold-Hurek and Hurek, 2011; Geurts et al., 2016].

**Catabolic Network**

The chromosomal loci of gene clusters for the anaerobic and aerobic degradation of aromatic and terpenoid growth-supporting substrates of *Ar. petrolei* ToN1T and *Ar. bremense* PbN1T are illustrated in Figure 2. The composite anaerobic catabolic network derived thereof (Fig. 3) was found to be largely composed of the reaction sequences known from *Ar. aromaticum* EbN1T [Rabus et al., 2005; Wöhlbrand et al., 2007; Trautwein et al., 2012]. Genomic data underlying the schemed network are compiled in online supplementary Table S2. About 57% of the 88 proteins constituting the network have orthologues in strains ToN1T/PbN1T. Overall, the genes underlying the catabolic networks cumulatively occupy 86.1 kbp (1.59%) and 76.4 kbp (1.75%) of the genomes of *Ar. petrolei* ToN1T and *Ar. bremense* PbN1T, respectively (online suppl. Table S3). In the following, selected key catabolic properties of these
Fig. 1. Occurrence of mobile genetic elements across the betaproteobacterial *Aromatoleum/Azoarcus/Thauera* cluster. Underlying data (incl. differentiation of intact, questionable, and incomplete phages) is compiled in online supplementary Table S1. The Bayesian 16S rRNA tree with clade credibility values is shown in online supplementary Fig. S1. *Az. pumilus, Az. nasutitermitis*, and *Az. rhizosphaerae* cluster with *Thauera hydrothermalis* are monophyletic with other species of the genus *Thauera*, and are thus probably misnamed members of this genus.

Fig. 2. Circular chromosomes of *Ar. petrolei* ToN1<sup>T</sup> and *Ar. bremense* PbN1<sup>T</sup> highlighting the positions of gene (clusters) for the anaerobic (red) and aerobic (blue) degradation of aromatic and terpenoid growth substrates. Compound numbering is as follows: 1, toluene; 2, benzyl alcohol; 3, benzaldehyde; 4, benzoate; 5, benzoyl-CoA; 6, ethylbenzene; 7, <i>n</i>-propylbenzene; 8, 1-phenylethanol; 9, 1-phenylpropanol; 10, acetophenone; 11, propiophenone; 12, phenylacetate; 13, phenylalanine; 14, tyrosine; 15, phenol; 16, <i>p</i>-cresol; 17, 4-hydroxybenzoate; 18, 2-aminobenzoate; 19, 3-hydroxybenzoate; 20, phenylpropanoate; 21, 3-(4-hydroxyphenyl) propanoate; 22, cinnamate; 23, cyclohexane carboxylate. Paralogs: <i>a</i><sub>ToN1</sub>, <i>bcrBCD1/2</i>; <i>a</i><sub>PbN1</sub>, <i>bcrABCD1/2</i>; <i>b</i><sub>ToN1</sub>, <i>aldA1–7</i>; <i>b</i><sub>PbN1</sub>, <i>aldA1–5</i>; <i>c</i>, <i>adh1/2</i>; <i>d</i>, <i>padJ1/2</i>; <i>e</i>, <i>padBCD1/2</i>; <i>f</i><sub>ToN1</sub>, <i>padFGH1/2</i>; <i>f</i><sub>PbN1</sub>, <i>padIE1</i>–<i>3</i>; <i>g</i>, <i>padFGH1</i>1/2; <i>h</i>, <i>paaABCDEK1/2</i>; <i>i</i>, <i>paaG1/2</i>; <i>j</i>, <i>paaJ1/2</i>; <i>k</i>, <i>paaJ1/2</i>; <i>l</i><sub>ToN1</sub>, <i>paaH1/2</i>; <i>l</i><sub>PbN1</sub>, <i>paaH1–3</i>.2
two strains will be presented, further evidencing the influence of HGT on the catabolic evolution of *Aromatoleum* spp.

(i) Ethyl-/*n*-Propylbenzene

The catabolic module specifying *Ar. bremense* PbN1^T concerns ethylbenzene and *n*-propylbenzene as growth-supporting hydrocarbon substrates [Rabus and Widdel, 1995]. The anaerobic degradation of ethylbenzene was previously shown in *Ar. aromaticum* EbN1^T to proceed via an upper reaction sequence from ethylbenzene to acetophenone (*ebd/ped* gene cluster) followed by a lower one from acetophenone to the central intermediate benzoyl-CoA (*apc/bal* gene cluster) [Kniemeyer and Heider, 2001a, 2001b; Rabus et al., 2002; Kühner et al., 2005; Jobst et al., 2010] (Fig. 4a, upper panel); further degradation of benzoyl-CoA proceeds via the classical *Thauera*-type ATP-dependent benzoyl-CoA pathway. The genome of *Ar. bremense* PbN1^T contains a single locus harbouring both gene clusters, which includes transposase-related genes implicating acquisition via HGT, as earlier observed with *Ar. aromaticum* EbN1^T [Rabus et al., 2002] (Fig. 4a, lower panel). The *ebd/ped* and *apc/bal* gene clusters share the same organization and high sequence identities (72–93%) between strains PbN1^T and EbN1^T, while the setting of genes for the presumptive substrate-specific 2-component sensory/regulatory systems differs markedly. In *Ar. aromaticum* EbN1^T, the gene pairs

---

**Fig. 3.** Synthetic catabolic network of *Ar. petrolei* ToN1^T and *Ar. bremense* PbN1^T. Colour coding: black, shared by both strains; orange, specific for *Ar. bremense* PbN1^T; green, specific for *Ar. petrolei* ToN1^T. Compound numbering is as detailed in legend of Fig. 2.
Fig. 4. Anaerobic degradation of alkylbenzenes and monoterpenes in
Ar. bremense PbN1T and Ar. aromaticum EbN1T, respectively.

a Degradation pathway (upper panel) and gene clusters (lower panel;
with sequence identities [%]) for ethylbenzene and n-propylbenzene.
Protein names: EbdA–D, ethyl-n-propylbenzene dehydrogenase; Ped, 1-phenylethanol/-propanol dehydrogenase; ApA–E, acetone/-propionate carboxylase; Bal, benzoylacetate/-propionate CoA-ligase; PbN1_04080/41990, putative thiolase. Nomenclature of genes for the 2-component sensory/regulatory systems follows the renaming by Heider et al. [2016]: adiSR (formerly tcs1/tcr1) and ediSR (formerly tcs2/tcr2). Genes for predicted integrases and transposases are indicated in green. Compound names: 5–11, see legend to Fig. 2; 27/28, benzoylacetate/-propionate CoA-ligase; 29/30, benzoylacetate/-propionyl-CoA.

b Degradation pathway (upper panel) and gene clusters (lower panel; with sequence identities [%]) for limonene. Protein names: CmdABC, p-cymene dehydrogenase; Iod, 4-isopropylbenzyl alcohol dehydrogenase; Iad, 4-isopropylbenzaldehyde dehydrogenase; Ibl, putative isopropylbenzoate CoA-ligase; BcrA1-D1, specific benzoyl-CoA reductase; BadK, enoyl-CoA hydratase; BadH, β-hydroxyacyl-CoA dehydratase; BadI, 2-oxo-acyl-CoA hydrolase. Compound names: 25, limonene; 31, perillyl alcohol; 32, perillaldehyde; 33, perillic acid; 34, perillyl-CoA; 35, 2-hydroxy-p-menth-8-en-7-oyl-CoA; 36, 2-oxo-p-menth-8-en-7-oyl-CoA; 37, 3-isopropenyl-pimeloyl-CoA.
adiSR (acetophenone-specific) and ediSR (ethylbenzene-specific) are positioned in-between the two catabolic gene clusters. By contrast, in *Ar. bremense* PbN1T two orthologous sets of ediSR are present up- and down-stream of the catabolic gene clusters, with no hint for adiSR orthologs. Thus, the sequential transcriptional regulation of the *ebd/ped* and *apc/bal* gene clusters by EdiSR and AdiSR, previously proposed for *Ar. aromaticum* EbN1T [Kühner et al., 2005], cannot be unambiguously transferred to *Ar. bremense* PbN1T. The presence of only a single pair of *ebd/ped* and *apc/bal* gene clusters in the genome of *Ar. bremense* PbN1T suggests that the encoded pathway is used for ethylbenzene as well as *n*-propylbenzene. This agrees with a previous study showing that cells of *Ar. bremense* PbN1T adapted to either alkylbenzene have similar specific activities for the dehydrogenation of the latter, while ethylbenzene-adapted cells of *Ar. aromaticum* EbN1T displayed an 85% lower specific activity for *n*-propylbenzene compared to ethylbenzene [Kniemeyer and Heider, 2001a]. These observations prompt questions about how the enzymes of the pathway in *Ar. bremense* PbN1T are adapted to the longer sidechain of *n*-propylbenzene and its degradation intermediates. One may speculate that the here observed sequence divergence (14%) of the catalytic subunit (EbdA) of ethylbenzene dehydrogenase (EBDH) of strain PbN1T could facilitate a more spacious end of the active centre pocket at the molybdenum ion and thereby a better accommodation of the *n*-propylbenzene sidechain as compared to EBDH of *Ar. aromaticum* EbN1T [Kloer et al., 2006]. Future structural studies are required to substantiate this speculation. Likewise, purified (S)-1-phenylethanol dehydrogenase (Ped) of *Ar. aromaticum* EbN1T displays a ~50% lower activity for 1-phenylpropanol as compared to its native substrate [Kniemeyer and Heider, 2001b], paralleling the sequence differences (12%) between the Ped proteins of both strains. By contrast, purified acetophenone carboxylase of *Ar. aromaticum* EbN1T, which has a heterooctameric Apcc(αα'ββ)2 core complex [Weidenweber et al., 2017], converts propiophenone with the same specific rate as its native substrate acetophenone [Jobst et al., 2010], even though substrate-binding subunit ApcA shows a similar sequence divergence (13%) between strains PbN1T and EbN1T as described above for EBDH and Ped preceding in the pathway. Taken together, one may speculate that in *Ar. bremense* PbN1T the ethylbenzene pathway was tuned to additionally accommodate *n*-propylbenzene mostly by specific structural adaptations in the active centre pockets of the first two enzymes, namely EBDH and Ped.

(ii) Toluene

The anaerobic degradation of toluene (online suppl. Fig. S3a) was previously shown in *Ar. aromaticum* EbN1T and *Th. aromaticica* K172T to involve the following genes: tdiiSR (toluene-responsive 2-component sensory regulatory system), bssA-H (benzylsuccinate synthase and accessory proteins), and bbsA-H (modified β-oxidation reaction sequence converting benzylsuccinate into the central intermediate benzoyl-CoA) [Leuthner and Heider, 1998; 2000; Hermuth et al., 2002; Kube et al., 2004]. In *Ar. petrolei* ToN1T the genes related to the anaerobic degradation of toluene are analogously arranged and share high sequence similarities of their encoded proteins (58–96% identities) as compared to *Ar. aromaticum* EbN1T [Kühner et al., 2004] (online suppl. Fig. S3b). A notable difference, however, concerns the tdiiSR genes. While they are localized directly upstream of *bssD* (bss operon) in *Ar. aromaticum* EbN1T, only a *tdiS* gene is found next to *bssD* and a complete *tdiSR* is present about 11 kbp upstream of the *bbs* operon (online suppl. Fig. S3b). Since the *bss* and *bbs* operons of *Ar. petrolei* ToN1T share the promoter consensus with other anaerobic toluene degraders (online suppl. Fig. S3c), one may nonetheless speculate that a coordinated transcriptional regulation of these two operons by TdiSR occurs in *Ar. petrolei* ToN1T, as was previously proposed for *Ar. aromaticum* EbN1T [Kube et al., 2004; Kühner et al., 2005].

(iii) Monoterpenes

Anaerobic growth with monoterpenes was originally not reported for *Ar. petrolei* ToN1T [Rabus and Widdel, 1995], but recently demonstrated for α-terpinene and limonene [Rabus et al., 2019]. The genes capacitating *Ar. petrolei* ToN1T accordingly are organized in several clusters as recently reported for monoterpane-degrading *Ar. aromaticum* pCyN1 [Harms et al., 1999; Strijkstra et al., 2014] (Fig. 4b, lower panel): (i) The *cmdABC* genes, encoding the dehydrogenase for initial hydroxylation of the allylic methyl group in α-terpinene and limonene, have the same order and share moderate sequence similarities (63–82% identities). In particular, the 31% sequence deviation between the catalytic CmdA subunits of both strains may point to differing substrate ranges. Predicted CmdABC of *Ar. petrolei* ToN1T and *Ar. aromaticum* pCyN1 phylogenetically affiliate with other iron-sulfur molybdooenzymes for anaerobic activation of hydrocarbons [Strijkstra et al., 2014]; they do not share domain similarities with the flavin-containing dehydrogenase CtmAB of the betaproteobacterium *Castellaniella defragrans*, likewise oxidizing limonene to perillyl alcohol [Puentes-Calca et al., 2018]. (ii) The *iad*, *iod*, and *ibl* genes...
coding for the enzymes conducting the subsequent conversion to the acyl-CoA intermediate [Strijkstra et al., 2014], are differently organized in the two strains, with the Iod proteins sharing high sequence similarity (97%).

(iii) Modified β-oxidation yielding the respective pimeloyl-CoA derivate is then presumably conducted by the BadH-K enzymes as recently proposed for Ar. aromatiquecum [Küppers et al., 2019]. In case of limonene the exocyclic double bond is probably retained along the hypothesized pathway to the level of the pimeloyl-CoA derivate (Fig. 4b, upper panel), while for α-terpinene the fate of the endocyclic double bond at position 3 cannot be predicted at present. Notably, the genome of Ar. petrolei ToN1T does not contain genes for a 4-isopropylbenzoyl-CoA-specific reductase [Küppers et al., 2019], agreeing with its incapacity to anaerobically degrade the aromatic monoterpenes p-cymene [Rabus et al., 2019]. These differences in the genetic equipment and thereof resulting range of utilizable monoterpenes between Ar. petrolei ToN1T and Ar. aromatiquecum pCyN1 further underpin the diversity and species-specific evolution of catabolic modules across the genus Aromatoleum.

(iv) Benzoate/Benzyol-CoA

Further degradation of the central intermediate benzyol-CoA, formed during anaerobic degradation of the vast majority of aromatic growth substrates by Ar. bre- mense PbN1T and Ar. petrolei ToN1T, proceeds in both strains via the classical Thauera-type benzyol-CoA pathway (bcrABCD, dch, had, oah genes) [Boll, 2005], as previously also reported for Ar. aromatiquecum EbN1T [Rabus et al., 2005].
In accord with their lifestyles as facultative anaerobes, Ar. bremente PbN1T and Ar. petrolei ToN1T are capable of aerobic degradation of benzoate. This involves the box pathway (boxABCD genes) known from Ar. evansii KB740T, which is characterized by the O2/NADH-dependent formation of 2,3-epoxybenzoyl-CoA as key intermediate followed by non-oxygenolytic ring cleavage yielding 3,4-dehydrodipicolyl-CoA semialdehyde and formate (Fig. 5, upper panel) [e.g., Rather et al., 2010]. The genomes of strains PbN1T and ToN1T harbour each a complete box gene cluster (Fig. 5), showing essentially the same gene order as previously reported for Ar. evansii KB740T [Gescher et al., 2002, 2005] as well as high sequence identities of the encoded proteins (e.g., 99% and 84% for BoxA of strains ToN1T and PbN1T, respectively). However, there are marked differences in the genetic neighbourhood of the box gene cluster between the two studied strains (Fig. 5, lower panel). In case of strain ToN1T, a complete ABC-transporter is encoded directly upstream of the gene for a benzoate CoA-ligase, as has been described for Ar. evansii KB740T and proposed to mediate uptake of benzoate [Gescher et al., 2002]. By contrast, in strain PbN1T only a gene for a periplasmic solute binding protein is present in proximity of the box gene cluster. Furthermore, strain PbN1T possesses genes for a formate dehydrogenase localized in the neighbourhood of the box gene cluster, which is not the case for strain ToN1T. Previous studies with the aerobic degradation specialist Burkholderia xenovorans LB400 indicated co-regulation of formate-DH and BoxA-D [Denev et al., 2005; Chain et al., 2006], agreeing with the release of formate during ring cleavage. Taken together, the here described variances of box gene cluster neighbourhood provide further evidence for the high genome plasticity among members of the genus Aromatoleum.

(v) Cyclohexane Carboxylate

In the alphaproteobacterium Rhodopseudomonas palustris anaerobic degradation of cyclohexane carboxylate involves the aliAB genes for activation to the CoA-ester and dehydrogenation to cyclohex-1-ene-1-carboxyl-CoA [Perrotta and Harwood 1994; Küver et al., 1995; Pelletier and Harwood, 2000] followed by β-oxidation involving the bad gene products. In Ar. petrolei ToN1T the co-localization of these genes is reminiscent of that previously determined in R. palustris [Larimer et al., 2004], but differs with respect to organization (e.g., aliAB oriented in opposite direction in Ar. petrolei ToN1T) and shares only moderate sequence similarities ranging between 49 and 65% (online suppl. Fig. S4). Evidence for a pathway involving formation of a diene via a 1,4-dehydrogenating cyclohexa-1,5-diene-1-carboxyl-CoA dehydrogenase feeding into the central benzoyl-CoA pathway, as reported for the deltaproteobacterium Geobacter metallireducens [Kung et al., 2014], was not revealed by analysing the genome of Ar. petrolei ToN1T.

Carbohydrate Metabolism

The metabolic potential for degradation and biosynthesis of glucose was compared between Ar. petrolei ToN1T, Ar. bremente PbN1T, Ar. aromaticum EbN1T, and “Aromatoleum” sp. CIB (Fig. 6; online suppl. Table S4). Ar. petrolei ToN1T stands out in possessing paralogues for many genes constituting glycolysis.

Ar. petrolei ToN1T can utilize glucose for growth, contrary to Ar. bremente PbN1T, Ar. aromaticum EbN1T [Rabus et al., 2019], and “Aromatoleum” sp. CIB [Martín-Moldes et al., 2015], and according to the present genome analysis employs the Entner-Doudoroff (ED) pathway for this purpose. The respective enzymes (Zwf, Pgl, Edd, and Eda) are not encoded in the genomes of Ar. bremente PbN1T and Ar. aromaticum EbN1T, but in “Aromatoleum” sp. CIB, for which the absence of a sugar uptake system was regarded as causative for the inability to utilize glucose [Martín-Moldes et al., 2015]. The four compared strains transform their aromatic/terpenoid growth substrates to acetyl-CoA, which is then funnelled into the TCA cycle for terminal oxidation (to CO2). Therefore, synthesis of glucose for anaerobic purposes is an essential metabolic trait of all four strains. To feed acetyl-CoA into anaerobic pathways, it is converted to pyruvate via the glyoxylate shunt, the genes for its enzyme constituents (aceA, isocitrate synthase; glcB, malate synthase; maeb, malic enzyme) are present in the genomes of strains ToN1T and PbN1T, which has previously also been shown for strains EbN1T [Rabus et al., 2005] and CIB [Martín-Moldes et al., 2015]. Furthermore, genome analysis revealed gluconeogenesis to proceed via the reverse Embden-Meyerhof-Parnass (EMP) pathway in the four strains [for “Aromatoleum” sp. CIB see also Martin-Moldes et al., 2015]. All four strains lack the pfk gene for phosphofructokinase, but possess the fbp and fbp genes coding for bidirectional pyrophosphate-dependent fructose-6-phosphate 1-phosphotransferase [Reshetnikov et al., 2008] and unidirectional fructose-1,6-bisphosphatase, respectively. Since the latter was identified during proteomic profiling of Ar. aromaticum EbN1T [Wöhlbrand et al., 2007], one may speculate Fbp to be operative in gluconeogenesis. For the direct conversion of pyruvate to phosphoenolpyruvate, all four strains possess bidirectional phosphoenolpyruvate
synthase (encoded by \textit{ppsA}). Moreover, \textit{Ar. petrolei} ToN1\textsuperscript{T} and “\textit{Aromatoleum}” sp. CIB can form phosphoenolpyruvate from pyruvate via oxaloacetate consecutively employing pyruvate carboxylase (encoded by \textit{pyc}) and phosphoenolpyruvate carboxykinase (encoded by \textit{pckG}).

Taken together, with respect to carbohydrate metabolism, \textit{Ar. petrolei} ToN1\textsuperscript{T} is more similar to “\textit{Aromatoleum}” sp. CIB than to \textit{Ar. bremense} PbN1\textsuperscript{T} and \textit{Ar. aromaticum} EbN1\textsuperscript{T}, which correlates with the phylogenetic relatedness [Martín-Moldes et al., 2015; Rabus et al., 2019].
Genomes of *Aromatoleum* spp. ToN1T and PbN1T

![Diagram](image-url)

Fig. 7. Search for potential *nif* genes for nitrogenase in *Ar. petrolei* ToN1T and *Ar. bremense* PbN1T. **a** Scheme of FeMo-co biosynthesis, nitrogenase maturation and activity (modified from Rubio and Ludden [2008]). Abbreviations: SAM, S-adenosylmethionine; RHC, R-homocitrate; FeMo-co, iron-molybdenum cofactor. **b** Organization of the *nif* gene cluster in *Az. olearius* BH72 [Krause et al., 2006] and comparative BLASTP-search across the genomes of six further selected organisms: *Ar. aromaticum* EbN1T [Rabus et al., 2005], "Aromatoleum" sp. CIB [Martín-Moldes et al., 2015], *A. vinelandii* [Setubal et al., 2009], *K. pneumoniae* [Arnold et al., 1988], *A. brasilense* SP7 [Fouts et al., 2008], and *R. leguminosarum* bv. *trifolii* WSM2304 [Reeve et al., 2010]. Results from BLASTP-searches are given as e-values (−, no Blast output). More details are provided in online supplementary Table S5.

Search for Capacities Enabling Azoarcus-Like Endophytic Lifestyle

The characteristic physiological attribute of the genus *Aromatoleum* is the specialisation on the anaerobic degradation of recalcitrant organic compounds, contrasted by the predominant absence of N₂-fixing capacities, a key trait of the endophytic *Azoarcus* spp. [Rabus et al., 2019]; the only currently known exception is "Aromatoleum" sp. CIB [Fernández et al., 2014; Martín-Moldes et al., 2015].

To reconcile these observations, the genome imprinted physiology of *Ar. petrolei* ToN1T and *Ar. bremense* PbN1T was assessed for clues on endophytic lifestyle capacities by searching for the presence of genes required or suspected for *Azoarcus*-type microbe-plant interaction, i.e., nitrogenase [Egener et al., 1999], type VI protein secretion systems (T6SSs) [Jiang et al., 2019], biosynthesis of indoleacetate [Fernández et al., 2014], pili, flagella, and the TonB-system [Reinhold-Hurek et al., 2006; Böhm et al., 2007; Buschart et al., 2012].

(i) Nitrogenase

Diazotrophic microorganisms fix nitrogen mostly by means of molybdenum nitrogenases composed of two protein complexes [for overview refer to Rubio and Ludden, 2005, 2008]. Dinitrogenase (NifDK; termed MoFe-protein or component I) has a heterotetrameric (α₂β₂) structure and carries two complex metalloclusters: the iron-molybdenum cofactor (FeMo-co), where the mo-
lybdenum is coordinated by an \([\text{Fe}_7\text{-S}_5]\) cluster in conjunction with a \(R\)-homocitrate moiety, and the \(P\)-cluster harbouring an \([\text{Fe}_8\text{-S}_7]\) cluster. Dinitrogenase catalyses the ATP-dependent reduction of the triple bond of nitrogen (\(N_2\)) forming ammonia (NH\(_3\)). Dinitrogen reductase (NifH; termed Fe-protein or component II) is homodimeric, carries a MgATP-binding site on each subunit, transfers electrons via its \([\text{Fe}_4\text{-S}_4]\) cluster to the P-cluster of dinitrogenase, and plays a role in FeMo-co biosynthesis and apoMoFe formation. The \(nif\) gene cluster codes for the nitrogenase (NifHDK) as well as for a suite of other proteins required for the complex process of FeMo-co biosynthesis and nitrogenase maturation, largely elucidated by studies on the gammaproteobacterial diazotrophs Azotobacter vinelandii and Klebsiella pneumoniae [for overview refer to Rubio and Ludden, 2005, 2008] (Fig. 7a): NifSU provide simple FeS clusters for the SAM radical protein NifB to transform itself into the NifB-cofactor carrying the \([\text{Fe}_6\text{-S}_6]\)-core (NifB is essential for Fe-Mo-co synthesis). The NifEN proteins are considered to establish a central node for completing the FeMo-co synthesis from building blocks delivered by accessory proteins: the \([\text{Fe}_6\text{-S}_6]\)-core is transferred from NifB-cofactor by NifX, the molybdenum is delivered by NifQ, and the \(R\)-homocitrate is furnished by NifV. Finally, the cofactor binding protein NifY is assumed to facilitate insertion of the FeMo-co into apo-NifDK. Expression of the \(nif\) genes is controlled by the NifLA system, which integrates several environmental signals [Martinez-Argudo et al., 2004]. NifA activates \(\sigma^{54}\)-dependent RNA-polymerase and is negatively modulated by protein-protein interaction with NifL, which responds to excess nitrogen (mediated, for example, by GlnK) and oxygen (mediated by the oxidation state of bound flavin).

The organisation of the \(nif\) gene cluster in diazotrophic Azotobacter olearius BH72 as inferred from its genome [Krause et al., 2006] is shown in Figure 7b (online suppl. Table S5). The occurrence of genes encoding \(Nif\) proteins in the genomes of Ar. petrolei ToN1\(^T\) and Ar. bremense PbN1\(^T\) was assessed by using predicted \(Nif\) protein sequences from Az. olearius BH72 as query in biBLAST analyses. For the purpose of a broader comparison, also the genomes of Ar. aromaticum EbN1\(^T\), “Aromatoleum” sp. ClB, as well as the well-studied diazotrophs Azotobacter vinelandii DJ, Klebsiella pneumoniae (here only \(nif\) gene cluster reported), Azospirillum brasilense SP\(^T\), and Rhizobium leguminosarum bv. trifolii WSM2304 were included in this analysis. The genomes of Ar. petrolei ToN1\(^T\), Ar. bremense PbN1\(^T\), and Ar. aromaticum EbN1\(^T\) did not reveal any candidate genes coding for a nitrogenase (NifDKH) or proteins essential for cofactor biosynthesis, i.e., SAM radical protein NifB, FeS-core transferring NifX, and central node forming NifEN. Thus, these three degradation specialists do not possess the genetic equipment required for diazotrophy.

(ii) Type-6 Secretion Systems (T6SS)

Bacteria have evolved a range of secretion systems to excrete proteins (virulence factors or effector proteins) across the cell envelope into the extracellular space or to translocate them directly into eukaryotic and prokaryotic cells. In many pathogens and symbionts, type-3/-4 protein secretion systems are used to deliver effector proteins to plants, while these systems are not common in endophytes in contrast to widespread T6SS [Reinhold-Hurek and Hurek, 2011]. These secretion systems span the inner (IM) as well as outer membrane (OM) of Gram-negative cells [Costa et al., 2015]. The T6SS has broadly been recognized to translocate antibacterial proteins [Hachani et al., 2016], but was recently also implicated in the endophytic lifestyle of Az. olearius BH72 [Shidore et al., 2012; Jiang et al., 2019]. T6SSs are composed of 13 conserved, essential components and structurally organized in two complexes (Fig. 8a) [Costa et al., 2015]. The membrane complex (TssLM) is embedded in the IM and prostrudes to the inner face of the OM. The tail complex is related to contractile phage tails [Basler et al., 2012] and accordingly structured into three parts: (i) the baseplate (TssKE, VgrG [also termed TssI], and probably other proteins), which establishes the physical contact to the membrane complex and serves as assembly platform for the tail; (ii) the tail tube, which consists of haemolysin co-regulated protein (Hcp; also termed TssD) polymerized antipodal to the VgrG spike, and (iii) the tail sheath (TssBC), which is assembled around the expanding tail tube. Following translocations of the effector protein via the expelled tail tube, the contracted tail sheath is thought to be disassembled by the ATPase ClpV (also termed TssH). Activity of T6SS can be posttranslationally modulated via a threonine phosphorylation (TPP) dependent (involving TagEGH) [Mougous et al., 2007] and an independent (involving TagF) [Silverman et al., 2011] pathway. The recent study by Jiang et al. [2019] revealed the presence of two gene clusters for T6SS in Az. olearius BH72, termed \(tss\)-1 and \(tss\)-2 (Fig. 8b), which were proposed to serve different functions. Both \(tss\) clusters were expressed and translated in pure culture; however, \(tssl\) only under conditions of \(N_2\) fixation. Curiously, only T6SS-2 was found to be active based on Hcp detection in the secretome, posttranslationally regulated via phosphorylation by TagEG, and inhibited by TagF.
The occurrence of genes for Az. olearius BH72-type T6SSs in the genomes of Ar. petrolei ToN1T and Ar. bremense PbN1T was assessed by comparison to Ar. aromaticum EbN1T and “Aromatoleum” sp. CIB, as well as to the further selected bacteria (see legend to Fig. 7). Results from BLASTP-searches are given as e-values (−, no Blast output). More details are provided in online supplementary Table S6. c Chromosomal locations and organizations of the tss and tag genes in Ar. petrolei ToN1T and “Aromatoleum” sp. CIB.

Fig. 8. Search for potential tss genes encoding type-6 secretion systems (T6SSs) in Ar. petrolei ToN1T and Ar. bremense PbN1T.

- Scheme of the subcellular architecture of T6SSs (modified from Costa et al. [2015] and Jiang et al. [2019]).
- Organization of the two paralogous tss gene clusters in Az. olearius BH72 [Jiang et al., 2019] and comparative BLASTP-search across the genomes of five well-documented T6SS-employing pathogens Vibrio cholerae [Pukatzki et al., 2006], Pseudomonas aeruginosa [Mougous et al., 2006], and Agrobacterium tumefaciens [Bondage et al., 2016]. Ar. petrolei ToN1T, as previously Genomes of Aromatoleum spp. ToN1T and PbN1T
reported for “Aromatoleum” sp. CIB [Martin-Moldes et al., 2015], possesses an operon-like gene cluster coding for the essential components of a complete T6SS sharing (very) high similarities with their counterparts from Az. olearius BH72 (Fig. 8b; online suppl. Table S6): membrane complex TssJLM (e.g., e-value of e^{-131} for TssM), baseplate TssKEI (e.g., e-value of e^{-94} for TssK), tail sheath TssBC (e.g., e-value of 0 for TssC), and tail tube TssD (e.g., e-value of e^{-27}). The Tag system for posttranslational modulation of T6SS is potentially also encoded in the genome of Ar. petrolei ToN1; however, by contrast to Az. olearius BH72, the tagEFGH genes do not co-localize with the tss-cluster and are scattered across the chromosome as is also the case with “Aromatoleum” sp. CIB (Fig. 8c). Notably, a T6SS system is apparently not encoded in the genomes of Ar. bremenu PbN1T and Ar. aromaticum EbN1T. A noteworthy exception is TssH, which has orthologs in both strains; however, they may function in a different more general role in protein quality control as typical of Hsp100/Clp proteins [Sauer et al., 2004].

(iii) Indoleacetate (IAA) Synthesis

Growth and development of plants are facilitated by the plant hormone indole-3-acetic acid (IAA) and its production by plant-associated bacteria promotes health of the host plants [Patten et al., 2013]. Accordingly, Azospirillum brasilense, which forms IAA also under stress conditions [Malhotra and Srivastava, 2008], was shown to exchange IAA for tryptophan and vitamins with the microalgae Chlorella sorokiniana [Palacios et al., 2016]. Formation of IAA by “Aromatoleum” sp. CIB was previously reported for cultures growing in tryptophan-containing medium [Fernández et al., 2014], while genome analysis suggested Az. olearius strains DQS-4T and BH72 not to be able to synthesize IAA [Faoro et al., 2017].

The formation of IAA from tryptophan can proceed via three different proposed, partly ambiguous pathways, which are named by their respective key intermediate and are illustrated in online supplementary Figure S5. The indoleacetronitrile pathway is supposed to be initiated by a yet unaccounted enzyme decarboxylating tryptophan to indole-3-acetonitrile, which is converted to indole-3-acetamide by the specific dehydratase Oxd [Kato et al., 2000]. The final step, i.e., hydrolytic removal of ammonia, may be catalysed by various enzymes. The indole-3-acetamide pathway is known rather from phytopathogens and involves two enzymes: decarboxylating monooxygenase IaaM [Gaweska et al., 2013] and deaminating hydrolase IaaH [Comai and Kosuge, 1980]. The indole-3-pyruvate pathway proceeds via a deaminating aminotransferase, followed by indole-3-pyruvate decarboxylase IpdC [Ryu and Patten, 2008] and an unaccounted aldehyde dehydrogenase.

The search for genes potentially encoding enzymes of IAA synthesis in the genomes of Ar. petrolei ToN1 and Ar. bremenu PbN1T was conducted with published query sequences and compared to the findings with Ar. aromaticum EbN1T, “Aromatoleum” sp. CIB, and Az. olearius BH72 (online suppl. Table S7). The indoleacetronitrile pathway was not found using comparative genome analysis, since no conclusive hits for the indole-3-acetonitrile forming aldoxime dehydratase and most of the subsequent enzymes could be found in Ar. petrolei ToN1 and Ar. bremenu PbN1T. Similar negative results were obtained for the other three studied strains. Support for the presence of the indole-3-acetamide pathway proved ambiguous as hits for the initial monooxygenase (IaaM) were weak across all five studied strains and those for the subsequent hydrolase (IaaH) represented aminoaacetyl-tRNA aminotransferase (except for Az. olearius BH72 harbouring a predicted iaaH gene). Moreover, these best hits did not co-localize in any of the five studied genomes, while this is the case for functional iaaMH in the gall inducing Erwinia herbicola pv. gypsophilae [Manulis et al., 1998]. The results for the indole-3-pyruvate pathway were also not conclusive. While robust hits for orthologs of indole-3-pyruvate decarboxylase (IpdC) were detected in all five strains, they represent phenylpyruvate decarboxylases in Ar. petrolei ToN1T, Ar. bremenu PbN1T, and Ar. aromaticum EbN1T, assigned to anaerobic degradation of phenylalanine (see online suppl. Table S2) [Schneider et al., 1997], and only in “Aromatoleum” sp. CIB was an IpdC predicted. In case of subsequent aldehyde dehydrogenase, multiple hits were observed with all five strains, hindering a specific assignment to IAA formation. Taken together, the comparative genome analyses indicate that Ar. petrolei ToN1T, Ar. bremenu PbN1T, and Ar. aromaticum EbN1T are not able to synthesize IAA.

(iv) Others

The endophytic lifestyle of Az. olearius BH72 has previously been proposed to require flagella and type-IV pili to achieve endophytic colonization of the host plant [Dörr et al., 1998; Reinhold-Hurek et al., 2006; Böhm et al., 2007; Buschart et al., 2012]. Ar. petrolei ToN1T and Ar. bremenu PbN1T, but not Ar. aromaticum EbN1T, possess all genes for a functional flagellum [Minamino et al., 2019] with analogous gene organization as known from “Aromatoleum” sp. CIB and Az. olearius BH72 (online
Type-IV pili are thin proteinaceous polymers protruding from the cell envelope of Gram-negative bacteria and serve various functions, such as surface attachment for host interaction or cell-to-cell contact for DNA exchange [Hospenthal et al., 2017]. As illustrated in online supplementary Figure S7 (underlying data in online suppl. Table S9), Ar. petrolei ToN1T and Ar. bremense PbN1T possess the genetic equipment for type-IV pili, with highly similar gene arrangements and encoded protein sequences compared to Az. olearius BH72 as well as “Aromatoleum” sp. CIB and Ar. aromaticum EbN1T. In case of the non-endophytic degradation specialists the type-IV pili may well facilitate HGT.

Uptake of iron-chelates, vitamin B12 and other molecules across the outer membrane by TonB-dependent transporters is energized by the proton motive force, which is transduced by the TonB/ExbBD-complex residing in the cytoplasmic membrane [Noinaj et al., 2010; Celia et al., 2020]. Ar. petrolei ToN1T and Ar. bremense PbN1T possess all genes for the TonB-mediated uptake of iron sharing highly similar organization and sequences (online suppl. Fig. S8, underlying data in online suppl. Table S10).

The predicted endonuclease EglA was previously shown by means of transposon mutagenesis to be required for infection of rice roots by Az. olearius BH72 [Reinhold-Hurek et al., 2006]. Notably, an orthologue of eglA was detected in Ar. bremense PbN1T (pbN1_39100, e-value 1e−105, 54% identities), but not in either of the three other investigated strains, adding to the genome-imprinted physiological heterogeneities across the genus Aromatoleum.

**Conclusions**

The present study on the complete genomes of Ar. petrolei ToN1T and Ar. bremense PbN1T underpins the current view on the key properties of Aromatoleum spp., i.e., pronounced genome plasticity due to a multitude of mobile genetic elements, strain-specific gene clusters for typifying degradation capacities, and the predominant incapacity of a diazotrophic endophytic lifestyle (Table 2). One may speculate that Ar. petrolei ToN1T, Ar. bremente PbN1T, Ar. aromaticum EbN1T, “Aromatoleum” sp. CIB, and Az. olearius BH72 represent distinct evolutionary transition states between straight degradation specialists (strain EbN1T) via the chimera (strain...
CIB) to the true diazotrophic endophytes (strain BH72). Furthermore, shared cellular components (flagella, pili, TonB-system), while required for the endophytic lifestyle of Azoarcus sensu strictu, may be employed for different purposes in non-endophytic Aromatoleum spp. For example, type-IV pili can be assumed to facilitate association of Az. olearius BH72 with its host plant, whereas they promote DNA exchange among degradation specialists of the genus Aromatoleum.

Future research on genomes from further isolates as well as metagenome-assembled genomes (MAGs) [Bowers et al., 2017] is needed to assess the occurrence of diazotrophy, endophytic potential, and other niche-shaping physiological capacities across the members of the genera Aromatoleum and Thauera. Since Ar. petrolei ToN1\textsuperscript{T} already possesses a complete tss-cluster, this strain could represent a promising platform for synthetic approaches to transform a degradation specialist into an N\textsubscript{2}-fixing endophyte and thereby learning about the evolutionary mechanisms and paths involved in such a process. Reversely, availability of more genomes from the Aromatoleum/Azoarcus/Thauera cluster will provide new insights into the constraints involved in shaping the different genome architectures, viz. high plasticity (Aromatoleum and Thauera) versus stability (Azoarcus).

**Materials and Methods**

**Strains, Medium, and General Cultivation Conditions**

Aromatoleum petrolei ToN1\textsuperscript{T} and Ar. bremense PbN1\textsuperscript{T} were originally isolated with toluene and n-propylbenzene, respectively, under nitrate-reducing conditions from freshwater mud [Rabus and Widdel, 1995], maintained in the laboratory since then and recently deposited at the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ), Braunschweig, Germany, under nitrate-reducing conditions from freshwater mud [Rabus et al., 2019]. Both strains were cultivated at 28°C under nitrate-reducing conditions in a defined, ascorbate-reduced mineral medium with benzoate (4 mM) as sole source of organic carbon and energy as previously described [Rabus and Widdel, 1995]. All chemicals were of analytical grade.

Harvesting of cultures of both Aromatoleum strains was conducted at half-maximal optical density as follows. Essentially, the complete 400-mL cultures were centrifuged (14,334×g, 30 min, 4°C), the pellets washed in 250 mL washing buffer (100 mM Tris/HCl, 5 mM MgCl\textsubscript{2} × 6 H\textsubscript{2}O, pH 7.5) and re-suspended in 0.8 mL of the same washing buffer. Following anew centrifugation (20,000×g, 10 min, 4°C), pellets were shock frozen in liquid N\textsubscript{2} and stored at −80°C until further analysis.

**DNA Sequencing, Assembly, Annotation**

Isolation of genomic DNA was carried out using the Genomic DNA kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions. Recombinant plasmid and fosmid shotgun libraries were constructed, and plasmid libraries were generated from sonified DNA as previously described [Rabus et al., 2005]. Additionally, a fosmid library was constructed (>40-fold physical coverage) for data finishing and assembly confirmation (Epiconic 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 (ThermoFisher Scientific, Waltham, MA, USA). PHRAP assembly programme 1999, http://www.phrap.org/phredphrapconsed.html and Consed [Gordon, 2003] were used to assess sequence quality and perform the assembly (>25-fold coverage) with a quality of less than 1 error in 100,000 bases.

Structural rRNAs and tRNAs were determined using RNAmmer [Lagesen et al., 2007] and tRNAscan-SE [Lowe and Eddy, 1997]. Protein-coding sequences (CDS) were predicted by the ORF-finding programme Glimmer\textsuperscript{3} [Delcher et al., 1999] and manually revised and curated using Artemis (v.12.0) [Rutherford et al., 2000] and InterPro [Mitchell et al., 2019]. The generated ORF dataset was screened against nonredundant protein databases (SWISSPROT and TREMBL) [Bairoch and Apweiler, 2000]. Genomic islands and islets (less than 10 kbp) were predicted applying IslandViewer 3 [Dhillon et al., 2015]. The genome was screened for phage-like regions by PHASTER [Arndt et al., 2016] and CRISPR recognition tool [Bland et al., 2007] served in detection of CRISPR sequences. EggNOG database [Huerta-Cepas et al., 2019] was consulted for orthology prediction and functional categorization. The genome data of Ar. aromaticum EbN1\textsuperscript{T}, “Aromatoleum” sp. CIB, and Az. olearius BH72 used for genomic comparison were obtained from the previous publications [Rabus et al., 2005; Krause et al., 2006; Martin-Moldes et al., 2015].

**Phylogenetic Analysis**

The 40 16S rRNA gene sequences were aligned using SILVA Incremental Aligner (SINA, www.arb-silva.de), which takes into account ribosomal secondary structures when aligning sequences [Pruesse et al., 2012]. The alignment had 1,458 positions after manual refinement. Bayesian analysis was performed with NG-phylogeny.fr using the generalized time reversible (GTR) substitution model with default settings [Huelsenbeck and Ronquist, 2001; Lemoine et al., 2019]. The consensus tree from Bayesian analysis was plotted with Bayesian clade credibility values indicated at nodes. The majority-rule consensus tree was visualized using MEGA X [Kumar et al., 2018].

**Sequence Accession Numbers**

The genome sequences of Ar. petrolei ToN1\textsuperscript{T} and Ar. bremense PbN1\textsuperscript{T} have been submitted to GenBank under the BioProjects PRJNA645075 and PRJNA645050, respectively, with accession numbers CP059560 and CP059467, respectively.

**Acknowledgements**

We are grateful to D. Thies (Bremen) for cultivation work.

**Statement of Ethics**

Ethic approval was not required for this type of study.
Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

This study was supported by the Max Planck Society and the Deutsche Forschungsgemeinschaft (SFB TRR 51 and RTG 2226).

Author Contributions

R.Ra. conceived the study; A.W. and K.K. manually annotated and analysed the genomes of Ar. petrolei ToN1T and Ar. bremense PbN1T, respectively; P.B. conducted comparative mobilome and synteny analyses; R.Re. determined the genomes of both Aromatoleum strains; B.R.-H. selected the genes relevant for the life style of Azoarcus spp.; T.H. constructed phylogenetic trees; R.Ra. wrote the manuscript with contributions from A.W., K.K., P.B., and B.R.-H. All authors agreed to the final version of the manuscript.

References

Anders H-J, Kaetze A, Kämper P, Ludwig W, Fuchs G. 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 Bacteriol. 1995 Apr; 45(2):327–33.

Arndt D, Grant JR, Marco A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016 Jul;44(W1):W16–21.

Arnold W, Rump A, Klipp W, Priever UB, Pühler A. Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of Klebsiella pneumoniae. J Mol Biol. 1988 Oct;203(3):715–38.

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

Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TriEMBL.in 2000. Nucleic Acids Res. 2000 Jan; 28(1):45–8.

Basler M, Pilhofer M, Henderson GP, Jensen GI, Mekalanos JJ. Type VI secretion requires a dynamic contractile phage tail-like structure. Mekalanos JJ. Type VI secretion requires a dynamic contractile phage tail-like structure. Int J Mol Sci. 2020 Feb;21(2):375.

Chain PSG, Denef VJ, Konstantinidis KT, Vergez LM, Aguillo L, Reyes VL, et al. Burkholderia xenovarior LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. Proc Natl Acad Sci USA. 2006 Oct;103(12):4529–70.

Comai L, Kosuge T. Involvement of plasmid deoxyribonucleic acid in indoleacetic acid synthesis in Pseudomonas savastanoi. J Bacteriol. 1980 Aug;143(2):950–7.

Costa TRD, Felsiberto-Rodrigues C, Meir A, Prevoit MS, Redzej A, Trokter M, et al. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nat Rev Microbiol. 2015 Jun;13(6):343–59.

Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 1999 Jun;27(23):4636–41.

de Leeuw JW, Versteegh GJM, van Bergen PF. Identification and mechanistic insights. Nat Rev Microbiol. 2015 Jun;13(6):343–59.

Dhillon BK, Laird MR, Shay JA, Winsor GL, Lo R, Badger JH, et al. Complete genome sequence of the N2-fixing broad host range endophyte Klebsiella pneumoniae 342 and virulence predictions verified in mice. PLoS Genet. 2008 Jul;4(7):e1000411.

Gaweska HM, Taylor AB, Hart PJ, Fitzpatrick PF. Structure of the flavoprotein tryptophan 2-monooxygenase, a key enzyme in the formation of galls in plants. Biochemistry. 2013 Apr;52(15):2620–6.

Gescher J, Eisenreich W, Wörth J, Bacher A, Fuchs G. Aerobic benzoyl-CoA catabolic pathway in Azoarcus evansii: studies on the non-oxynolytic ring cleavage enzyme. Mol Microbiol. 2005 Jun;56(6):1586–600.

Gescher J, Zaar A, Mohamed M, Schägger H, Fuchs G. Genes coding for a new pathway of aerobic benzoate metabolism in Azoarcus evansii. J Bacteriol. 2002 Nov;184(21):6301–15.

Geurts R, Xiao TT, Reinhold-Hurek B. What does it take to evolve a nitrogen-fixing endosymbiosis?. Trends Plant Sci. 2016 Mar;21(3):199–208.

Gibson J, Harwood CS. Metabolic diversity in aromatic compound utilization by anaerobic microbes. Annu Rev Microbiol. 2002 May;56:631–59.

Gordon D. Viewing and editing assembled sequences using Consed. Curr Protoc Bioinformatics. 2003;Chapter 11:Unit11–2.
Jobst B, Schühle K, Linne U, Heider J. ATP-dependent carboxylation of acetophenone by a novel type of carboxylase. J Bacteriol. 2010 Mar;192(5):1387–94.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X. Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018 Jun;35(6):1547–9.

Kung JW, Meier A-K, Mergelsberg M, Boll M. Enzymes involved in a novel anaerobic cyclohexane carboxylic acid degradation pathway. J Bacteriol. 2014 Oct;196(20):3667–74.

Küppers J, Becker P, Jarling R, Dörries M, Cakić N, Schmidt-Matthiesen M, et al. Stereoechemical insights into the anaerobic degradation of 4-isopropylbenzoyl-CoA in the denitrifying bacterium strain pCyN1. Chem. Eur. J. 2019 Mar; 25(18):4722–31.

Küver J, Xu Y, Gibson J. Metabolism of cyclohexane carboxylic acid by the photosynthetic bacterium Rhodopseudomonas palustris. Arch Microbiol. 1995 Nov;164(5):337–45.

Lagenk S, Hallin P, Redland E, Aa, Starfeldt H-H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007 May;35(9):3100–8.

Larimer FW, Chain P, Hauser L, Lamerdin J, Malinconico M, Gifford CA, et al. A virulence locus of Pseudomonas aeruginosa encodes a protein secretion apparatus. Science. 2006 Jun;312(5779):1526–30.

Mougdou JD, Girard CA, Ramsdell TL, Mekalanos JJ. Threonine phosphorylation post-translationally regulates protein secretion in Pseudomonas aeruginosa. Nat Cell Biol. 2007 Jul;9(7):797–803.

Noinaj N, Guillier M, Barnard TJ, Buchanan SK. TonB-dependent transporters: regulation, structure, and function. Annu Rev Microbiol. 2010;64:43–60.

Oliveira PH, Touchon M, Cury J, Rocha EPC. The chromosomal organization of horizontal gene transfer in bacteria. Nat Commun. 2017;8(1):84.

Palacios OA, Gomez-Anduro G, Bassan Y, de-Bashan LE. Cryptothan, thiamine and indole-3-acetic acid exchange between Chlorellasorokiniana and the plant growth-promoting bacterium Azospirillum brasilense. FEBS Microbiol Ecol. 2016 Jun;92(6):fiw077.

Patten CL, Blakney AJC, Coulson TJD. Activity, distribution and function of indole-3-acetic acid biosynthetic pathways in bacteria. Crit Rev Microbiol. 2013 Nov;39(4):395–415.

Pelletier DA, Harwood CS. 2-Hydroxycyclohexene-1-carboxylic acid promotes A. brasilense growth, as an enzyme characteristic of the aromatic benzene degradation pathway used by Rhodopseudomonas palustris. J Bacteriol. 2000 May;182(10):2753–60.

Perrotta JA, Harwood CS. Anaerobic metabolism of cyclohexene-1-carboxylic acid biodegraded by Rhodopseudomonas palustris. Appl Environ Microbiol. 1994 Jun;60(6):1775–82.

Pichersky E, Raguso RA. Why do plants produce so many terpenoid compounds? New Phytol. 2018 Nov;220(3):692–702.

Pruesse E, Peplies J, Glöckner FO. SINA: accurate taxonomic assignment of single protein sequence annotations. Nucleic Acids Res. 2007 Mar;35(7):7–9.

Quast C, Yilmaz P, Prüse E, Peplies J, Glöckner FO. SINA: accurate taxonomic assignment of single protein sequence annotations. Nucleic Acids Res. 2013 May;41(9):527–35.

Ragusa RA. Why do plants produce so many terpenoid compounds? New Phytol. 2018 Nov;220(3):692–702.

Riquelme M, Muñoz-Trigo C, Martínez C, Vázquez-Valencia S, Arias A, et al. Whole-genome analysis of Azorarcus sp. strain CIB provides genetic insights to its different life-styles and predicts novel metabolic features. Syst Appl Microbiol. 2015 Oct;38(7):462–71.

Runde S, Mitef N, Seifert F, Schütz E, Mücke H, et al. A novel plant growth-promoting bacterium, strain EbN1. Arch Microbiol. 2004 Mar;181(3):182–94.

Kühner S, Wöhlbrand L, Fritz I, Wruck W, Hufnagel P, et al. Substrate-dependent regulation of anaerobic degradation pathways for toluene and ethylbenzene in a denitrifying bacterium, strain EbN1. J Bacteriol. 2005 Feb;187(4):1493–503.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X. Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018 Jun;35(6):1547–9.

Lowe TM, Eddy SR. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997 Mar;25(6):595–600.

Malhotra M, Srivastava S. Organization of the ipdC region regulates IAA levels in different Azospirillum brasilense strains: molecular and functional analysis of ipdC in strain SM. Environ Microbiol. 2008 May;10(5):1365–73.

Manulis S, Haviir-Chesner A, Brandt MT, Lindow SE, Barash I. Differential involvement of indole-3-acetic acid biosynthetic pathways in pathogenicity and epiphytic fitness of Erwinia herbicola pv. glycopathiae. MPML. 1998 Jul;11(7):634–42.

Kubo M, Heider J, Amann J, Hufnagel P, kühner S, Beck A, et al. Genes involved in the anaerobic degradation of toluene in a denitrifying bacterium, strain EbN1. Arch Microbiol. 2004 Mar;181(3):182–94.

Martin-Moldes Z, Zamaro MT, del Cerro C, Valencina A, Gómez MJ, Arcas A, et al. Whole-genome analysis of Azorarcus sp. strain CIB provides genetic insights to its different life-styles and predicts novel metabolic features. Syst Appl Microbiol. 2015 Oct;38(7):462–71.

Martinez-Ardieu I, Little R, Shearer N, Johnson P, Dixon R. The NiF-NiFA system: a multidomain transcriptional regulatory complex that integrates environmental signals. J Bacteriol. 2004 Feb;186(3):601–10.

Martín-Moldes Z, Zamaro MT, del Cerro C, Valencina A, Gómez-M, Arcas A, et al. Whole-genome analysis of Azorarcus sp. strain CIB provides genetic insights to its different life-styles and predicts novel metabolic features. Syst Appl Microbiol. 2015 Oct;38(7):462–71.

Martinez-Ardui I, Little R, Shearer N, Johnson P, Dixon R. The NiF-NiFA system: a multidomain transcriptional regulatory complex that integrates environmental signals. J Bacteriol. 2004 Feb;186(3):601–10.

Matos-Moldes Z, Zamaro MT, del Cerro C, Valencina A, Gómez-M, Arcas A, et al. Whole-genome analysis of Azorarcus sp. strain CIB provides genetic insights to its different life-styles and predicts novel metabolic features. Syst Appl Microbiol. 2015 Oct;38(7):462–71.
Puentes-Cala E, Liebeke M, Markert S, Harder J. Limonene dehydrogenase hydroxylates the allylic methyl group of cyclic monoterpenes in the anaerobic terpene degradation by Castellaniella defragrans. J Biol Chem. 2018 Jun; 293(24):9520–9.

Pukatzki S, Ma AT, Sturtevant D, Krastins B, Sarracino D, Nelson WC, et al. Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model system. Proc Natl Acad Sci USA. 2006 Jan;103(5):1528–33.

Rabus R, Boll M, Heider J, Meckenstock RU, Buckel W, Einsle O, et al. Anaerobic microbial degradation of hydrocarbons: from enzymatic reactions to the environment. J Mol Microbiol Biotechnol. 2016 Mar;26(1-3):5–28.

Rabus R, Kube M, Heider J, Beck A, Heitmann K, Rabus R, Kube M, Beck A, Widdel F, Reinhardt R. Genes involved in the anaerobic degradation of ethylbenzene in a denitrifying bacterium, strain EbN1. Arch Microbiol. 2002 Dec; 178(6):506–16.

Rabus R, Kube M, Heider J, Beck A, Heitmann K, Widdel F, et al. The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1. Arch Microbiol. 2005 Jan;183(1):27–36.

Rabus R, Trautwein K, Wöhlbrand L. 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. 2014 Apr;98(8):3371–88.

Rabus R, Widdel F. Anaerobic degradation of ethylbenzene and other aromatic hydrocarbons by new denitrifying bacteria. Arch Microbiol. 1995 Feb;163(2):96–103.

Rabus R, Wöhlbrand L, Thies D, Meyer M, Reinhold-Hurek B, Köppler P. Aromatoleum gen. nov., a novel genus accommodating the phylogenetic lineage including Azoarcus evansi and related species, and proposal of Aromatoleum aromaticum sp. nov., Aromatoleum petrolei sp. nov., Aromatoleum bremense sp. nov., Aromatoleum toloulicum sp. nov., and Aromatoleum diliscum sp. nov. Int J Syst Evol Microbiol. 2019 Apr;69(4):982–97.

Rather LJ, Knapp B, Haehnel W, Fuchs G. Coenzyme A-dependent aerobic metabolism of benzoate via epoxide formation. J Biol Chem. 2010 Jul; 285(27):20615–24.

Reeve W, O’Hara G, Chain P, Arldley J, Bräu L, Nandesena K, et al. Complete genome sequence of Rhizobium leguminosarum bv trifolii strain WSM2304, an effective mycorrhizal symbiont of the South American clover Trifolium polymorphum. Stand Genomic Sci. 2010 Jun; 2(1):66–76.

Reinhold-Hurek B, Hurek T. Living inside plants: bacterial endophytes. Curr Opin Plant Biol. 2011 Aug;14(4):435–43.

Reinhold-Hurek B, Hurek T, Gillis M, Hoste B, Vancanneyt M, Kersters K, et al. Azoarcus gen. nov., nitrogen-fixing proteobacteria associated with roots of Kallar grass (Leptochloa fusca (L.) Kunth), and description of two species, Azoarcus indigenus sp. nov. and Azoarcus communis sp. nov. Int J Syst Bacteriol. 1993 Jul;43(3):574–84.

Reinhold-Hurek B, Maes T, Gemmer S, Van Montagu M, Hurek T. An endoglucanase is involved in infection of rice roots by the not cellulose-metabolizing endophyte Azoarcus sp. strain BH72. MPMI. 2006 Feb;19(2):181–8.

Reshetnikov AS, Rozova ON, Khmelenina VN, Reshetnikov AA, Smolyarenko NA, et al. Artemis: sequence visualization and annotation. Bioinformatics. 2000 Oct;16(10):944–5.

Ryu RJ, Patten CL. Aromatic amino acid dependent expression of indole-3-pyruvate decarboxylase is regulated by TyrR in Enterobacter cloacae UW5. J Bacteriol. 2008 Nov;190(21):7200–8.

Sauer RT, Bolon DN, Burton BM, Burton RE, Flynn JM, Grant RA, et al. Sculpting the proteome with AAA+ proteases and disassembly machines. Cell. 2004 Oct;119(1):9–18.

Schneider S, Mohamed ME-S, Fuchs G. Anaerobic metabolism of L-phenylalanine via benzo-yl-CoA in the denitrifying bacterium Thauera aromatica. Arch Microbiol. 1997 Sep;168(4):310–20.

Setubal JC, dos Santos P, Goldman BS, Ertesvåg H, Espin G, Rubio LM, et al. Genome sequence of Azotobacter vinelandii, an obligate aerobe specialized to support diverse anaerobic metabolic processes. J Bacteriol. 2009 Jul; 191(14):4534–45.

Shidore T, Dinse T, Öhrlein J, Becker A. Reinhold-Hurek B. Transcriptomic analysis of responses to exudates reveal genes required for rizosphere competence of the endophyte Azoarcus sp. strain BH72. Environ Microbiol. 2012 Oct;14(10):2775–87.

Sikkema J, de Bont JAM, Poolman B. Mechanisms of membrane toxicity of hydrocarbons. Microbiol Rev. 1995 Jun;59(2):201–22.

Silverman JM, Austin LS, Hsu F, Hicks KG, Hood RD, Mougous JD. Separate inputs modulate phosphorylation-dependent and -independent type VI secretion activation. Mol Microbiol. 2011 Dec;82(5):1277–90.

Strijkstra A, Trautwein K, Jarling R, Wöhlbrand L, Dörries M, Reinhardt R, et al. Anaerobic activation of p-cymene in denitrifying beta-proteobacteria: methyl group hydroxylation versus addition to fumarate. Appl Environ Microbiol. 2014 Dec;80(24):7592–603.

Trautwein K, Wilkes H, Rabus R. Proteogenomic evidence for β-oxidation of plant-derived 3-phenylpropanoids in “Aromatoleum aromaticum” EbN1. Proteomics. 2012 May; 12(9):1402–13.

Weidenweber S, Schühle K, Demmer U, Warkentin E, Ermier U, Heider J. Structure of the acetotphenone carboxylase core complex: prototype of a new class of ATP-dependent carboxylases/hydrolases. Sci Rep. 2017 Jan;7:39674.

Wilkes H, Schwarzbauer J. Hydrocarbons: An introduction to structure, physico-chemical properties and natural occurrence. In: Timmes KN, editor. Handbook of hydrocarbon and lipid microbiology. Springer-Verlag Berlin Heidelberg: 2010:3–48.

Wöhlbrand L, Kallerhoff B, Lange D, Hufnagel P, Thiermann J, Reinhardt R, et al. Functional proteomic view of metabolic regulation in “Aromatoleum aromaticum” strain EbN1. Proteomics. 2007 Jun;7(13):2222–39.