The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1

Ralf Rabus · Michael Kube · Johann Heider · Alfred Beck · Katja Heitmann · Friedrich Widdel · Richard Reinhardt

Archive of Microbiology DOI 10.1007/s00203-004-0742-9
Received: 10 September 2004 Revised: 5 October 2004 Accepted: 8 October 2004

R. Rabus (✉) · F. Widdel
Max Planck Institut für Marine Mikrobiologie,
Celsiusstrasse 1, 28359 Bremen, Germany
e-mail: rrabus@mpi-bremen.de
Tel.: +49-421-2028736, Fax: +49-421-2028580

M. Kube · A. Beck · K. Heitmann · R. Reinhardt
Max Planck Institut für Molekulare Genetik,
Ihnestrasse 73, 14195 Berlin, Germany

J. Heider
Universität Freiburg, Institut für Biologie II, Mikrobiologie,
Schänzlestrasse 1, 79104 Freiburg, Germany

Dedicated to Prof. Dr. Dr. h.c. Gerhard Gottschalk on the occasion of his 70th birthday.
Abstract Recent research on microbial degradation of aromatic and other refractory compounds in anoxic waters and soils has revealed that nitrate-reducing bacteria belonging to the Betaproteobacteria contribute substantially to this process. Here we present the first complete genome of a metabolically versatile representative, strain EbN1, which metabolizes various aromatic compounds including hydrocarbons. A circular chromosome (4.3 Mb) and two plasmids (0.21 and 0.22 Mb) encode 4603 predicted proteins. Ten anaerobic and four aerobic aromatic degradation pathways were recognized, with the encoding genes mostly forming clusters. Presence of paralogous gene clusters (e.g., for anaerobic phenylacetate oxidation), high sequence similarities to orthologs from other strains (e.g., for anaerobic phenol metabolism) and frequent mobile genetic elements (e.g., more than 200 genes for transposases) suggest high genome plasticity and extensive lateral gene transfer during metabolic evolution of strain EbN1. Metabolic versatility is also reflected by the presence of multiple respiratory complexes. A large number of regulators including more than 30 two-component and several FNR-type regulators indicate a finely tuned regulatory network to respond to the fluctuating availability of organic substrates and electron acceptors in the environment. The absence of genes required for nitrogen fixation and specific interaction with plants separate strain EbN1 ecophysiologically from the closely related nitrogen-fixing plant symbionts of the Azoarcus cluster. Supplementary material on sequence and annotation are provided at the Web page http://www.micro-genomes.mpg.de/ebn1/. Electronic Supplementary Material is available for this article at http://dx.doi.org/10.1007/s00203-004-0742-9.

Keywords Complete genome · Anaerobic degradation · Aromatic compounds · Denitrifying bacterium · Betaproteobacteria
Introduction

Aromatic compounds are abundant products in higher plants and major constituents of crude oil. Anthropogenic release of fuel-derived alkylbenzenes or toxic phenolic compounds is frequently of environmental concern. Therefore, microbial degradation of aromatic compounds is of long-standing ecological and microbiological interest. Due to the structural diversity of aromatic compounds in the environment, their degradation involves a multitude of compound-specific enzymes which also harbor resources for biotechnological applications (Wackett and Hershberger 2001).

Aerobic degraders of aromatic compounds and their metabolic pathways for initial hydroxylation steps and ring cleavage with O₂-activating enzymes are known since many decades (Harayama et al. 1992). The anaerobic degradation of aromatic compounds is a more recently discovered process, and most insights into the involved enzyme mechanisms have been achieved only within the past decade (Heider and Fuchs 1997; Gibson and Harwood 2002). Since these enzymes cannot make use of a reactive oxygen species, their mechanisms differ fundamentally from those of the aerobic enzymes and were often found to be unprecedented in biochemistry (Heider and Fuchs 1997; Gibson and Harwood 2002). An overall metabolic principle in the degradation of aromatic compounds is the channeling of various phenolic, carboxylic and hydrocarbon substrates into the pool of a central intermediate, mostly benzoyl-CoA (Heider et al. 1998), which subsequently undergoes reductive deaeromatization and hydrolytic ring cleavage (Boll et al. 2002; Harwood et al. 1999). Additionally, these organisms also use unusual pathways and novel types of oxygenases for aerobic aromatic metabolism (Zaar et al. 2001; Mohamed et al. 2002). Enzymatic reactions and pathways were mostly studied in organisms from a group of denitrifying Betaproteobacteria, such as Thauera aromatica and Azoarcus-related species (Heider and Fuchs 1997), and in the phototrophic Rhodopseudomonas palustris (Gibson and Harwood 2002). The special pathways of these organisms are generally not present in other aerobic bacteria with broad degradative capacities, such as Pseudomonas putida (Nelson et al. 2002), Burkholderia spp. (http://genome.jgi-psf.org/microbial/) or Ralstonia spp. (http://genome.jgi-psf.org/microbial/).

A phylogenetic group of increasing interest with respect to the anaerobic degradation of aromatic compounds is related to the plant-associated Azoarcus species. The number of isolates of this group has been steadily increasing during the past decade and includes many strains degrading various hydrocarbons anaerobically (Spormann and Widdel 2000; Widdel and Rabus 2001). None of these aromatics-degrading strains is plant-associated. A metabolically versatile representative of these organisms is the denitrifying strain EbN1. It degrades either of two aromatic hydrocarbons, toluene or ethylbenzene via different, strictly anaerobic pathways. Several polar aromatic compounds such as benzoate or phenylacetate as well as aliphatic acids, ketones and alcohols are degraded anaerobically and aerobically (Rabus and Widdel 1995, 1996). Because of its versatile metabolism including mechanistically intriguing anaerobic enzymes, the genome of strain EbN1 was analyzed, being the first of a member of this phylogenetic group. The availability of the full genome sequence allows to predict more pathways for the degradation of aromatic compounds than previously known in this strain and to differentiate this strain functionally from the related plant-associated Azoarcus species.
Moreover, it provides evidence for the existence of high genome plasticity and extensive regulatory networks in strain EbN1.

**Materials and methods**

**Sequencing strategy**

Strain EbN1 was cultivated and genomic DNA was prepared as previously described (Rabus and Widdel 1995; Rabus et al. 2002). Two whole-genome shotgun libraries with average insert sizes of 1.5 and 3.5 kb were generated from sonified DNA. The fragments were ligated into pUC19 vector (Fermentas, St. Leon-Rot, Germany) and electroporated into *Escherichia coli* (strain DH10B; Invitrogen, Karlsruhe, Germany). Additionally, a cosmid library was constructed for data finishing and assembly confirmation (at least 28-fold physical genome coverage, minimal tailing path reconstructed; pWEB vector, Epicentre Technologies, Madison, WI, USA). Templates for sequencing were obtained by insert amplification via PCR or by plasmid isolation. DNA was sequenced using ABI3730 capillary sequencer systems (ABI, Weiterstadt, Germany). Gaps and regions of weak quality were improved by resequencing and primer walking with plasmides, and sequencing of long-range PCR amplicons and cosmids. The quality of raw sequence data was determined with PHRED (Ewing and Green 1998; Ewing et al. 1998). Sequences were assembled with phrap2gap (http://www.sanger.ac.uk/Software/sequencing/docs/phrap2gap/) and GAP4 of the Staden Package (Staden et al. 1999) was used for final editing of the sequences. The sequence data have a quality of less than 1 error within 100,000 bases and have been deposited in EMBL (accession numbers CR555306, CR555307 and CR555308).

**Genome analysis**

ORPHEUS (Frischman et al. 1998) was used for the prediction of genes on the finished chromosome and the two plasmids. The program was adjusted to predict open reading frames (ORFs) ≥105 bp and an extended ORF-set was created that included all other possible ORFs ≥105 bp. All ORFs with an HTGA score ≥5 (see below) were merged into the first data set. ORF prediction was manually refined using ARTEMIS (Rutherford et al. 2000). Overlapping ORFs without functional assignment or BLAST hits were removed from the list of predicted genes. To identify ORFs that were not noticed during initial prediction, the translated nucleotide sequence of the genome was screened against protein databases (SWISSPROT, TREMBL and PIR).

Similarity searches were carried out by the BLAST programs (Altschul et al. 1997) and translated amino acid sequences of predicted ORFs were screened against a non-redundant protein database (compiled from SWISSPROT, TREMBL and PIR). Functional assignments were performed with the INTERPRO system (Apweiler et al. 2001) using the modules (Baxevanis 2002) PROSITE, Pfam, PRINTS, ProDom, SMART, TIGRFAMs and SIGNALP (Nielsen et al. 1997) and including direct links to the gene ontology system (The Gene Ontology Consortium 2001). In addition, predicted
ORFs were screened against the Clusters of Orthologous Groups of proteins (COGs; ref. Baxevanis 2002). These methods were implemented in the annotation platform HTGA (High Throughput Genome Annotation; ref. Rabus et al. 2002), which allowed automated annotation of the predicted ORFs. As an additional feature of the HTGA system, all information generated during functional assignment is loaded into an SQL-database and edited manually on a web-front-end for final annotation. More details on the design of the HTGA system are provided on the web page http://www.micro-genomes.mpg.de/ebn1/.

Results

General genome features

The general characteristics of the genome (1 chromosome and 2 plasmids) of strain EbN1 are summarized in Table 1 and Fig. 1. Proteins involved in anaerobic or aerobic aromatic degradation amount to more than 150, highlighting the importance of this nutritional capacity for strain EbN1. A striking feature of the genome is the high number (237) of transposon-related genes, which are distributed across the chromosome (Fig. 1) and the two plasmids. Only few prokaryotic genomes contain more of these genes (see Fig. 1.3 in supplementary material). There are also several genes in the genome coding for various “addiction systems”. These usually consist of two small genes, one coding for a toxin, the other for a specific antitoxin, and are often found on plasmids (Hayes 2003). Since the antitoxin is more labile than the toxin, bacteria containing such gene pairs only survive as long as both genes are present and expressed. Genes coding for known toxins or antitoxins are present at 14 sites on the chromosome of strain EbN1 and are often (11 cases) in close proximity to genes for transposases. Thus, the presence of such modules may enforce the maintenance of mobile genetic elements in the chromosome in a similar way as known for plasmid-based systems. In addition, the chromosome harbors 2 prophages. One is located between 3.26 and 3.29 Mb with a gene for a phage-related integrase at one side. The other is located between 4.24 Mb and 13 kb (extending over the ends of the sequence file) and flanked on both sides by integrase genes.

Both plasmids and the chromosome contain genes related to conjugal transfer and plasmid-specific functions; some other genes of plasmid 1 apparently code for a tellurite resistance system, and several genes of plasmid 2 are potentially related to aromatic or organic acid metabolism. Interestingly, plasmid 2 resembles the chromosome in G+C content and average gene length, whereas plasmid 1 differs profoundly with respect to these criteria (Table 1).

The main focus of this study is on aromatic metabolism as an ecophysiologicaly decisive capacity of strain EbN1 and its relatives. Details of general genome features, housekeeping functions and further metabolic properties are provided as supporting material. Important reference bacteria for genome analysis were strains representing three approximately equidistant phylogenetic branches of the Azoarcus/Thauera group, A. evansii (aromatics-degrading, non-plant-associated, same phylogenetic cluster as strain EbN1 and to be renamed), T. aromatic (aromatics-degrading; non-plant-associated, separate phylogenetic cluster) and Azoarcus sp. BH72 (plant-associated, separate phylogenetic cluster).
Metabolic functions

The chromosomal location of genes for anaerobic and aerobic degradation of aromatic compounds and a summary scheme of the involved pathways are shown in Figs. 1 and 2. Detailed pathway descriptions, schemes and annotations are provided as supplementary material.

Anaerobic aromatic degradation

Ten major peripheral anaerobic pathways of catabolism of aromatic compounds are encoded on the chromosome. All but one of these converge at the level of the central intermediate, benzoyl-CoA (Fig. 2). Predicted gene products for the enzymes of these pathways, which were detected in the genome of strain EbN1, are indicated in brackets in the following section.

**Phenylalanine.** Initial transamination of phenylalanine by an aminotransferase (Pat) yields phenylpyruvate, which is usually decarboxylated and oxidized via phenylacetaldheyde to phenylacetate (catalyzed by phenylpyruvate decarboxylase, Pdc, and phenylacetaldehyde dehydrogenase, Pdh). An alternative route from phenylpyruvate directly to phenylacetyl-CoA may proceed via an indolepyruvate:ferredoxin oxidoreductase (IorAB), which also accepts phenylpyruvate as substrate and oxidizes it directly to phenylacetyl-CoA. The genes coding for these catabolic enzymes are not clustered together.

**Phenylacetate.** Anaerobic phenylacetate degradation is initiated by a CoA ligase forming phenylacetyl-CoA (PadJ), which is then oxidized to phenylglyoxylate by a phenylacetyl-CoA:acceptor oxidoreductase (PadBCD). Phenylglyoxyxlate is oxidatively decarboxylated to benzoyl-CoA by a phenylglyoxylate:NAD oxidoreductase (PadEFGHI). The genes for these enzymes are organized in strain EbN1 in an apparent operon like that found for the orthologs of *A. evansii*, and the derived gene products are highly similar between these strains.

**Benzyl alcohol/Benzaldehyde.** Anaerobic oxidation of benzyl alcohol and benzaldehyde to benzoyl-CoA requires substrate-specific benzyl alcohol and benzaldehyde dehydrogenases (Adh and Ald, respectively). Potential genes for these enzymes are located at distant positions on the chromosome.

**p-Cresol.** Oxygen-independent hydroxylation of *p*-cresol to *p*-hydroxybenzaldehyde is catalyzed by the flavocytochrome enzyme *p*-cresol methylhydroxylase (PchCF). Further oxidation to *p*-hydroxybenzoate is catalyzed by *p*-hydroxybenzaldehyde dehydrogenase (PchA). The genes coding for both enzymes are organized in an operon-like structure. Interestingly, strain EbN1 contains a second set of *pchCF*-like genes located directly adjacent to a gene coding for a transposase.

**Phenol.** Anaerobic phenol carboxylation to 4-hydroxybenzoate is achieved via phenylphosphate synthase (PpsABC) and phenylphosphate carboxylase (PpcABCD). The operon-like organization of the corresponding genes is identical to that in *T. aromatica* (Breinig et al. 2000). The orthologs display a high degree of sequence similarity, e.g., 98% identical residues in case of PpcA. A gene for a potential regulator (PdeR) for the *pps* and *ppc* genes is located directly upstream of *ppsA*.

**4-Hydroxybenzoate.** The common intermediate of anaerobic *p*-cresol and phenol metabolism, 4-hydroxybenzoate, also serves as anaerobic growth substrate.
Correspondingly, a putative uptake carrier (PcaK) was detected. Initial activation of 4-hydroxybenzoate by a specific CoA ligase (HbcL) yields 4-hydroxybenzoyl-CoA. A heterotrimERIC molybdenum-enzyme, 4-hydroxybenzoyl-CoA reductase (HcrCBA), then reductively removes the hydroxyl-group, yielding benzoyl-CoA.

**Toluene.** Anaerobic toluene degradation is initiated by a radical-catalyzed addition of the methyl group to fumarate, yielding (R)-benzylsuccinate, which is converted to benzoyl-CoA and succinyl-CoA via a β-oxidation-like reaction sequence. Genes coding for the enzymes of this pathway in strain EbN1 (bss and bbs operons) have been reported recently (Kube et al. 2004).

**Ethylbenzene.** Anaerobic oxidation of ethylbenzene to (S)-1-phenylethanol as initial step of ethylbenzene catabolism is performed by the novel molybdenum enzyme ethylbenzene dehydrogenase (EbdABC). Further degradation to benzoyl-CoA involves a further oxidation step by (S)-1-phenylethanol dehydrogenase (Ped), carboxylation of the resulting acetoephone to benzoylacetate, activation to a CoA-thioester and thiolytic removal of acetyl-CoA. Genes of strain EbN1 involved in this pathway have been reported recently (Rabus et al. 2002). Interestingly, the chromosome harbors a second version of ebd- and ped-like genes. The gene order of these paralogs is reversed and the sequence identities range from 47.2 to 83.5%. Thus, the second operon may code for enzymes for the metabolism of hydrocarbons other than ethylbenzene via a similar pathway.

**Benzoate.** Benzoate is an excellent substrate for anaerobic growth of strain EbN1. Genes for benzoate uptake and metabolism are organized in one large cluster, as known from *T. aromatica* (Boll et al. 2001) and *A. evansii* (GeneBank AJ428529). Benzoate uptake may either proceed via an H⁺-driven symporter (BenK) or an ABC-transporter (EbA5303 to EbA5309), which is probably used at low substrate concentrations. A benzoate-CoA ligase (BclA) synthesizes benzoyl-CoA, which is reductively dearomatized by an ATP-dependent benzoyl-CoA reductase (BcrCBAD). A ferredoxin (Fdx) probably serves as primary electron donor for reduction of the aromatic ring. Subsequent reactions leading to hydrolytic ring cleavage involve an enoyl-CoA hydratase (Dch), a short chain alcohol dehydrogenase (Had) and a ring-opening hydrolase (Oah). The gene products of strain EbN1 are much more similar to that of *A. evansii* than to those of *T. aromatica, Magnetospirillum magnetotacticum* and *Rhodopseudomonas palustris*. Phylogenetic analysis of benzoyl-CoA reductases suggests that the enzymes from strain EbN1 and *A. evansii* form a distinct clade from those of *T. aromatica, M. magnetotacticum* and *R. palustris* (see Fig. 3.1.9C in supplementary material).

**3-Hydroxybenzoate.** A distinct pathway of anaerobic 3-hydroxybenzoate metabolism that does not overlap with benzoate metabolism is predicted in strain EbN1 from the presence of a specific gene cluster which has previously been identified in *T. aromatica* (Laempe et al. 2001). The 3-hydroxybenzoate-catabolic gene clusters of both organisms contain genes for a 3-hydroxybenzoate-CoA ligase (HbcL), enoyl-CoA hydratases and alcohol dehydrogenases. The gene cluster of strain EbN1 also harbors a gene for a putative H⁺-driven symporter for 3-hydroxybenzoate (EbA725) and paralogs of the *bcrABCD* genes coding for benzoyl-CoA reductase. Remarkably, the paralogous ATP-binding “activating” subunits BcrA and BcrD are 100% identical, while the substrate-binding “catalytic” subunits BcrB and BcrC are only 78 and 90% identical, respectively. Thus, strain EbN1 may use different isoenzymes for benzoyl-CoA and
3-hydroxybenzoyl-CoA reduction, in apparent contrast to *T. aromatica*, where only one benzoyl-CoA reductase isoenzyme reduces both intermediates to non-aromatic derivatives (Laempe et al. 2001).

**Aerobic aromatic degradation**

Genes of four different pathways for aerobic metabolism of aromatic compounds were identified in the genome of strain EbN1 (Fig. 2).

**Benzoate.** Genes for a novel pathway of aerobic benzoate degradation were recently described for *A. evansii* and other bacteria (Zaar et al. 2001; Gescher et al. 2002) and are also present in strain EbN1. They include genes for an aerobic benzoate-CoA ligase isoenzyme (BclA), an NADPH- and oxygen-dependent benzoyl-CoA dioxygenase (BoxAB) catalyzing a 2,3-dioxygenation of the aromatic ring, and an enoyl-CoA isomerase (BoxC) and aldehyde dehydrogenase (BoxZ) involved in hydrolytic ring opening. These genes form a cluster in strain EbN1 and their products are about equally similar to the orthologs from *A. evansii* and *T. aromatica*.

**Salicylate/Gentisate.** Genes for the enzymes of a conventional aerobic salicylate degradation pathway were detected in strain EbN1. Salicylate is apparently channeled into the aerobic gentisate pathway by salicylate 5-hydroxylase (S5h). Oxidative cleavage of the gentisate ring is then catalyzed by a gentisate dioxygenase (NagI), yielding maleylpyruvate, which is hydrolyzed to pyruvate and fumarate by maleylpyruvate isomerase (NagL) and fumarylpyruvate hydrolase (NagK). The corresponding genes are apparently duplicated in strain EbN1 (except for *nagK*), and the paralogs are organized in two operon-like structures. The two gene clusters are flanked by genes coding for different regulators and either a putative salicylate transporter or β-oxidation enzymes, suggesting the possible participation of the gene products of the second cluster in a different (still unknown) metabolic pathway.

**Phenylacetate.** Genes for the aerobic metabolism of phenylacetate are known from many bacteria and archaea and were also identified in strain EbN1. The first step is the formation of phenylacetyl-CoA by a specific aerobic phenylacetate-CoA ligase (PaaK). A specific phenylacetyl-CoA dioxygenase (PaaABCDE) then hydroxylates phenylacetyl-CoA to a dihydrodiol intermediate, which has recently been characterized in phenylacetate-utilizing *E. coli* cells (Ismail et al. 2003). The PaaGZ proteins are involved in subsequent double bond isomerization and hydrolytic ring opening. Finally, β-oxidation-like steps (by PaaJFH) lead to succinyl-CoA and acetyl-CoA. Remarkably, strain EbN1 contains two distinct clusters of *paa*-like genes. One of them (divergent *paaZY* and *paaGHIABCDE* genes, located around 2.1 Mb) is identically organized to the *paa* cluster of *A. evansii*, and individual *paa* gene products are highly similar (e.g., 95% identity for PaaA orthologs). The second *paa* cluster (*paaABCDEZJF* orthologs, located around 3.4 Mb) has a different gene content and organization. It is currently not known whether gene products of both clusters are involved in aerobic phenylacetate metabolism or in the degradation of different substrates.

**2-Aminobenzoate.** An unusual pathway for aerobic 2-aminobenzoate metabolism was previously detected in *A. evansii*. A CoA ligase initially forms 2-aminobenzoyl-CoA, followed by its transformation to 2-amino-5-oxo-cyclohex-1-ene-1-carbonyl-CoA by a bifunctional monooxygenase/reductase (Buder et al. 1989). Further degradation steps are unclear at present, but probably involve a β-oxidation-like pathway. An *abmA* gene
coding for the key enzyme of this metabolic pathway, 2-aminobenzoyl-CoA monooxygenase/reductase, is present in strain EbN1, but none of the other genes found in the correlated operons of *A. evansii* (Schühle et al. 2001). However, 2-aminobenzoate is known to be activated as alternate substrate by regular benzoate-CoA ligases (Schühle et al. 2003), and the missing genes for β-oxidation enzymes (Schühle et al. 2001) may also be complemented by genes from other degradative operons.

**Significance of paralogous gene products for aromatic degradation**

In several cases, paralogs were detected for genes involved in anaerobic (*p*-cresol, ethylbenzene and benzoate) and aerobic (salicylate and phenylacetate) aromatic degradation pathways. In all these cases, relatively large sequence deviations between the paralogous gene products (see supplementary material) implicate different substrate specificities or as yet unknown biochemical functions of the second copies, which probably reflects a much broader degradative potential of strain EbN1 than currently known.

**Aliphatic ketones**

Aliphatic ketones (acetone, butanone) as growth substrates of strain EbN1 (Rabus and Widdel 1995) are of particular interest because their degradation is analogous to that of the aromatic substrate and intermediate acetophenone (Fig. 3). However, genome analysis clearly indicates that a different set of enzymes is used for aliphatic ketone metabolism. Acetone metabolism is initiated by an ATP-dependent carboxylase (*ApxABC*) generating acetoacetate. The latter is activated to acetoacetyl-CoA by a succinyl-CoA:3-ketoacid CoA-transferase (*KctAB*) and thiolytically cleaved to two acetyl-CoA. In strain EbN1, the *apxABC* and *kctAB* genes are located in a large gene cluster (about 18 kb), which also contains a gene for a conserved σ^54^-dependent regulator (*AcxR*). The *apxABC* gene products of strain EbN1 show low similarity to the subunits of acetophenone carboxylase (*Apc2-4*; up to 31% similarity), but are much closer related to the subunits of authentic acetone carboxylase (Sluis et al. 2002) from the aerobic bacterium *Xanthobacter autotrophicus* strain Py2 (up to 69% similarity).

**General metabolism**

Degradation of alcohols and organic acids follows general routes, using dehydrogenases and either CoA ligases or CoA-transferases for initial activation of the organic acid intermediates. Propionate is metabolized via the methylmalonyl-CoA pathway, which is probably also involved in degradation of butanone or odd-chained fatty acids yielding propionyl-CoA as an intermediate. The aerobic or anaerobic metabolic pathways of most known substrates of strain EbN1, in particular those of aromatic compounds, yield acetyl-CoA units, which are further oxidized to CO₂ via the citric acid cycle, and the glyoxylate shunt apparently replenishes the citric acid cycle intermediates needed for gluconeogenesis or amino acid biosynthesis.

Strain EbN1 is an obligatory respiratory bacterium, capable of aerobic respiration with oxygen and anaerobic respiration of nitrate (denitrification). Besides the genes for a proton-pumping NADH:ubiquinone oxidoreductase (complex I), strain EbN1 possesses
genes for a non-proton-pumping NADH dehydrogenase (Ndh) and two poorly characterized potential NADH:ferredoxin oxidoreductase isoenzymes of the Rnf type which may be engaged in reverse electron transfer (Kumagai et al. 1997). A cytochrome \(bc_1\) complex III couples electron transfer from ubiquinol to periplasmic cytochromes \(c\) with proton pumping (Hunte et al. 2003) and provides reducing equivalents for the periplasmic reductases involved in denitrification. Strain EbN1 apparently contains an unusually large number of genes for alternative terminal oxidases: (i) four operons coding for proton translocating cytochrome \(aa_3\)-type cytochrome \(c\) oxidases, which typically operate at high oxygen concentrations; (ii) two operons coding for cytochrome \(cbb_3\)-type cytochrome \(c\) oxidases; and (iii) one operon for a cytochrome \(bd\)-type quinol oxidase. The latter two types of oxidases have high affinities for oxygen, qualifying them for microaerobic conditions (Pitcher et al. 2002). The enzymes involved in denitrification are highly similar to those in other denitrifiers (Zumft 1997). They include nitrate/nitrite transporters, a membrane bound nitrate reductase with its active center at the cytoplasmic side of the membrane, a periplasmic cytochrome \(cd_1\)-type nitrite reductase, a membrane-bound nitric oxide reductase whose active center is oriented towards the periplasm, and a periplasmic nitrous oxide reductase (Zumft 1997). The proton gradient generated by the enzymes of the various respiratory systems is exploited by an \(F_1F_0\)-type ATP-synthase for the regeneration of ATP. The availability of several NADH-oxidizing isoenzymes and terminal oxidases may enable strain EbN1 to finely tune its respiratory chain according to the environmental conditions.

All genes for the components required for uptake of inorganic nutrients (e.g., ammonia, nitrate, sulfate or phosphate) and for amino acid and nucleotide biosynthesis and most genes for the enzymes of cofactor biosynthesis were identified on the chromosome. The genome of strain EbN1 contains a large number of genes coding for metallo-enzymes.

Several detoxification systems were detected, e.g., a flavorubredoxin (NorVW) reducing toxic NO to \(N_2O\), a superoxide dismutase and a catalase, and potential systems for arsenate and tellurite resistance.

**Lack of nitrogenase and components for symbiotic association with plants**

The apparent inability of strain EbN1 to fix nitrogen (unpublished) correlates with the absence of genes for nitrogen fixation (Brill 1980), except for regulatory proteins such as NifLA (Egener et al. 2002) which probably serve additional functions in ammonia assimilation. This is in marked contrast to the diazotrophic growth of the closely related plant-associated *Azoarcus* sp. strain BH72 which contains typical genes for \(N_2\)-fixation (Egener et al. 2001). Strain EbN1 also differs from *Azoarcus* sp. strain BH72 by its ability to initiate symbiosis with plants. A genome-based explanation for that phenotypic difference may be the apparent absence of homologous cell surface proteins in strain EbN1. The presence of such proteins on cells of strain BH72 is a prerequisite for association with the mycelium of an ascomycete isolated from the rhizosphere of Kallar grass and the subsequent colonization of host plants associated with nitrogen fixation (Dörr et al. 1998).
Cellular functions

Details on cellular functions (DNA replication, recombination and repair; transcription; translation; cell envelope biosynthesis; cell division; chemotaxis; motility and secretion) are provided as supplementary material.

Sensory functions and regulation of gene expression

About 120 genes for DNA-binding regulatory proteins belonging to 18 different protein classes are present in the genome of strain EbN1. Among these are genes for seven members of the Fnr/Crp family (Körner et al. 2003), two of them coding for genuine orthologs of the oxygen sensor Fnr and the catabolite sensor Crp, respectively. The five other gene products are similar to Dnr and Nnr regulators, which regulate expression of denitrification genes in pseudomonads (Zumft 2002); these gene products may have a similar role in strain EbN1, as suggested by the proximity of some of their genes to those for denitrification enzymes. Genes for at least 30 two-component regulatory systems were detected on the chromosome. Some of them can be correlated to specific functions, such as NtrBC to ammonium assimilation, the DctRS to dicarboxylate uptake, or NarXL to nitrate sensing. For three other two-component systems, regulatory functions in anaerobic toluene, ethylbenzene or acetophenone degradation were recently suggested (Rabus et al. 2002; Kube et al. 2004; Kühner et al. 2004).

Other cellular functions

With only few exceptions, all genes required for prototrophic growth have been identified (for details refer to supporting information). A type II ribonucleotide reductase (Stubbe et al. 2001) synthesizes deoxyribonucleotides under aerobic or anoxic conditions. Another notable feature of strain EbN1 is the apparent lack of genes required for assembly and function of flagella, which agrees with previous microscopic observation of the cells as non-motile. On the other hand, about 30 genes for structural and assembly components of a type IV pilus-dependent twitching motility system and associated chemotactic functions were detected in the genome of strain EbN1 (Nudleman and Kaiser 2004).

Discussion

The genome sequence of strain EbN1 is the first of an anaerobic aromatics-degrading member of the Betaproteobacteria and may provide a model of heuristic value for the study of the genomic organization and evolution of genes for anaerobic aromatic degradation in bacteria. The detection of genes for anaerobic phenol or 3-hydroxybenzoate metabolism or for 2-aminobenzoyl-CoA monooxygenase/reductase expands hitherto known degradative capabilities of strain EbN1 (Rabus and Widdel 1995). Most catabolic genes involved in a common catabolic pathway cluster together. In contrast, many of the genes for anabolic enzymes are scattered across the chromosome. The enzymes of central anaerobic benzoyl-CoA metabolism are most similar to those of the closely related A. evansii, whereas many enzymes of degradative peripheral pathways
(e.g., for anaerobic toluene or phenol metabolism) show a similar degree of conservation with those of the more distantly related *Thauera* strains. Particularly the latter observation suggests that some catabolic gene clusters may have been acquired by horizontal gene transfer. Such an assumption is supported by several lines of evidence. (i) A strikingly high number of genes for transposases and integrases were detected in the genome of strain EbN1, sometimes closely adjacent to catabolic gene clusters. (ii) Plasmid 2 harbors several genes coding for potential enzymes of aromatic metabolism and β-oxidation. (iii) The proximity of the genes for toxin/antitoxin proteins of “addiction systems” to genes for transposases may be interpreted as additional safeguarding system to stabilize maintenance of acquired mobile genetic elements. One may speculate that horizontal gene transfer is a general strategy of bacteria with broad degradative capacities to achieve rapid pathway evolution and hence nutritional competitiveness. A comparative analysis of the genomes of other anaerobic aromatics-degrading members of the *Betaproteobacteria*, but also of the *Deltaproteobacteria*, where aromatics degradation is rather common, will be highly interesting as soon as genomes with comparable gene products become available.

Strain EbN1, its close relatives and many *Thauera* spp. have been isolated from freshwater and soil habitats, while the *Azoarcus* spp. (*sensu stricto*) have been found in symbiotic association with plants (Reinhold-Hurek and Hurek 2000). The complete absence in strain EbN1 of genes related to nitrogen fixation and adhesins needed for symbiotic interaction documents profound ecophysiological differences between strain EbN1 and the N₂-fixing plant symbiont *Azoarcus* sp. BH72. Because of such fundamentally different life styles, strain EbN1 and the closely related aromatics degraders should not be classified as *Azoarcus* spp. but rather as an own new genus (name to be agreed upon). Above all, a future comparison of the genomes of strain EbN1 and a plant-associated *Azoarcus* species such as strain BH72 (Reinhold-Hurek and Hurek 2000) is desirable since it is expected to provide new insights into mechanisms that may have allowed transition from a free-living soil bacterium to a plant-associated symbiont.

**Acknowledgements** We acknowledge Katja Borzym, Ines Müller, Janina Thiel, Susan Böhm, Sven Klages, Beatrice Baumann, Daniela Gröger, Verena Gimmel, Mario Sontag, Maximilian Weiss (Berlin) and Daniela Lange (Bremen) for technical assistance, and Dirk Jacobs (Berlin) for computational support with the HTGA system. This study was supported by the Max Planck Society.
References


Complete genome sequence of strain EbN1
Complete genome sequence of strain EbN1


Table 1 General features of the strain EbN1 genome

<table>
<thead>
<tr>
<th>Genome features</th>
<th>Chromosome</th>
<th>Plasmid 1</th>
<th>Plasmid 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (bp)</td>
<td>4,296,230</td>
<td>207,355</td>
<td>223,670</td>
</tr>
<tr>
<td>G+C content</td>
<td>65.12</td>
<td>57.63</td>
<td>63.11</td>
</tr>
<tr>
<td>Stable RNAs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rRNAs</td>
<td>4 rrn operons</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>tRNAs</td>
<td>58</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Coding regions(^a)</td>
<td>4,133</td>
<td>274</td>
<td>196</td>
</tr>
<tr>
<td>Coding (%)(^b)</td>
<td>90.9</td>
<td>89.4</td>
<td>85.1</td>
</tr>
<tr>
<td>Average size (bp)(^c)</td>
<td>945</td>
<td>676</td>
<td>972</td>
</tr>
<tr>
<td>Assigned function(^d)</td>
<td>2,560</td>
<td>71</td>
<td>141</td>
</tr>
<tr>
<td>Conserved unknown(^e)</td>
<td>628</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>Predicted novel(^f)</td>
<td>945</td>
<td>178</td>
<td>45</td>
</tr>
</tbody>
</table>

\(^a\)Number of predicted ORFs for that DNA molecule  
\(^b\)Percentage of coding sequence for that DNA molecule  
\(^c\)Average ORF size  
\(^d\)Number of predicted ORFs with assigned functions  
\(^e\)Number of predicted ORFs with strong similarity to genes with unknown functions from other organisms  
\(^f\)Novel ORFs without assigned functions
Fig. 1 Structural representation of the chromosome. Circles (from outside to inside): First, distribution of catabolic genes (red, anaerobic aromatic degradation; blue, aerobic aromatic degradation; black, other metabolic functions); Second, integrases (green), insertion sequences (brown) and transposases (orange); Third, predicted orfs (COG-colours, or black if no COG was assigned); Fourth, G+C skew; Fifth, tRNAs (red), rRNAs (green) and ribosomal proteins (blue); Sixth, scale [Mb] of the chromosome