Whole-genome analysis of *Azoarcus* sp. strain CIB provides genetic insights to its different lifestyles and predicts novel metabolic features

Zaira Martín-Moldes[a1], María Teresa Zamarro[a1], Carlos del Cerro[a1], Ana Valencia[a], Manuel José Gómez[b2], Aida Arcas[b3], Zulema Udaondo[a4], José Luis García[a], Juan Nogales[a], Manuel Carmona[a], and Eduardo Díaz[a*]

[a] *Centro de Investigaciones Biológicas-CSIC, 28040 Madrid, Spain*
[b] *Centro de Astrobiología, INTA-CSIC, 28850 Torrejón de Ardoz, Madrid, Spain*

[1] These authors contributed equally to this work

[2] Present address: Centro Nacional de Investigaciones Cardiovasculares, ISCIII, Madrid, Spain
[3] Present address: Instituto de Neurociencias, UMH-CSIC, Alicante, Spain
[4] Present address: Abengoa Research, Sevilla, Spain

[*]Corresponding author. Tel.: +34915611800. *E-mail address*: ediaz@cib.csic.es (E.Díaz)

Abbreviations: ANI, average nucleotide identity; BIMEs, bacterial interspersed mosaic elements; IAA, indoleacetic acid; ICE, integrative and conjugative element; REPs, repeated extragenic palindrome sequences; ROS, reactive oxygen species; TAS, toxin-antitoxin system; TMAO, trimethylamine *N*-oxide.

ABSTRACT

The genomic features of *Azoarcus* sp. CIB reflect its most distinguishing phenotypes as a diazotroph, facultative anaerobe, capable of degrading either aerobically and/or anaerobically a wide range of aromatic compounds, including some toxic hydrocarbons such as toluene and *m*-xylene, as well as its endophytic lifestyle. The analyses of its genome have expanded the catabolic potential of strain CIB towards common natural compounds, such as certain diterpenes, that were not anticipated as carbon sources. The high number of predicted solvent efflux pumps and heavy metal resistance gene clusters has provided the first evidence for two environmentally-relevant features of this bacterium that remained unknown. Genome mining has revealed several gene clusters likely involved in the endophytic lifestyle of strain CIB, opening the door to the molecular characterization of some plant growth promoting traits. Horizontal gene transfer and mobile genetic elements appear to have played a major role as a mechanism of adaptation of this bacterium to different lifestyles. This work paves the way for a systems biology-based understanding of the abilities of *Azoarcus* sp. CIB to integrate aerobic and anaerobic metabolism of aromatic compounds, tolerate stress conditions, and interact with plants as an endophyte of great potential for phytostimulation and phytoremediation strategies. Comparative genomics provides an *Azoarcus* pan genome that confirms the global metabolic flexibility of this genus, and suggests that its phylogeny should be revisited.

Keywords: *Azoarcus*, aromatic compounds, endophyte, metals resistance, mobile genetic elements, comparative genomics

Introduction

Azoarcus is a genus of betaproteobacteria that belongs to the family *Rhodocyclaceae*, a physiologically versatile group encompassing bacteria with diverse functions [60]. The environmental relevance of *Azoarcus* strains is supported by their frequent detection in diverse soils, sludges, and wastewaters [39]. The *Azoarcus* genus, that includes ten recognized species, namely *A. indigens* (type species of the genus), *A. communis, A. tolulyticus, A. toluvorans, A. toluclasticus, A. evansii, A. anaerobius, A. buckelii, A. olearius, and A. taiwanensis* [2,31,42,49], was shown to comprise bacteria that fit into one of two major phylogenetic and eco-physiological groups [22,28,38,49]. One group includes free-living bacteria that usually inhabit waters and soils and participate in the biogeochemical cycling of a large number of organic and inorganic metabolites [31,40,43,49]. Many strains of this group have been described and/or isolated by their ability to degrade aromatic compounds in anoxic conditions, being strain EbN1 (currently *"Aromatoleum aromaticum"* EbN1) [38,39] and *A. evansii* KB740 [2,14] the two most studied. The other group includes *Azoarcus* strains such as *A. communis* strain SWub3, *A. indigens* strain VB32 or the well-characterized *Azoarcus* sp. strain BH72, that invade roots of Kallar grass and rice, living as endophytic bacteria [42]. Interestingly, the free-living *Azoarcus* strains that are anaerobic biodegraders of aromatic compounds, and whose prototype is the EbN1 strain, have exclusively received particular attention for their degradation and biotransformation abilities [22,38,39].

Azoarcus sp. CIB (CECT#5669) was isolated from a DSMZ 12184 culture (not available anymore), which was supposed to be *Azoarcus* sp. strain M3, isolated from a diesel fuel-contaminated aquifer at Menziken (Switzerland) [20,32]. *Azoarcus* sp. CIB is a free-living bacterium with the ability to degrade a high number of aromatic compounds under aerobic and/or anaerobic conditions, including some toxic hydrocarbons such as toluene and *m*-xylene [9,25,32,57]. Recently, we have demonstrated that *Azoarcus* sp. CIB has also the ability to grow in association with plants, colonizing the intercellular spaces of the exodermis of rice roots. In addition, the strain CIB displays plant growth promoting traits such the ability to uptake insoluble phosphorous, production of indoleacetic acid (IAA) or nitrogen fixation [16]. Thus, *Azoarcus* sp. CIB may represent the prototype of a subgroup of *Azoarcus* strains that share the anaerobic biodegradation of aromatic hydrocarbons with a facultative endophytic lifestyle [16]. Since *Azoarcus* sp. CIB presents a robust growth and it is susceptible of genetic manipulation, it became a model system to study the complex regulatory networks that control the expression of the aerobic and anaerobic aromatic degradation clusters [5,9,13,57,58], and some recombinants strains have been engineered for biotechnological prospects [63].

In this work, we sequenced the whole genome of *Azoarcus* sp. CIB and accomplished a comparative analysis with the genomes of other strains, such as *Azoarcus* sp. BH72 and strain EbN1, that are the prototypes of obligate endophytes and free-living strains, respectively. This work provides new insights into the genetic determinants that may account for some of the reported metabolic abilities of the CIB strain, and offers information on genetic characteristics that may be relevant for the adoption of a particular lifesytle or that can be of biotechnological interest.

Materials and Methods

Genome sequence, contigs assembling, and gaps filling

Azoarcus sp. CIB was anaerobically grown at 30 ºC in MC medium [32] containing 3 mM benzoate as sole carbon and energy source and 10 mM nitrate as electron acceptor. Cultures were collected when they reached the early stationary phase and genomic DNA was extracted using previously published protocols [32].

The genome sequencing of *Azoarcus* sp. CIB was carried out using the 454 Life Sciences high-density pyrosequencing methodology in a GSFLX sequencer from Roche at LifeSequencing (Valencia, Spain). FASTQ reads (about 250-nt long) were assembled in contigs by using the Newbler software from Roche. Contigs were ordered in scaffolds by performing a long-tag paired-end sequencing according to Roche protocols at LifeSequencing (Valencia, Spain). Gap filling on the scaffolds was performed by manual assembly of FASTQ reads with BioEdit (Ibis Biosciences) and by conventional sequencing methods (ABI Prism 377; Applied Biosystems) of PCR products (purified with Gene Clean Turbo, Q-BIOgene) spanning the regions between flanking contigs.

Gene prediction and genome annotation

The genome of *Azoarcus* sp. CIB was annotated by means of a bacterial genome annotation pipeline [56], which used tRNAscan-SE to predict tRNA genes, RNAmmer to predict rRNA genes and Glimmer to predict coding sequences. Functional annotations for proteins were generated by comparison against several protein sequence and protein family databases (SwissProt, NCBI protein, COG, Pfam, Smart, Prk) with BLAST and RPS-BLAST [1]. Annotations were summarized in different output formats. One of them was used as input for Pathway Tools, for automatic metabolic reconstruction [27].

Transposase and integrase encoding genes were manually annotated with the assistance of the ISFinder database (http://www-is.biotoul.fr/).

Comparative genomics

The complete genome of *Azoarcus* sp. CIB was compared with that of all currently sequenced strains: "*Aromatoleum aromaticum*" strain EbN1 (NC_006513.1; NC_006823.1; NC_006824.1); *Azoarcus* sp. BH72 (NC_008702.1); *Azoarcus* sp. KH32C (NC_020516.1; NC_020548.1); and *Azoarcus toluclasticus* strain MF63 (NZ_ARJX00000000.1). The closely related *Thauera* sp. strain MZ1T (CP001281.2, CP001282.1) was used as outgroup. Comparative genomic analyses, including Venn diagrams, synteny analyses and phylogenetic trees were performed with EDGAR [6]. Average nucleotide identity among the genomes based on MUMmer (ANIm) was calculated with Jspecies [44].

Nucleotide sequence accession number

The *Azoarcus* sp. strain CIB whole genome sequence and annotation has been deposited in GenBank and is available under accession number CP011072.

Substrate diversity studies

Azoarcus sp. CIB was tested for its ability to utilize various carbon sources in MC medium in aerobic conditions or anaerobiosis (10 mM nitrate as electron acceptor) when cultured for 48 h at 30ºC. For each substrate, two replicates and a control without inoculation were included. Aromatic compounds that had not been tested previously and that allowed growth were: cinnamate and *p*-coumarate (anaerobiosis); gentisate and cumene (aerobiosis). Other aromatics that did not allow growth were: 2-hydroxybenzoate, 3-hydroxyphenylacetate, 2-hydroxyphenylacetate, 2-hydroxyphenylpropionate, protocatechuate, catechol, homogentisate, resorcinol, ferulate, vanillin, vanillate, nicotinate, isonicotinate, *o*-phthalate, mandelate, phenylglyoxylate, tyramine, 2-aminobenzoate, 4-aminobenzoate, 3-fluorobenzoate, 2-chlorobenzoate, 3-chlorobenzoate, benzene, propylbenzene, styrene, biphenyl. Non aromatic carbon sources tested that provided growth were: acetate, citrate (aerobiosis) propionate, lactate (anaerobiosis), pyruvate, butyrate, 3-hydroxybutyrate, isobutyrate, valerate, isovalerate, succinate, fumarate, malate, glutarate, adipate, pimelate, Ala, Pro, Glu, Asp, Arg, Asn, Gln, Ile, Leu, Gly, Val, His, Ser, and abietic acid (aerobiosis). Non aromatic carbon sources tested that did not allow growth were: fructose, maltose, glycerol, galactose, arabinose, sacarose, manitol, ribose, xylose, Thr, Trp, Lys, maleate, quinate, limonene, geraniol, citronellol, cyclohexanol, cyclohexanone, cholesterol, isopropanol, butanol, propanol, octanoate, decanoate.

Results and discussion

Genome organization

The complete genome of *Azoarcus* sp. CIB was constructed as a single circular chromosome consisting of 5,257,030-bp by assembling sequence data from pyrosequencing and Sanger sequencing of PCR products. A total of 872,867 sequencing reads were obtained from three DNA pyrosequencing runs and a paired-end run, providing a 42-fold coverage of the genome. The sequence was assembled into 645 contigs with an average length of 9748 bp, which were, in turn, organized in 38 scaffolds. The predicted size of the *Azoarcus* sp. CIB chromosome was 5,145,765 bp, based on those sequences. Manual curation of the assembly allowed to fill-in 349 out of 655 gaps. The remaining 306 gaps were filled-in by conventional sequencing of PCR products obtained by using primers that were designed to the ends of each of the remaining contigs. Gap filling added 111,265 bp to the initial genome assembly.

Table 1 shows basic characteristics of the genome of *Azoarcus* sp. strain CIB and their comparison with those of the other four sequenced *Azoarcus* strains, i.e., EbN1 [38], BH72 [28], KH32C [37], and MF63 (BioProject PRJNA199393). The chromosome contains 4 complete 16S-tRNA^{Ile}-tRNA^{Ala}-23S-5S operons (Fig. 1). A total of 2347 genes (49.5 % of the total protein-coding genes) of *Azoarcus* sp. CIB were assigned to a functional category of clusters of orthologous groups of proteins.

The genomic structures of *Azoarcus* sp. CIB and *Azoarcus* sp. KH32C strain were syntenic, except in the region from position 1.0 to 3.0 Mb (Fig. S1). Many of the breakpoints in synteny corresponded to the presence or absence of integrated elements. However, synteny was not extended to the EbN1 and BH72 strains (data not shown), suggesting a more distant phylogenetic relation for these two strains.

The mobilome of Azoarcus sp. CIB

A notable genomic feature of strain CIB is the presence of an extensive mobilome. Thus, the genome of *Azoarcus* sp. CIB contains a high number of full-length and partial genes encoding putative transposases (85) and phage-related integrases/recombinases (16) that belong to a wide variety of families and many of which are paralogs (Table S1). All these genes are distributed across the chromosome but appear to be particularly densely clustered at some genome hot spots (Fig. 1).

Out of the 14 toxin-antitoxin systems (TASs) present in the genome of strain CIB, seven are located in close proximity to genes encoding transposases (Fig. 1), which suggests a major role in the maintenance of mobile genetic elements. Other TASs may play important roles in cell physiology of strain CIB under stress conditions, as it has been shown in other bacteria [59].

Several site-specific integrated elements have been identified in the genome of strain CIB (Table 2). Some of them contain site-specific recombinases or transposases, and/or they are targeted to tRNAs. Moreover, most of these integrated elements show a GC content that is significantly lower or higher than the average GC content (65.8%) of the *Azoarcus* sp. CIB genome. All these features suggest that these integrated elements are genomic islands that were acquired by horizontal gene transfer. These genomic islands are likely involved in the adaptation of *Azoarcus* sp. CIB to the use of different carbon sources or the endophytic lifestyle (see below). The presence of orthologs of some of the genes contained in the genomic islands of strain CIB in the genome of *Azoarcus* strains BH72, EbN1, KH32C and MF63, is summarized in Table 2.

Repeated extragenic palindrome (REPs) sequences, often clustered as two tandem inverted copies to form bacterial interspersed mosaic elements (BIMEs) [53], are present in high copy number in the genome of strain CIB (Table S2). REPa/BIMEa elements are frequently associated to translation and/or transcription termination signals (Table S2). We have identified two similar REPa sequences in the gene clusters for the aerobic degradation of benzoate and 2-aminobenzoate in *A. evansii* [19,46], and there are five REPa sequences in the chromosome of strain EbN1 [38]. Interestingly, a couple of REPa elements flank transposase $TnpA_{REP}$ (AzCIB_4604) which is associated to insertion sequences of the IS*200*/IS*605* family (Fig. 1). $TnpA_{REP}$, which belongs to the Y1-superfamily of transposases, occurs in a variety of bacteria and appears to be responsible for REP proliferation throughout their host genome [53]. The $TnpA_{REP}$ from *Azoarcus* sp. CIB constitutes the first one to be identified within the *Rhodocyclaceae* family.

BIMEb elements are commonly overlapping the stop codons of adjacent genes (Table S2). We have identified four REPb sequences in the gene clusters for the aerobic degradation of benzoate [19] and anaerobic degradation of phenylacetate (Acc. No. AJ428571) in *A. evansii*, and there are 15 REPb sequences in the genome of strain MF63. Most BIMEc elements are located in intergenic regions (Table S2). The BIMEc element is also present in high copy number in *Azoarcus* sp. BH72, and a truncated REPc sequence (34-nt long) can be identified in the genome of strains of *Vibrio, Variovorax, Rhanella, Rhodopseudomonas*, etc., suggesting that this element is a new REP/BIME sequence widely spread in proteobacteria.

Phylogeny of Azoarcus strains

Comparison of the whole genome of *Azoarcus* sp. CIB with that of other strains whose genome is known, i.e., *Azoarcus toluclasticus* MF63, *Azoarcus* sp. KH32C,

Azoarcus/"Aromatoleum" strain EbN1, and *Azoarcus* sp. BH72, gives average nucleotide identity (ANIm) values [44] of 90.35% (percentage of genome aligned 73.73%), 85.96% (aligned 34.76%), 85.22 (aligned 24.72%), and 84.25% (aligned 15.48%), respectively. As a reference, the ANIm value between *Azoarcus* sp. CIB and a strain of the closely related *Thauera* genus, *T. aminoaromatica* MZ1T strain [23], was 84.34% (aligned 18.47%). Therefore, these data strongly suggest that strain CIB is indeed a bacterial species different to those sequenced so far. Taken into account previous phylogenetic analyses of the 16S rDNA sequence of strain CIB and that of the other *Azoarcus* species known so far [16,32], *Azoarcus* sp. CIB would be closely related to *A. evansii/A. toluvorans* species, which constitute the *A. toluvorans* species group.

Comparative genomic analyses were performed to determine putative orthology relations between the sequenced *Azoarcus* strains. A Venn diagram representing the core genome and the pan genome of the *Azoarcus* genus is shown in figure 2. 1701 genes are shared among the five *Azoarcus* strains. The strain CIB shares 3643 genes with strain MF63, 3060 genes with strain KH32C, 2395 genes with strain EbN1, and 2360 genes with strain BH72 (Fig. 2). A total of 475 genes were identified as exclusive, singletons, to strain CIB.

A core genome tree among the five compared strains placed *Azoarcus* sp. CIB closest to *A. toluclasticus* MF63, with strains BH72 and EbN1 as the most distantly related to strain CIB (Fig. 3A). This result is supported by the analysis of the average amino acid identities of the *Azoarcus* core genome (Fig. 3B). Taken together, these data confirm the ANI analyses shown above, and suggest that the phylogeny of the *Azoarcus* genus should be revisited in the near future to presumably re-classify the strains into at least three different genera, i.e., a first genus comprising the strains closely related to species *A. buckelii, A. anaerobius, A. taiwanensis*, and "*Aromatoluem aromaticum*", a second genus comprising the strains closely related to species *A. tolulyticus, A. toluclasticus*, and *A. evansii/A. toluvorans*, and a third genus comprising the strains closely related to species *A. communis, A. indigens*, and *A. olearius*. Further genome analyses have been focused on genetic determinants involved in the major ecophysiological capacities of strain CIB.

Genes for nitrogen metabolism

Azoarcus sp. CIB is able to fix nitrogen as a diazotroph [16], which was confirmed through genome analyses by the presence of a nitrogen fixation regulon. Thus, a *nif* gene cluster encoding the synthesis and maturation of the nitrogenase required to fix nitrogen is closely linked to another cluster that likely encodes a RnfI membrane complex driving a reverse electron flow from NADH to reduce ferredoxin FdxN, which serves as the electron donor to nitrogenase [45]. Both the *nif* and *rnfI* gene clusters are located within genomic island III of strain CIB (Table 2, Fig. 1). The ability of *Azoarcus* sp. CIB to fix nitrogen should give this organism an advantage for survival in environments poor in nitrogen, and should favor plant associations (see below). By growing the strain CIB aerobically on a minimal medium containing nitrate as sole nitrogen source, we have confirmed here the existence of a functional assimilatory nitrate reductase in *Azoarcus* sp. CIB (data not shown). Genome analyses revealed the existence of potential assimilatory nitrate (AzCIB_0744) and nitrite (AzCIB_0740-0741) reductases, respectively, encoded in a *nas* gene cluster [33].

On the other hand, *Azoarcus* sp. CIB is able to use nitrate as terminal electron acceptor in the absence of oxygen, when coupled to the oxidation of organic compounds [32]. Based on genome analyses we predict a complete denitrification that generates

nitrogen and that involves a nitrate reductase (*nar* genes; AzCIB_2181-2184), a nitrite reductase (*nir* genes, AzCIB_3596-3608), a nitric oxide reductase (*nor* genes, AzCIB_1316-1320), and a nitrous oxide reductase (*nos* genes, AzCIB_2188-2193) [65].

Genes for aromatic compounds degradation

The capability of *Azoarcus* sp. strain CIB to degrade a wide variety of aromatic compounds, both aerobically and anaerobically, is one of most relevant metabolic features of this strain and it has been subject of several studies [9,25,32]. This metabolic potential was confirmed here through genome sequence analyses. About 6% of the total protein-encoding genes are related to the catabolism of aromatic compounds or chemicals that generate aromatic compounds during their degradation.

Three different central pathways for anaerobic degradation of aromatic compounds can be identified in strain CIB. Thus, in addition to the previously reported gene clusters for anaerobic degradation of benzoate (*bzd* genes) and 3-methylbenzoate (*mbd* genes) through the bzd and mbd central pathways, respectively (Fig. 4) [25,32], the strain CIB contains an additional gene cluster (*hbd* genes) orthologous to that responsible for the anaerobic degradation of 3-hydroxybenzoate in *T. aromatica* [30], and strain EbN1 [38] and, hence, likely to be devoted to 3-hydroxybenzoate degradation in *Azoarcus* sp. CIB (Fig. 4).

A number of peripheral pathways that converge into the anaerobic benzoyl-CoA central pathway can be identified in the genome of strain CIB. The genes encoding the putative peripheral pathways for the anaerobic degradation of 4-hydroxybenzoate (*hcr*), phenol (*pps/ppc*), *p*-cresol (*pch/pdh1*), phenylacetate/4-hydroxyphenylacetate (*pad*), 2-phenylethylamine (*pea*), phenylalanine (*pat/pdc*), 2-phenylethanol (*ped*), phenylacetaldehyde (*pdh/aor*), phenylpropionate/cinnamate/*p*-coumarate (*cou*) are found distributed along the *Azoarcus* sp. CIB chromosome (Fig. 4) [4, 9, 38, 39]. The *cou* genes are also responsible of funneling 3-hydroxyphenylpropionate to the hbd central pathway (Fig. 4). Interestingly, a gene cluster orthologous to the *bss/bbs* cluster encoding the anaerobic peripheral pathway that funnels the aromatic hydrocarbons toluene and *m*-xylene to the bzd and mbd central pathways, respectively [9,25,29], has been found within a putative integrative and conjugative element, ICE_{XTD}, in the genome of *Azoarcus* sp. CIB (Figs. 1 and 4, Table 2).

Azoarcus sp. CIB is also able to degrade aromatic compounds under aerobic conditions [32]. The aerobic degradation of benzoate via the benzoyl-CoA hybrid pathway is encoded by the *box* gene cluster [57]. Genome mining revealed a number of additional gene clusters likely encoding different central pathways for the aerobic degradation of aromatic compounds. The *paa* gene cluster encoding a second aerobic hybrid pathway, i.e., the phenylacetyl-CoA pathway for the aerobic degradation of phenylacetate [52], was identified (Fig. 4). On the other hand, gene clusters orthologous to those encoding a catecholic *meta*-cleavage pathway for aerobic degradation of toluene/cumene (*tod*) [10], and the central gentisate pathway for 3-hydroxybenzoate degradation (*nag*) [64], were also present in the chromosome of strain CIB (Fig. 4). It should be noted that the *tod* genes are also present within the ICE_{XTD} element, suggesting that the ability of strain CIB to degrade either aerobically or anaerobically toxic aromatic hydrocarbons has been acquired by horizontal gene transfer. Genes that might encode peripheral pathways that channel other aromatic compounds, e.g., 2-phenylethylamine (*pea*) and 3-hydroxyphenylpropionate (*cou*), to the paa and nag aerobic central pathways, respectively, are shared with the corresponding anaerobic degradation pathways (Fig. 4).

In bacteria, some cyclic non-aromatic compounds, e.g., cyclohexane carboxylate, diterpenes, steroids, are channeled to aromatic central pathways through their degradation [18,34,48]. *Azoarcus* sp. CIB is able to degrade cyclohexane carboxylate [5], and genes homologous to the *bad* genes responsible of bacterial cyclohexane degradation via an alternative anaerobic central pathway [9,15] have been identified in the strain CIB (Fig. 4). On the other hand, we have confirmed here (see the "Substrate diverstity studies" section in Materials and Methods) that the CIB strain uses abietic acid, a model diterpene, as sole carbon source under aerobic conditions. It is known that the *dit* genes are responsible for diterpenes degradation in some bacteria via initial hydroxylating cytochrome P450s (peripheral pathway) and further *meta*-cleavage of the aromatic intermediate (central pathway) [48]. The *dit* genes are usually located within ICE elements [34], and we have found orthologous *dit* genes within a putative ICE_{DIT} element in the genome of strain CIB (Figs. 1 and 4, Table 2). Nevertheless, further work should be done to demonstrate that the predicted *dit* genes are involved in abietic acid degradation in strain CIB.

Steroids have been shown to generate aromatic intermediates during their degradation [18]. In the genome of strain CIB we have found orthologs of the *tesEFG* genes (Fig. 4A), which encode a *meta*-cleavage lower pathway [21], linked to genes encoding several putative steroid dehydrogenases. Since these genes are located at genomic island VI of *Azoarcus* sp. CIB (Table 2, Fig. 1), they appear to be recruited by this strain and some related *Azoarcus* strains, e.g, MF63 (Table 2), as an adaptation mechanism to metabolize certain steroids.

Genes for the metabolism of other carbon sources

Substrate diverstity studies accomplished here (see the "Substrate diverstity studies" section in Materials and Methods) revealed that the major carbon sources, other than aromatic compounds, for *Azoarcus* sp. CIB are monocarboxylic acids (acetate, propionate, lactate, pyruvate, butyrate, 3-hydroxybutyrate, isobutyrate, valerate, isovalerate), dicarboxylic acids (succinate, fumarate, malate, glutarate, adipate, pimelate), and amino acids (Ala, Pro, Glu, Asp, Arg, Asn, Gln, Ile, Leu, Gly, Val, His, Ser). The genes encoding all enzymes and isoenzymes of the tricarboxylic acid cycle are distributed across the entire genome of strain CIB. Moreover, *Azoarcus* sp. CIB contains the *aceA* and *aceB* genes encoding the isocitrate lyase and malate synthase, respectively, responsible for a glyoxylate shunt that is required when growth is based on the utilization of substrates, such as fatty acids or aromatic compounds, that generate acetyl-CoA units, which are then channeled into anabolic pathways such as gluconeogenesis.

The genome of *Azoarcus* sp. CIB appears to be devoid of a classical glycolysis pathway (Embden-Meyerhoff pathway) because a *pfk* gene encoding phosphofructokinase could not be identified. However, all genes needed for a gluconeogenesis pathway are present in the CIB genome. Interestingly, a cluster containing the *zwf* (glucose-6-P 1-dehydrogenase), *pgl* (6-phosphogluconolactonase), *edd* (phosphogluconate dehydratase), and *eda* (2-dehydro-3-deoxyphosphogluconate aldolase) genes that encode an Entner-Doudoroff pathway for the degradation of glucose-6-phosphate, is present in the genome of strain CIB. A similar gene cluster is also present in *Azoarcus* sp. KH32C and MF63 strains, but it is lacking in BH72 and EbN1 strains. Although the *gnd* gene of the oxidative branch of the pentose phosphate pathway appears to be absent, the genes encoding the non-oxidative branch of the pentose phosphate pathway are present in the genome of strain CIB, as expected from

its requirement for the biosynthesis of pentose sugars. However, despite the enzymatic potential of strain CIB to degrade certain sugars (e.g., glucose, gluconate, glycerol), analysis of its genome revealed the absence of genes encoding active transporters for these carbon sources, which would explain why *Azoarcus* sp. CIB does not generally utilize carbohydrates as growth substrates. The inability to utilize exogenous carbohydrates appears to be a common feature within some strains of the *Azoarcu*s genus [28,38]. The fact that these *Azoarcus* strains contain genes likely to encode the degradation of some sugars but they do not use such compounds when supplied as carbon sources could reflect that these genes are devoted to use sugar-phosphates derived of the turnover of endogenous polysaccharides from cellular envelopes or storage material.

Two different gene clusters, AzCIB_4593-4594 and AzCIB_0059-0062, encoding putative formate dehydrogenases for the oxidation of formate to CO_2 [24] have been identified in the genome of strain CIB. The role of these predicted formate dehydrogenases could be to recycle the formate generated during the metabolism of certain carbon sources, such as in the aerobic degradation of benzoate [19,57].

Genes coding for potential solvent and heavy metals resistance

We have identified in the genome of strain CIB several gene clusters that may encode RND-type efflux systems analogous to the AcrAB/TolC-like systems of strain EbN1 that were shown to be induced in the presence of aromatic solvents [62] and to the TtgABC-like systems involved in aromatic hydrocarbons tolerance in *Pseudomonas* strains [41] (Table S3). The AcrAB/TolC-like encoding genes are located next to the gene clusters for diterpenes (*dit*) and phenylethylamine (*pea*) degradation (Table S3, Fig. 4). Two additional gene clusters, one that shows similarity to a solvent-induced RND-type gene cluster in *P. putida* [12], and other analogous to the *ttg2ABCDEF* genes that encode an efflux pump of the ATP-binding cassette family involved in multidrug resistance and toluene tolerance in *P. putida* DOT-T1E [17], were also identified in *Azoarcus* sp. CIB (Table S3). The extensive repertoire of gene clusters encoding putative solvent efflux pumps in *Azoarcus* sp. CIB may reflect an adaptive response of this bacterium when exposed to a wide range of toxic substances that were likely present in the diesel fuel-contaminated aquifer from which it was isolated [20]. Nevertheless, we cannot discard that some of these efflux pumps are involved in extrusion of natural compounds, such as plant produced flavonoids (see below), and therefore their participation in solvent tolerance will require further experimental demonstration.

One interesting outcome of the *Azoarcus* sp. CIB genome analysis is the evidence of a high number of gene clusters encoding potential resistance to heavy metals (Table S4). It is worth noting that the majority of the predicted heavy metal resistance genes in *Azoarcus* sp. CIB are clustered in genomic island VII (Table 2, Fig. 1). The high number of putative solvent and heavy metal resistance gene clusters identified in the genome of strain CIB predicts an environmentally-relevant feature of this bacterium that remained unknown, and suggest promise in using this organism to treat sites containing mixed wastes of aromatic solvents and metals.

Genes for an endophytic lifestyle

Rhizosphere is a nutrient-rich microbial hotspot. To gain a competitive advantage, some of the rhizosphere bacteria penetrate plant organs and share both saprophytic and endophytic lifestyles. Recent studies have revealed that the capacity to colonize plants endophytically cannot be reduced to a few genetic traits and that different bacteria have evolved differently in their adaptation to the plant environment [36]. In this sense, it appears that plant colonization ability involves the combination of different genetic determinants, some of which are also present in non-endophytic bacteria, while other may be missing in well-characterized endophytes. So far, no unique gene cluster could be exclusively linked to the endophytic lifestyle [36]. We have searched the genome of *Azoarcus* sp. CIB to identify genes similar to those that were shown to encode functions relevant for an endophytic lifestyle in selected endophyte genomes such as that of the closely related species *Azoarcus* sp. BH72 [28].

The root surface colonization is guided by plant-released compounds, i.e., root exudates, that serve as chemoattractants, and bacterial motility is generally achieved by flagella and type IV pili (twitching motility) [7]. In *Azoarcus* sp. CIB, the genes involved in the biosynthesis and function of flagella and pili are organized in three noncontiguous gene clusters, respectively (Table S5). Although there is a conserved *che* cluster involved in the chemotaxis apparatus, an additional *cheII* cluster (Table S5), that is not conserved in most *Azoarcus* strains, is located within genomic island IX (Table 2) near to the ICE_{DIT} element (Fig. 1).

Flagella and type IV pili are also involved in bacterial adhesion to roots, which is a major feature for successful plant colonization [36]. Other frequent adhesins for root surface colonization are capsular material and exopolysaccharides. *Azoarcus* sp. CIB contains several gene clusters located within genomic islands (Table 2) distributed along the chromosome and likely devoted to the synthesis of these adhesins (Table S5, Fig. 1). Motility and the production of plant cell wall-degrading enzymes, e.g. glycoside hydrolases, might be involved in the colonization and spreading of the endophytes within the plant. Once inside plants, endophytes must have the capacity for quick adaptation to a highly different environment and they need also to overcome plant defense responses to the invasion, e.g., production of reactive oxygen species (ROS) [36]. In this sense, the genome of *Azoarcus* sp. CIB contains genes likely coding for some glycoside hydrolases and for detoxification of ROS, such as catalases, superoxide dismutases, peroxidases, hydroperoxide reductases and glutathione-S-transferases. Moreover, some of the RND-family efflux pumps that were suggested to be involved in solvent resistance (Table S3) may be also required for the export of plant toxins [8]. It is worth noting that the CIB genome contains several copies of the *phb* genes encoding proteins for the synthesis and degradation of polyhydroxybutyrate (PHB), a carbon storage compound that may enhance survival during starvation and tolerance to stress [26,54], and that could serve also as a redox regulator for the removal of growth inhibitory plant secondary metabolites [3].

Rapid adaptation of bacteria to increased osmolarity may aid plant colonization, and thus endophytes are usually provided with hyperosmotic stress response mechanisms [35]. *Azoarcus* sp. CIB contains genes orthologs to those involved in the transport and metabolism of different types of bacterial osmolytes, such as trimethylamine *N*-oxide (TMAO), polyamines and sarcosine (Table S5), located within genomic island VIII (Table 2, Fig. 1). Protein secretion systems, particularly type II, IV and VI systems, play an important role in beneficial plant-microbe interactions by triggering defense responses and by supporting colonization, nutrition and proliferation

of the bacteria [11,47,55]. We have found at least four different secretion systems, i.e., type I, II, IV and VI secretion systems, encoded in the genome of *Azoarcus* sp. CIB (Table S5). Type VI secretion systems are frequently present in genomic islands [11]. In this sense, a predicted type VI secretion system is encoded within genomic island XI of *Azoarcus* sp. CIB (Table 2, Fig. 1). In plant/endophyte interactions the microorganisms gain nutrients, e.g. plant-material derived aromatic compounds such as cinnamate, *p*-coumarate, phenylpropionate (Fig. 4), and a protected niche, whereas the host benefits from bacterial activities resulting in plant growth promotion, control of plant pathogens and induction of systemic resistance [50]. Bacterial phytostimulation can be carried out by the production/modulation of plant hormones, such as IAA, ethylene, etc., bio-fertilization through nitrogen fixation and enhancement of the uptake of mineral nutrients [36]. We have shown previously that *Azoarcus* sp. CIB fix nitrogen, produces IAA, and solubilizes insoluble inorganic phosphate compounds [16]. The genes predicted to encode some of these plant growth promoting functions, e.g. the *nif/rnfI* gene cluster for nitrogen fixation (see above), have been detected in the genome of *Azoarcus* sp. CIB (Table S5, Fig. 1).

Conclusions

The genome sequence of *Azoarcus* sp. CIB is a valuable addition to the body of knowledge about a unique bacterium that combines the ability to degrade anaerobically a high number of aromatic compounds [32,25] with the ability to colonize two different ecological niches, i.e., soil and water as free-living bacteria, and the inner tissues of the rice roots as a facultative endophyte [32]. Analysis of the genome of strain CIB strongly suggests that horizontal gene transfer and mobile genetic elements played a major role as a mechanism of adaptation of this bacterium to its different lifestyles. Comparative genomics among the five sequenced *Azoarcus* strains available so far provides a pan-genome that confirms the global metabolic flexibility of this bacterial genus, improves the characterization of the core genome and the phylogenetic relationships among these bacteria, and suggests that the phylogeny of the *Azoarcus* genus should be revisited in the near future. Unraveling the complex history and role of mobile genetic elements in the diversification and evolution of the *Azoarcus* species will require further investigation.

Analyses of the *Azoarcus* sp. CIB genome confirmed that this strain can be regarded as a degradation specialist feeding on compounds that range from plant exudates, e.g., dicarboxylic acids, to naturally as well as anthropogenic released toxic aromatic compounds, e.g., toluene and *m*-xylene. The potential genetic determinants that encode the peripheral and central pathways for the aerobic and anaerobic degradation of aromatic compounds have been identified. Moreover, these analyses expanded the catabolic potential of strain CIB towards common natural compounds, such as certain diterpenes and probably some steroids, that were not anticipated as carbon sources within the *Azoarcus* genus.

The presence of a high number of putative solvent efflux pumps and heavy metal resistance gene clusters in the genome of *Azoarcus* sp. CIB revealed two potential environmentally-relevant features of this bacterium that remained unknown and that deserve further research. The combination of a wide metabolic versatility with stress resistance properties suggest promise in using *Azoarcus* sp. CIB to treat sites containing mixed wastes of aromatic solvents and metals, and as a suitable biocatalyst in defined communities of degradation specialists.

Genome mining revealed several gene clusters likely involved in the endophytic lifestyle of *Azoarcus* sp. CIB, opening the door to the molecular characterization of some plant growth promoting traits [16] which makes this endophyte as a potential candidate for phytostimulation, biofertilization and biocontrol in a more sustainable agricultural practice [50]. Since *Azoarcus* sp. CIB is able to degrade toxic aromatic compounds and resist metals, the use of this bacterium in association with plants could offer an efficient, economic and sustainable phytoremediation technology [61].

This work paves the way for a systems biology-based understanding of the abilities of *Azoarcus* sp. CIB to integrate aerobic and anaerobic metabolism of aromatic compounds, tolerate stress conditions, and interact with plants as an endophyte. This will ultimately lead to an increased understanding of the adaptation processes of strain CIB to its different lifestyles, and will allow to elucidate further the role of the individual gene clusters described in this work in the proposed target functions.

Acknowledgements

We thank Lifesequencing (Valencia, Spain) for the genome sequencing of strain CIB, and Dr. J Blom (Justus-Liebig-University Giessen, Germany) for creating and allowing us to use the EDGAR project with the five annotated *Azoarcus* strains. This work was supported by grants BIO2009-10438, CSD2007-00005, BIO2012-39501, European Union FP7 Grant 311815. Z.M.-M. was the recipient of Research Personnel Formation (FPI) fellowship from the Ministry of Economy and Competitiveness of Spain.

References

1. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-3402.

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

3. Aneja, P., Dai, M., Lacorre, D.A., Pillon, B., Charles, T.C. (2004) Heterologous complementation of the exopolysaccharide synthesis and carbon utilization phenotypes of *Sinorhizobium meliloti* Rm1021 polyhydroxyalkanoate synthesis mutants. FEMS Microbiol. Lett. 239, 277-283.

4. Arias, S., Olivera, E.R., Arcos, M., Naharro, G., Luengo, J.M. (2008) Genetic analyses and molecular characterization of the pathways involved in the conversion of 2-phenylethylamine and 2-phenylethanol into phenylacetic acid in *Pseudomonas putida* U. Environ. Microbiol. 10, 413-432.

5. Blázquez, B., Carmona, M., García, J.L., Díaz, E. (2008) Identification and analysis of a glutaryl-CoA dehydrogenase-encoding gene and its cognate transcriptional regulator from *Azoarcus* sp. CIB. Environ. Microbiol. 10, 474-482.

6. Blom, J., Albaum, S.P., Doppmeier, D., Pühler, A., Vorhölter, F.J., Zakrzewski, M., Goesmann, A. (2009) EDGAR: a software framework for the comparative analysis of prokaryotic genomes. BMC Bioinformatics. 10,154.

7. Böhm, M., Hurek, T., Reinhold-Hurek, B. (2007). Twitching motility is essential for endophytic rice colonization by the N_2- fixing endophyte *Azoarcus* sp. strain BH72. Mol. Plant Microbe Interact. 20, 526–533.

8. Burse, A., Weingart, H., Ullrich, M.S. (2004) The phytoalexin-inducible multidrug efflux pump AcrAB contributes to virulence in the fire blight pathogen, *Erwinia amylovora*. Appl. Environ. Microbiol. 70, 693-703.

9. Carmona, M., Zamarro, M.T., Blázquez, B., Durante-Rodríguez, G., Juárez, J.F., Valderrama, A., Barragán, M.J.L., García, J.L., Díaz, E. (2009) Anaerobic catabolism of aromatic compounds: a genetic and genomic view. Microbiol. Mol. Biol. Rev. 73, 71-133.

10. Choi, E.N., Cho, M.C., Kim, Y., Kim, C.K., Lee, K. (2003) Expansion of growth substrate range in *Pseudomonas putida* F1 by mutations in both *cymR* and *todS*, which recruit a ring-fission hydrolase CmtE and induce the *tod* catabolic operon, respectively. Microbiology. 149, 795-805.

11. Coulthurst, S.J. (2013) The Type VI secretion system - a widespread and versatile cell targeting system. Res. Microbiol. 164, 640-654.

12. Domínguez-Cuevas, P., González-Pastor, J.E., Marqués, S., Ramos, J.L., de Lorenzo, V. (2006) Transcriptional tradeoff between metabolic and stress-response programs in *Pseudomonas putida* KT2440 cells exposed to toluene. J. Biol. Chem. 281, 11981-11991.

13. Durante-Rodríguez, G., Valderrama, J.A., Mancheño, J.M., Rivas, G., Alfonso, C., Arias-Palomo, E., Llorca, O., García, J.L., Díaz, E., Carmona M. (2010) Biochemical characterization of the transcriptional regulator BzdR from *Azoarcus* sp. CIB. J. Biol. Chem. 285, 35694-35705.

14. Ebenau-Jehle, C., Thomas, M., Scharf, G., Kockelkorn, D., Knapp, B., Schühle, K., Heider, J., Fuchs, G. (2012) Anaerobic metabolism of indoleacetate. J. Bacteriol. 194, 2894-2903.

15. Egland, P. G., Pelletier, D.A., Dispensa, M., Gibson, J., Harwood, C.S. (1997) A cluster of bacterial genes for anaerobic benzene ring biodegradation. Proc. Natl. Acad. Sci. USA 94, 6484-6489.

16. Fernández, H., Prandoni, N., Fernández-Pascual, M., Fajardo, S., Morcillo, C., Díaz, E., Carmona, M. (2014) *Azoarcus* sp. CIB, an anaerobic biodegrader of aromatic compounds shows an endophytic lifestyle. PLoS One. 9, e110771.

17. García, V., Godoy, P., Daniels, C., Hurtado, A., Ramos, J.L., Segura, A. (2010) Functional analysis of new transporters involved in stress tolerance in *Pseudomonas putida* DOT-T1E. Environ. Microbiol. Rep. 2, 389-395.

18. García, J.L., Uhía, I., Galán, B. (2012) Catabolism and biotechnological applications of cholesterol degrading bacteria. Microb. Biotechnol. 5, 679-699.

19. Gescher, J., Zaar, A., Mohamed, M., Schägger, H., Fuchs, G. (2002) Genes coding for a new pathway of aerobic benzoate metabolism in *Azoarcus evansii*. J. Bacteriol. 184, 6301-6315.

20. Hess, A., Zarda, B., Hahn, D., Häner, A., Stax, D., Höhener, P., Zeyer, J. (1997) *In situ* analysis of denitrifying toluene- and *m*-xylene-degrading bacteria in a diesel fuel-contaminated laboratory aquifer column. Appl. Environ. Microbiol. 63, 2136-2141.

21. Horinouchi, M., Hayashi, T., Kudo, T. (2012) Steroid degradation in *Comamonas testosteroni*. J. Steroid. Biochem. Mol. Biol. 129, 4-14.

22. Hurek, T., Reinhold-Hurek, B. (1995) Identification of grass-associated and toluene-degrading diazotrophs, *Azoarcus* spp., by analyses of partial 16S ribosomal DNA sequences. Appl. Environ. Microbiol. 61, 2257-2261.

23. Jiang, K., Sanseverino, J., Chauhan, A., Lucas, S., Copeland, A., Lapidus, A., Del Rio, T.G., Dalin, E., Tice, H., Bruce, D., Goodwin, L., Pitluck, S., Sims, D., Brettin, T., Detter, J.C., Han, C., Chang, Y.J., Larimer, F., Land, M., Hauser, L., Kyrpides, N.C., Mikhailova, N., Moser, S., Jegier, P., Close, D., Debruyn, J.M., Wang, Y., Layton, A.C.,

Allen, M.S., Sayler, G.S. (2012) Complete genome sequence of *Thauera aminoaromatica* strain MZ1T. Stand. Genomic Sci. 6, 325-335.

24. Jormakka, M., Byrne, B., Iwata, S. (2003) Formate dehydrogenase--a versatile enzyme in changing environments. Curr. Opin. Struct. Biol. 13, 418-423.

25. Juárez, J.F., Zamarro, M.T., Eberlein, C., Boll, M., Carmona, M., Díaz, E. (2013) Characterization of the *mbd* cluster encoding the anaerobic 3-methylbenzoyl-CoA central pathway. Environ. Microbiol. 15, 148-166.

26. Kadouri, D., Jurkevitch, E., Okon, Y. (2003) Involvement of the reserve material poly-beta-hydroxybutyrate in *Azospirillum brasilense* stress endurance and root colonization. Appl. Environ. Microbiol. 69, 3244-3250.

27. Karp, P.D., Paley, S., Romero, P. (2002) The Pathway Tools software. Bioinformatics. 18 Suppl 1, S225-232.

28. Krause, A., Ramakumar, A., Bartels, D., Battistoni, F., Bekel, T., Boch. J., Böhm, M., Friedrich, F., Hurek, T., Krause, L., Linke, B., McHardy, A.C., Sarkar, A., Schneiker, S., Syed, A.A., Thauer, R., Vorhölter, F.J., Weidner, S., Pühler, A., Reinhold-Hurek. B., Kaiser, O., Goesmann, A. (2006) Complete genome of the mutualistic, N2-fixing grass endophyte *Azoarcus* sp. strain BH72. Nat. Biotechnol. 24, 1385-1391.

29. Kühner, S., Wöhlbrand, L., Fritz, I., Wruck, W., Hultschig, C., Hufnagel, P., Kube, M., Reinhardt, R., Rabus, R. (2005) Substrate-dependent regulation of anaerobic degradation pathways for toluene and ethylbenzene in a denitrifying bacterium, strain EbN1. J. Bacteriol. 187, 1493-1503.

30. Laempe, D., Jahn, M., Breese, K., Schägger, H., Fuchs, G. (2001) Anaerobic metabolism of 3-hydroxybenzoate by the denitrifying bacterium *Thauera aromatica*. J. Bacteriol. 183, 968-979.

31. Lee, D.J., Wong, B.T., Adav, S.S. (2014) *Azoarcus taiwanensis* sp. nov., a denitrifying species isolated from a hot spring. Appl. Microbiol. Biotechnol. 98, 1301-1307.

32. López-Barragán, M.J., Carmona, M., Zamarro, M.T., Thiele, M., Boll, M., Fuchs, G., García, J.L., Díaz, E. (2004) The *bzd* gene cluster, coding for anaerobic benzoate catabolism, in *Azoarcus* sp. strain CIB. J. Bacteriol. 186, 5762-5774.

33. Luque-Almagro, V.M., Gates, A.J., Moreno-Vivián, C., Ferguson, S.J., Richardson, D.J., Roldán, M.D. (2011) Bacterial nitrate assimilation: gene distribution and regulation. Biochem. Soc. Trans. 39, 1838-1843.

34. Mathee, K., Narasimhan, G., Valdes, C., Qiu, X., Matewish, J.M., Koehrsen, M., Rokas, A., Yandava, C.N., Engels, R., Zeng, E., Olavarietta, R., Doud, M., Smith, R.S., Montgomery, P., White, J.R., Godfrey, P.A., Kodira, C., Birren, B., Galagan, J.E., Lory, S. (2008) Dynamics of *Pseudomonas aeruginosa* genome evolution. Proc. Natl. Acad. Sci. U S A. 105, 3100-3105.

35. Miller, K.J., Wood, J.M. (1996) Osmoadaption by rhizosphere bacteria. Annu. Rev. Microbiol. 50, 101–136.

36. Mitter, B., Petric, A., Shin, M.W., Chain, P.S., Hauberg-Lotte, L., Reinhold-Hurek, B., Nowak, J., Sessitsch, A. (2013) Comparative genome analysis of *Burkholderia phytofirmans* PsJN reveals a wide spectrum of endophytic lifestyles based on interaction strategies with host plants. Front Plant. Sci. 4,120.

37. Nishizawa, T., Tago, K., Oshima, K., Hattori, M., Ishii, S., Otsuka, S., Senoo, K. (2012) Complete genome sequence of the denitrifying and N_2O-reducing bacterium *Azoarcus* sp. strain KH32C. J. Bacteriol. 194, 1255.

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

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

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

41. Ramos, J.L., Duque, E., Gallegos, M.-T., Godoy. P., Ramos-González. M.I., Rojas. A., Terán, W., Segura, A. (2002) Mechanisms of solvent tolerance in Gram-negative bacteria. Annu. Rev. Microbiol. 56, 743–768.

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

43. Rhine, E.D., Phelps, C.D., Young, L.Y. (2006) Anaerobic arsenite oxidation by novel denitrifying isolates. Environ. Microbiol. 8, 899-908.

44. Richter, M., Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc. Nat. Acad. Sci. U. S. A. 106, 19126–19131.

45. Sarkar, A., Köhler, J., Hurek, T., Reinhold-Hurek, B. (2012) A novel regulatory role of the Rnf complex of *Azoarcus* sp. strain BH72. Mol. Microbiol. 83, 408–422.

46. Schühle, K., Jahn, M., Ghisla, S., Fuchs, G. (2001) Two similar gene clusters coding for enzymes of a new type of aerobic 2-aminobenzoate (anthranilate) metabolism in the bacterium *Azoarcus evansii*. J. Bacteriol. 183, 5268-5278.

47. Sessitsch, A., Hardoim, P., Döring, J., Weilharter, A., Krause, A., Woyke, T., Mitter, B., Hauberg-Lotte, L., Friedrich, F., Rahalkar, M., Hurek, T., Sarkar, A., Bodrossy, L., van Overbeek, L., Brar, D., van Elsas, J.D., Reinhold-Hurek, B. (2012) Functional

characteristics of an endophyte community colonizing rice roots as revealed by metagenomic analysis. Mol. Plant Microbe Interact. 25, 28–36.

48. Smith, D.J., Patrauchan, M.A., Florizone, C., Eltis, L.D., Mohn, W.W. (2008) Distinct roles for two CYP226 family cytochromes P450 in abietane diterpenoid catabolism by *Burkholderia xenovorans* LB400. J. Bacteriol. 190, 1575-1583.

49. Song, B., Häggblom, M.M., Zhou, J., Tiedje, J.M., Palleroni, N.J. (1999) Taxonomic characterization of denitrifying bacteria that degrade aromatic compounds and description of *Azoarcus toluvorans* sp. nov. and *Azoarcus toluclasticus* sp. nov. Int. J. Syst. Bacteriol. 49, 1129-1140.

50. Sturz, A.V., Christie, B.R., Nowak, J. (2000) Bacterial endophytes: potential role in developing sustainable systems of crop production. Crit. Rev. Plant Sci. 19, 1-30.

51. Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30, 2725-2729.

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

53. Ton-Hoang, B., Siguier, P., Quentin, Y., Onillon, S., Marty, B., Fichant, G., Chandler, M. (2012) Structuring the bacterial genome: Y1-transposases associated with REP-BIME sequences. Nucleic Acids Res. 40, 3596-3609.

54. Trautwein, K., Kühner, S., Wöhlbrand, L., Halder, T., Kuchta, K., Steinbüchel, A., Rabus, R. (2008) Solvent stress response of the denitrifying bacterium "*Aromatoleum aromaticum*" strain EbN1. Appl. Environ. Microbiol. 74, 2267-2274.

55. Tseng, T.-T., Tyler, B.M., Setubal, J.C. (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. BMC Microbiol. 9 (Suppl. 1):S2. doi: 10.1186/1471-2180-9-S1-S2.

56. Udaondo, Z., Molina, L., Daniels, C., Gómez, M.J., Molina-Henares, M.A., Matilla, M.A., Roca, A., Fernández, M., Duque, E., Segura, A., Ramos, J.L. (2013) Metabolic potential of the organic-solvent tolerant *Pseudomonas putida* DOT-T1E deduced from its annotated genome. Microb. Biotechnol. 6, 598-611.

57. Valderrama, J.A., Durante-Rodríguez, G., Blázquez, B., García, J.L., Carmona, M., Díaz, E. (2012) Bacterial degradation of benzoate: cross-regulation between aerobic and anaerobic pathways. J. Biol. Chem. 287, 10494-10508.

58. Valderrama, J.A., Shingler, V., Carmona, M., Díaz, E. (2014) AccR is a master regulator involved in carbon catabolite repression of the anaerobic catabolism of aromatic compounds in *Azoarcus* sp. CIB. J. Biol. Chem. 289, 1892-1904.

59. Van Melderen, L. (2010) Toxin-antitoxin systems: why so many, what for? Curr. Opin. Microbiol. 13, 781–785.

60. Watanabe, T., Kojima, H., Fukui, M. (2014) Complete genomes of freshwater sulfur oxidizers *Sulfuricella denitrificans* skB26 and *Sulfuritalea hydrogenivorans* sk43H: genetic insights into the sulfur oxidation pathway of betaproteobacteria. Syst. Appl. Microbiol. 37, 387-395.

61. Weyens, N., Beckers, B., Schellingen, K., Ceulemans, R., van der Lelie, D., Newman, L., Taghavi, S., Carleer, R., Vangronsveld, J. (2015) The potential of the Ni-resistant TCE-degrading *Pseudomonas putida* W619-TCE to reduce phytotoxicity and improve phytoremediation efficiency of poplar cuttings on a Ni-TCE co-contamination. Int. J. Phytoremediation. 17, 40-48.

62. Wöhlbrand, L., Wilkes, H., Halder, T., Rabus, R. (2008) Anaerobic degradation of *p*-ethylphenol by "*Aromatoleum aromaticum*" strain EbN1: pathway, involved proteins and regulation. J. Bacteriol. 190, 5699–5709.

63. Zamarro, M.T., López-Barragán, M.J., de la Peña, F., Prieto, M.A., Carmona, M., García, J.L., Díaz, E. **(**2013) A procedure to obtain polyhydroxybutyrate by fermenting *Azoarcus* sp. CIB and a organic acid as carbon source. Spanish Patent P201330102.

64. Zhou, N.Y., Fuenmayor, S.L., Williams, P.A. (2001) *nag* genes of *Ralstonia* (formerly *Pseudomonas*) sp. strain U2 encoding enzymes for gentisate catabolism. J. Bacteriol. 183, 700-708.

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

Figure legends

Fig. 1. Distribution of some mobile genetic elements, toxin-antitoxin systems (TASs) and regions of unusual GC content in the *Azoarcus* sp. CIB genome. The AzCIB_xxxx locus number in the annotated CIB genome is detailed in brackets. The color code for transposases is: black, IS*4*-family transposases without paralogs; light yellow and orange, IS*4*-family transposases with paralogs; dark yellow, IS*3*-family transposases; dark blue, IS*91*-family transposases; ligth blue, IS*30*-family transposases; light green, IS*21*-family transposases; dark green, IS*5*-family transposases; brown, IS*L3*-family transposase; rose, IS*110*-family transposases; grey, IS*66*-family transposases; white, IS*1595*-family transposase; red, IS*200*/IS*605*-family transposase (Y1 family); violet, other transposases. The 14 predicted TASs are numbered. Abbreviations used are detailed in Supplementary data.

Fig. 2. Venn diagram representing the core genome and pan genome of the five sequenced *Azoarcus* strains. The core and pan genome was calculated with the EDGAR program [6] by iterative pairwise comparison of all genomes taking as reference genome that of strain CIB. The Venn diagram shows the number of reciprocal best hits between a subset of genomes. Therefore, the number of exclusive genes in each genome is always higher than the number of singletons (a gene without any hit against any other genome) for each strain.

Fig. 3. (A) Core genome tree generated using EDGAR showing the phylogenetic positions of the five sequenced *Azoarcus* strains. The phylogenetic tree was constructed based on the nucleotide sequences of the 1701 genes that constitute the *Azoarcus* core genome. The tree was visualized by using MEGA 6 [51]. (B) Average amino acid identities (AAI, in %) calculated from the core genome data set obtained by comparing the five sequenced *Azoarcus* strains. *Thauera aminoaromatica* MZ1T strain was used as reference strain from a closely related bacterial genus.

Fig. 4. Pathways for the catabolism of aromatic compounds in *Azoarcus* sp. CIB.
(A) The location of genes and genes clusters predicted to encode the aromatic catabolic pathways is indicated (plain numbers) on the complete *Azoarcus* sp. CIB genome. The AzCIB_xxxx locus number in the annotated CIB genome is detailed in brackets. The ICE_{DIT} and ICE_{XTD} elements are represented by open and filled bars, respectively. Genes responsible for anaerobic or aerobic pathways are shown in blue or red, respectively. Genes encoding central pathways are indicated in bold and italics; genes encoding peripheral pathways that funnel to central routes are indicated in plain text. Genes names and their predicted functions are detailed in Supplementary data.
(B) Scheme of the predicted peripheral and central pathways for the anaerobic (in blue) or aerobic (in red) catabolism of aromatic compounds in *Azoarcus* sp. CIB. Central pathways are shown as thick arrows. Peripheral pathways are shown as thin arrows with continuous (represent one biochemical step) or discontinuous (represent several biochemical steps) lines. The names of the genes are indicated in panel A. The *hbcL, mbdA, bzdA, paaK*, and *bclA* genes are those encoding the corresponding aromatic acid

CoA ligases within the *hbd, mbd, bzd, paa* and *box* clusters, respectively. The names of the compounds are detailed in Supplementary data.

Table 1

Genome features of *Azoarcus* sp. strain CIB in comparison to those of strains EbN1, BH72, KH32C, and MF63.

Genome features	CIB	EbN1	BH72	KH32C	MF63
Total size (bp)*	5,257,030	4,727,255	4,376,040	5,818,755	5,925,983
GC%	65.8	65.1	67.9	65.1	66.0
Protein	4,739	4,603	3,992	5,188	5,432
rRNA	4	4	4	5	4(?)
tRNA	57	58	56	64	48
Plasmids	0	2	0	1	ND

* Total genome size, including plasmids.

Table 2

Some genomic islands in *Azoarcus* sp. CIB and presence of ortholog genes in the genome of strains EbN1, BH72, KH32C, and MF63.

Genome island	Chromosomal location	Predicted function	*Azoarcus* genomes that contain ortholog genes
I	145 kb-261 kb	Unknown	MF63
II	589 kb-622 kb	Lipopolysaccharide synthesis	EbN1, BH72, KH32C, MF63
III	768 kb-784 kb	Nitrogen fixation	BH72, KH32C, MF63
IV	923 kb-941 kb	Exopolysaccharide synthesis	KH32C, MF63
V	1269 kb-1300 kb	Unknown	None
VI	1509 kb-1758 kb	Steroids metabolism	MF63
VII	1834 kb-2070 kb	Metals resistance	MF63
VIII	2326 kb-2394 kb	Osmotic stress resistance	KH32C, MF63
IX	2608 kb-2621 kb	Chemotaxis	MF63
X (ICE_{DIT})[a]	2636 kb-2925 kb	Diterpenes degradation	MF63 (lacks the ICE_{DIT} element)
XI	3620 kb-3635 kb	Type VI protein secretion	KH32C, MF63
XII	4083 kb- 4118 kb	Exopolysaccharide synthesis	KH32C
XIII	4262 kb-4292 kb	Capsular material synthesis	None
XIV(ICE_{XTD})[a]	4894 kb-5067 kb	Aromatic hydrocarbons degradation	EbN1, MF63 (lack the ICE_{XTD} element)
XV	5164 kb-5196 kb	Phage tail-like bacteriocin	KH32C, MF63

[a] ICE, integrative and conjugative element.

Figure 1

Click here to download high resolution image

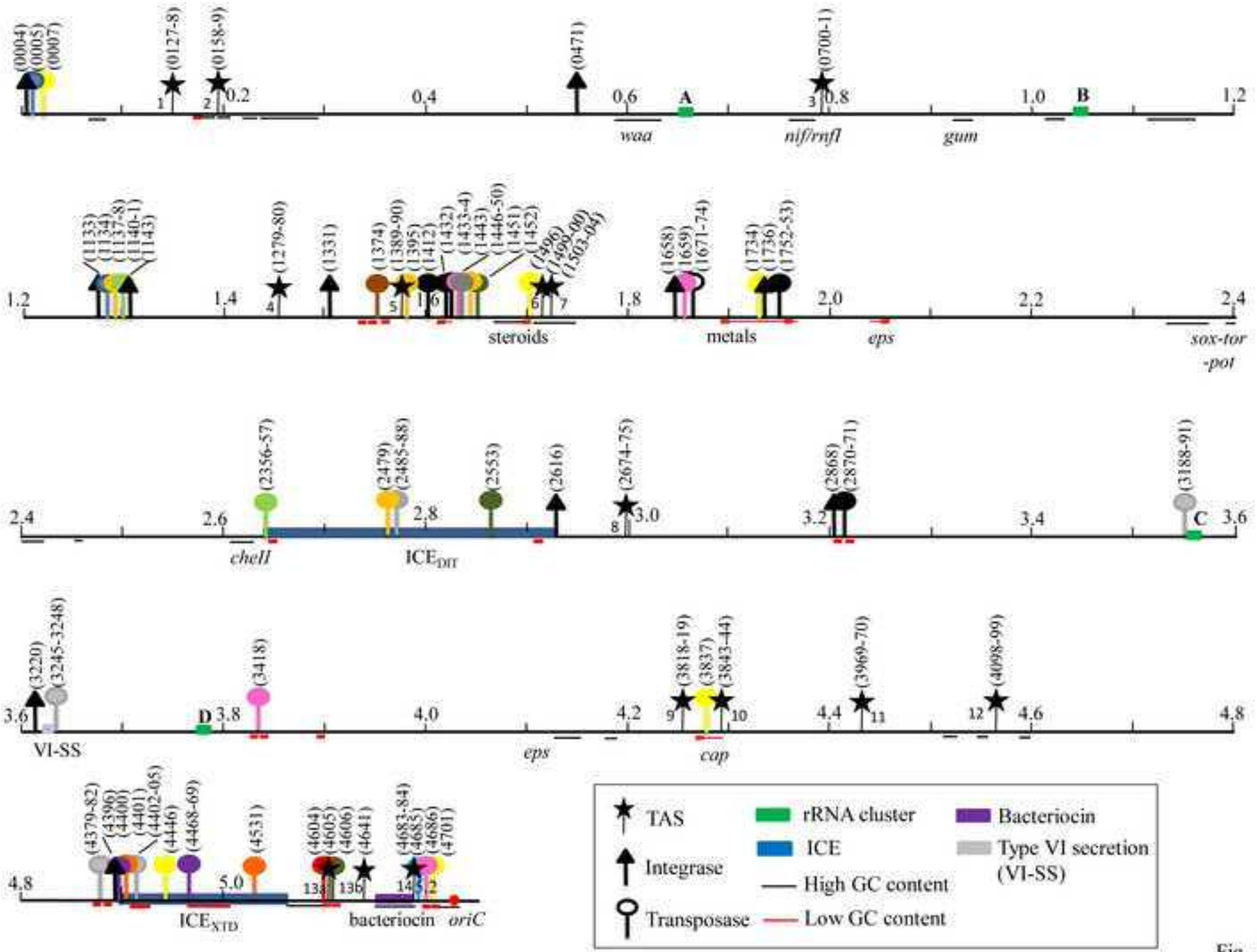


Fig. 1

Figure 2

Click here to download high resolution image

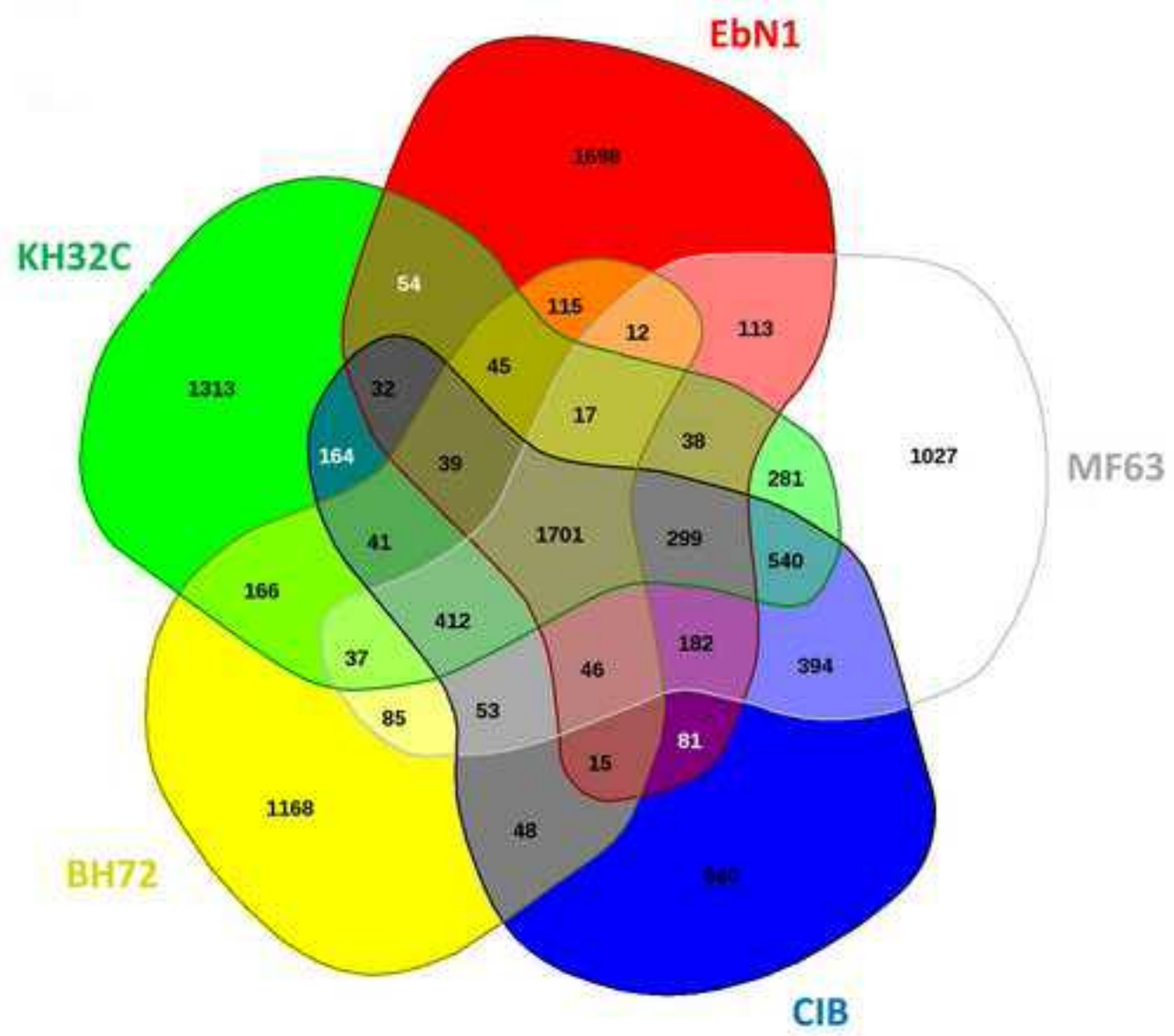
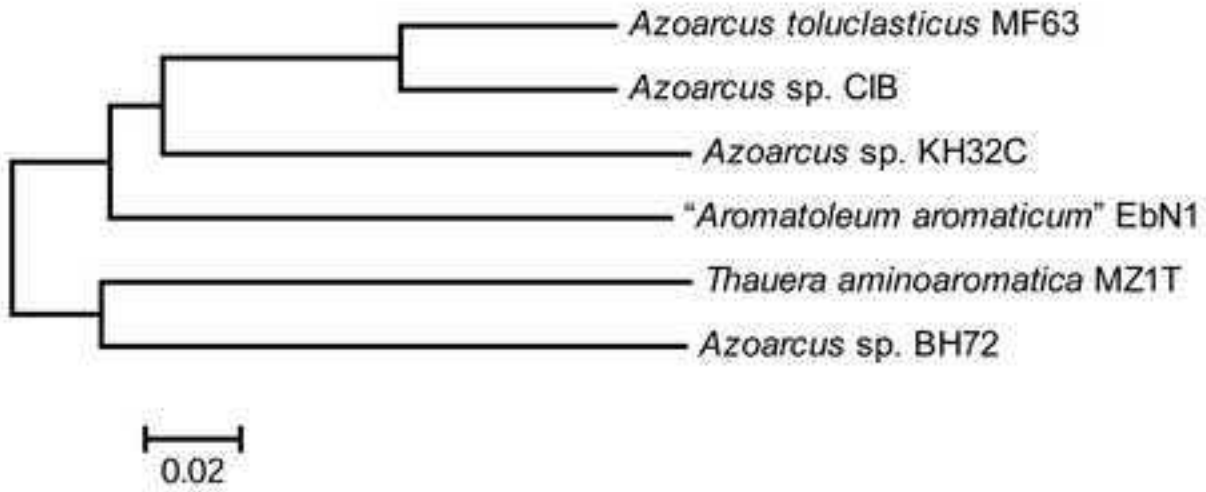


Fig. 2

Figure 3
Click here to download high resolution image

A



B

Azoarcus toluclasticus MF63	76.06	77.69	83.27	85.65	95.14	100
Azoarcus sp. CIB	75.84	77.61	83.31	85.63	100	
Azoarcus sp. KH32C	75.55	77.19	82.51	100		
"Aromatoleum aromaticum" EbN1	75.5	77.49	100			
Azoarcus sp. BH72	78.46	100				
Thauera aminoaromatica MZ1T	100					

Fig. 3

Figure 4

Click here to download high resolution image

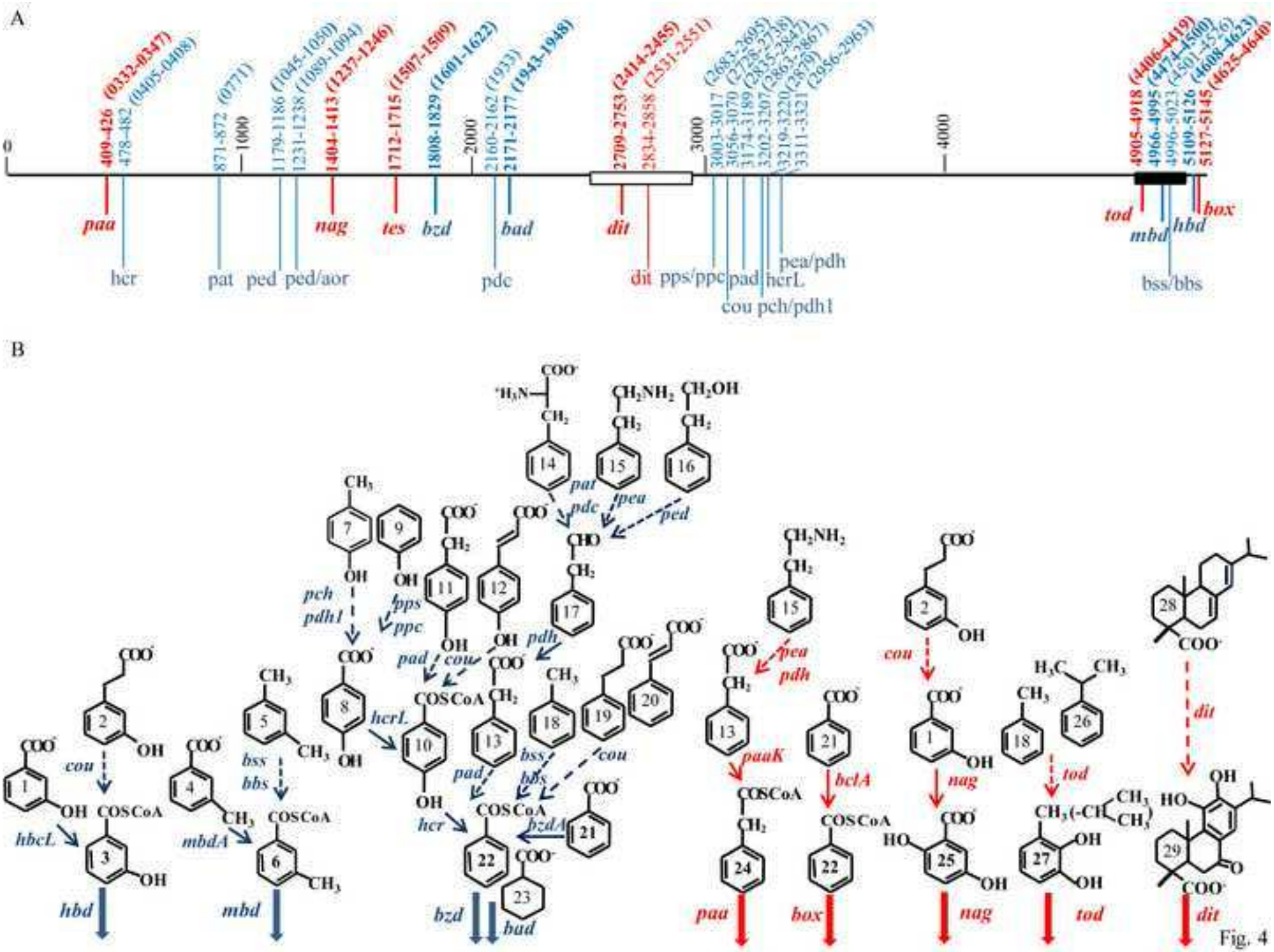


Fig. 4