Systematic and Applied Microbiology 38 (2015) 462–471

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

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

Zaira Martín-Moldes[a,1], María Teresa Zamarro[a,1], Carlos del Cerro[a,1], Ana Valencia[a], Manuel José Gómez[b,2], Aida Arcas[b,3], Zulema Udaondo[a,4], José Luis García[a], Juan Nogales[a], Manuel Carmona[a], Eduardo Díaz[a,*]

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

ARTICLE INFO

Article history:
Received 24 March 2015
Received in revised form 29 June 2015
Accepted 6 July 2015

Keywords:
Azoarcus
Aromatic compounds
Endophyte
Metals resistance
Mobile genetic elements
Comparative genomics

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 toward 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.

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 ecophysiological 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

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.

* Corresponding author. Tel.: +34 915611800.
E-mail address: ediaz@cib.csic.es (E. Díaz).

[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.

http://dx.doi.org/10.1016/j.syapm.2015.07.002
0723-2020/

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 °C 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 30 °C. 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