



# Genome of the candidate phylum *Aminicenantes* bacterium from a deep subsurface thermal aquifer revealed its fermentative saccharolytic lifestyle

Vitaly V. Kadnikov<sup>1</sup> · Andrey V. Mardanov<sup>1</sup> · Alexey V. Beletsky<sup>1</sup> · Olga V. Karnachuk<sup>2</sup> · Nikolai V. Ravin<sup>1</sup>

Received: 8 August 2018 / Accepted: 17 December 2018 / Published online: 1 January 2019  
© Springer Japan KK, part of Springer Nature 2019

## Abstract

Bacteria of candidate phylum OP8 (*Aminicenantes*) have been identified in various terrestrial and marine ecosystems as a result of molecular analysis of microbial communities. So far, none of the representatives of *Aminicenantes* have been isolated in a pure culture. We assembled the near-complete genome of a member of *Aminicenantes* from the metagenome of the 2-km-deep subsurface thermal aquifer in Western Siberia and used genomic data to analyze the metabolic pathways of this bacterium and its ecological role. This bacterium, designated BY38, was predicted to be rod shaped, it lacks flagellar machinery but twitching motility is encoded. Analysis of the BY38 genome revealed a variety of glycosyl hydrolases that can enable utilization of carbohydrates, including chitin, cellulose, starch, mannose, galactose, fructose, fucose, rhamnose, maltose and arabinose. The reconstructed central metabolic pathways suggested that *Aminicenantes* bacterium BY38 is an anaerobic organotroph capable of fermenting carbohydrates and proteinaceous substrates and performing anaerobic respiration with nitrite. In the deep subsurface aquifer *Aminicenantes* probably act as destructors of buried organic matter and produce hydrogen and acetate. Based on phylogenetic and genomic analyses, the novel bacterium is proposed to be classified as *Candidatus* Saccharicenans subterraneum.

**Keywords** Candidate phylum OP8 · Aminicenantes · Subsurface biosphere · Metagenome

## Introduction

The candidate phylum OP8 was originally identified in the sediments of the thermal spring Obsidian Pool in Yellowstone National Park (USA) among twelve first-described candidate phyla (Hugenholtz et al. 1998). In subsequent years, the 16S ribosomal RNA (rRNA) gene sequences

assigned to this candidate phylum were detected by molecular methods in various terrestrial and marine ecosystems, but in almost all of them they represented a minor portion of the communities, in 99% of the analyzed datasets not exceeding 1% (Farag et al. 2014). Members of OP8 are often present and abundant in hydrocarbon-impacted environments and deepwater marine hydrothermal systems, as well as in terrestrial aquatic ecosystems, such as hot springs and groundwater (Farag et al. 2014). Although OP8 bacteria are more frequent in habitats with low oxygen content and salinity, they were found in a wide range of oxygen concentrations, and at various salinities and temperatures (Farag et al. 2014). Currently, the SILVA database (Quast et al. 2013) contains about 5000 sequences of 16S rRNA genes assigned to the candidate phylum OP8.

The taxonomic status of OP8 is not clearly defined. Rinke et al. (2013) showed that OP8 is a sister lineage to the *Acidobacteria*, but proposed to classify it as a separate candidate phylum, and such assignments are consistent with the NCBI taxonomy database (Federhen 2012). The SILVA database and the recently described Genome Taxonomy Database

Communicated by S. Albers.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00792-018-01073-5>) contains supplementary material, which is available to authorized users.

✉ Nikolai V. Ravin  
nravin@biengi.ac.ru

<sup>1</sup> Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Leninsky Prospekt 33-2, Moscow 119071, Russia

<sup>2</sup> Laboratory of Biochemistry and Molecular Biology, Tomsk State University, 634050 Tomsk, Russia

(GTDB) based on genome phylogeny (Parks et al. 2018) consider the OP8 division as a class *Aminicenantia* in the phylum *Acidobacteria*.

Although OP8 to date has no cultured members, some information about their biology was obtained by sequencing metagenomes and single-cell genomes. At present, 64 genomes of OP8 bacteria are available in GenBank, but all of them are incomplete and represented by a set of contigs. The first data on the genetic potential of this candidate phylum were obtained upon sequencing of 36 single-cell genomes of OP8 bacteria from sediments of the brackish Sakinaw Lake in Canada (Rinke et al. 2013). An analysis of these genomes has shown that OP8 bacteria can use amino acids as substrates, so the name *Aminicenantes* was proposed for this candidate phylum (Rinke et al. 2013). Enzymes of the Wood–Ljungdahl pathway of autotrophic CO<sub>2</sub> fixation were found in these genomes, and it was suggested that *Aminicenantes* can run this pathway in the reverse direction to drive acetate oxidation into hydrogen and CO<sub>2</sub> in a syntrophic association with hydrogenotrophic methanogens (Gies et al. 2014). Sequencing of the sediment metagenome of an aquifer adjacent to the Colorado River (Rifle, Colorado, USA) allowed to assemble the composite genome of the representative of *Aminicenantes*. Analysis of this metagenome-assembled genome (MAG) revealed genes encoding glycosyl hydrolases and the aerobic respiratory chain, which led to the conclusion that these *Aminicenantes* may degrade organic substrates through either fermentation or aerobic respiration (Sharon et al. 2015). For another representative of *Aminicenantes*, a draft genome with an estimated completeness of 88% was recovered from the metagenome of the formation waters of hydraulically fractured coal bed methane production wells (Robbins et al. 2016). Reconstruction of metabolic pathways of this bacterium, designated *Aminicenantes*-PK28, has shown that it can use proteins and various polysaccharides as substrates, fermenting them under anaerobic conditions, but that it has no pathways for aerobic and anaerobic respiration (Robbins et al. 2016).

Deep terrestrial subsurface environments are extreme habitats characterized by a combination of high temperature, high pressure and sometimes high salinity. Microbial communities of the deep subsurface biosphere include representatives of various uncultured groups of prokaryotes. For many of them draft and even complete genomes were determined by metagenomics or single-cell genome sequencing (Anantharaman et al. 2016; HERNSDORF et al. 2017; Magnabosco et al. 2016; Probst et al. 2016). The oil and gas basin in the region of Western Siberia, Russia, was formed from sediments of marine origin of the Mesozoic period. In addition to oil bodies, reservoirs of underground thermal water were found at depths of 1–3 km. Some of them are available for study via the oil exploration boreholes through which

groundwater flows to the surface under natural pressure. Previously, we have studied the composition of the microbial community and sequenced the metagenome of subsurface thermal waters flowing out through the 1-R borehole in the Tomsk region (Kadnikov et al. 2018). This microbial community mainly consisted of sulfate-reducing *Firmicutes* and *Deltaproteobacteria*, as well as uncultured lineages of the phyla *Chloroflexi*, *Ignavibacteriae*, and *Aminicenantes* (Kadnikov et al. 2018). In particular, pyrosequencing of the 16S rRNA gene fragments showed that the frequency of *Aminicenantes* in the community was about 10%.

Taking advantage of the fact that *Aminicenantes* represents an unusually large fraction in this microbial community, in this study we used metagenomic sequencing to assemble the near-complete genome of thermophilic bacterium of the candidate phylum *Aminicenantes*. Genome data were used to reconstruct its metabolic pathways and gain insights into the ecological role of *Aminicenantes* in the deep subsurface aquifer.

## Materials and methods

### Sampling and determination of physicochemical characteristics of water

The oil exploration borehole 1-R is located in the town Byelii Yar in the Tomsk region of the Russian Federation (coordinates 58.4496 N, 85.0279 E). The borehole was drilled in 1961–1962 to a depth of 2563 m. The water flowing out of the borehole under natural pressure originated from a depth of 1997–2005 m, where sedimentary rocks of the Cretaceous period are located (Banks et al. 2014).

Samples of water were collected on 4–5 August 2014. The water had a temperature of 43 °C, which corresponds to a depth of about 2 km. It had a slightly alkaline pH (8.5) and a negative redox potential (Eh – 341 to – 279 mV). Among the ions, sodium and chloride prevailed; the chemical composition of the salts indicated their origin was from the ancient ocean that covered this region in the Mesozoic era (Banks et al. 2014). However, the total mineralization was about 1.8 g/l, accounting for only 5% salinity of the seawater. This indicates that most of the water is derived from meteoric recharge with a minor fraction of connate water from the ancient ocean (Banks et al. 2014; Kadnikov et al. 2018).

Microorganisms were collected from 50 L of water by filtration through 0.22 µm cellulose nitrate filters. The filters were homogenized by grinding with liquid nitrogen, and then the obtained powder was dissolved in TE buffer in a 37 °C water bath. The total community DNA was extracted using the CTAB/NaCl method (Wilson 2001).

## Sequencing and assembly of the metagenome, contig binning, and genome analysis

A sample of metagenomic DNA was sequenced on an Illumina HiSeq 2500 [250 nucleotides (nt) single-end reads] according to the manufacturer's protocols (Illumina Inc., USA) as described in Kadnikov et al. 2018. Primer sequences and low-quality reads were removed using Cutadapt (Martin 2011) and Sickle (<https://github.com/najoshi/sickle>), respectively. A total of 86.5 million high-quality reads (18.5 Gbp) were obtained. Contig assembly was carried out using the SPAdes Genome Assembler, in the metagenome assembly mode (“-meta” parameter).

Binning of contigs longer than 2000 nt into MAGs was carried out using the program CONCOCT (Alneberg et al. 2014). The completeness of the obtained metagenome-assembled genomes (MAGs) and their possible “contamination” (i.e., the presence of contigs representing genomes of other microorganisms in a given MAG) was assessed using CheckM v. 1.05 (Parks et al. 2015) and Anvi'o v. 4 (Eren et al. 2015). The 16S rRNA genes in the contigs were identified using CheckM.

Gene search and annotation of the MAG assigned to *Aminicenantes* were performed using the RAST server 2.0 (Brettin et al. 2015), followed by manual correction of the annotation by the comparison of predicted protein sequences with the National Center for Biotechnology Information (NCBI) databases. Signal peptides were predicted by Signal P v.4.1 for Gram-negative bacteria (<http://www.cbs.dtu.dk/services/SignalP/>) and PRED-TAT (<http://www.compgen.org/tools/PRED-TAT/>), and the presence of transmembrane helices was predicted by TMHMM v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>). The iRep software was used to calculate the index of genomic DNA replication (Brown et al. 2016).

## Genome-to-genome distance evaluation

To find genome assemblies closely related to BY38, homologs of 157 conserved marker genes found by CheckM in the BY38 genome were identified in the NCBI non-redundant database using a BLASTP search. Genomes corresponding to the top 15 hits for each of the marker genes (a total of 1047 genomes) were selected. Then, the average amino acid identity (AAI) between BY38 and the selected genomes was calculated using the *aai.rb* script from the *envomics* collection (Rodriguez-R and Konstantinidis 2016). Likewise, AAI was calculated between BY38 and two members of *Aminicenantes* described earlier (Sharon et al. 2015; Robbins et al. 2016).

The values of DNA–DNA hybridization *in silico* were calculated using GGDC2 tool (Meier-Kolthoff et al. 2013), available at <http://ggdc.dsmz.de/>.

## Phylogenetic analysis

GTDB-Tk v.0.1.3 toolkit (Parks et al. 2018) was used to find 120 single-copy bacterial marker genes in the assembled BY38 MAG and to construct a multiple alignment of concatenated single-copy gene sequences, comprising those from BY38 and all species from the GTDB. Previously sequenced partial genome of *Aminicenantes*-PK28 (Robbins et al. 2016), the *Aminicenantes* genome from the Rifle site (Sharon et al. 2015), and several other *Aminicenantes* genomes from NCBI database (with high AAI values) were additionally included in this analysis. Selected part of the multiple alignment built in GTDB-Tk was used to construct a phylogenetic tree in PhyML v. 3.3 (Guindon et al. 2010) with default parameters. The level of support for internal branches was assessed using the Bayesian test in PhyML.

Sequences of the 16S rRNA genes were aligned using Mothur v. 1.35.1 (Schloss et al. 2009). The maximum likelihood phylogenetic tree was computed by RAxML v. 8.2.8 (Stamatakis 2014), using the GTRGAMMA substitution model. Bootstrap tests were performed with 100 resamplings.

## Nucleotide sequence accession numbers

The sequence of the 16S rRNA gene and the annotated sequence of the BY38 bacterium of the candidate phylum *Aminicenantes* were deposited in the GenBank database under the accession numbers MH712740 and QUAH00000000, respectively.

## Results and discussion

### Genome assembly of a member of the candidate phylum *Aminicenantes*

To obtain MAGs of the members of the microbial community, metagenomic sequences with a total length of about 18.5 Gbp were generated and assembled into contigs. Binning of contigs was performed with CONCOCT using nucleotide composition and coverage data (Kadnikov et al. 2018). One of the obtained MAGs, BY38, was represented by 25 contigs with a total length of 2,899,266 nt, sequenced to 341-fold average coverage. The relative abundance of this genotype in the community, defined as a fraction of this MAG in the metagenome, was about 6.6%. Analysis of the presence of a set of 188 conservative single-copy marker genes in the BY38 genome with CheckM estimated that the completeness of this MAG was 95%, with a possible 5% contamination. Similar estimates, 96% completeness and 2% contamination, were obtained using Anvi'o. Thus, the BY38 genome met the recently proposed criteria (Bowers

et al. 2017) for a “high-quality” MAG (> 90% completeness and < 5% contamination).

An operon comprising the 16S and 23S rRNA genes, the 5S rRNA gene, and 46 transfer RNA (tRNA) genes capable of encoding all 20 amino acids were identified in the BY38 genome. As a result of the genome annotation, 2,509 potential protein-coding genes were identified and functions of 1,609 (64%) of them were assigned. Interestingly, despite the absence of a CRISPR (clustered regularly interspaced short palindromic repeats) system that provides protection from viruses and mobile elements, only two genes apparently associated with mobile elements and prophages were found in the BY38 genome. Perhaps the BY38 bacterium possesses other mechanisms, ensuring efficient protection against foreign DNA.

BY38 bacteria lacked genes encoding flagellar machinery and chemotaxis; however, a set of genes necessary for the generation of type IV pili have been found. Such pili enable twitching motility and attachment of the bacterium to solid surfaces, including insoluble growth substrates (Mandlik et al. 2008). BY38 cells are predicted to be rod shaped, based on the identification of genes encoding the rod shape-determining proteins MreBCD, PBP-2, and RodA (Supplemental Table S1).

The rate of DNA replication in situ was evaluated using the iRep software (Brown et al. 2016). The iRep replication index of 1.13 was calculated for the BY38 genome, indicating that these bacteria are slow growers (about 13% of cells were actively replicating at the time of sampling), but nevertheless are metabolically active.

### Phylogenetic placement of BY38

The search for relatives of BY38 on the basis of genome-to-genome distance evaluation revealed that BY38 is most related to *Aminicenantes*-PK28 (Robbins et al. 2016) and uncultured bacterium UBA10528 with an average AAI of 85% and 83%, respectively (Table 1). According to the AAI thresholds proposed by Konstantinidis et al. (2017) for uncultivated microorganisms, these three genomes represented different species in a single genus. Consistently, the degree of *in silico* DNA–DNA hybridization between BY38 and *Aminicenantes*-PK28 was estimated to be about 25%, indicating that these genomes represented different species. The AAI values between BY38 and four other *Aminicenantes* genomes, ARK-02, UBA1061, UBA2191, and UBA6228 were in the range of 69–72%. Probably these organisms belong to the same family as BY38, but the lack of near full size 16S rRNA gene sequences in these four assemblies prevented direct phylogenetic comparison.

To determine the phylogenetic position of the BY38 bacterium, a phylogenetic tree based on concatenated sequences of conservative marker genes was constructed. The results

confirmed that BY38 belongs to *Aminicenantes*, and placed BY38 within the family OPB95 of the order *Aminicenantales*, as defined by the GTDB taxonomy (Fig. 1). All classes, orders, and families proposed by GTDB corresponded to well-separated monophyletic branches (Fig. 1). Consistent with the AAI data, *Aminicenantes*-PK28 and UBA10528 appeared to cluster with BY38, and a sister lineage comprising ARK-02, UBA1061, UBA2191, and UBA6228.

The 16S rRNA gene found in the BY38 genome had only 82% sequence identity with the nearest cultured bacterium (*Thermodesulfobrio hydrogeniphilus*); however, many 16S rRNA sequences with up to 100% identity were described as members of the candidate phylum *Aminicenantes*. A search of its 16S rRNA gene against the SILVA database (Quast et al. 2013) also classified BY38 as a member of *Aminicenantes*. Previous phylogenetic studies of *Aminicenantes* identified four proposed classes in this candidate phylum (Frag et al. 2014). Phylogenetic analysis of the 16S rRNA gene sequences revealed that BY38 belongs to the class OP8-1 which seems to be an equivalent of the order *Aminicenantales* in the GTDB taxonomy (Fig. 2). However, the absence of near-complete 16S rRNA genes in most MAGs shown in Fig. 1 does not allow for a detailed comparison of genomic and 16S rRNA phylogenies. The status of *Aminicenantes* as a class of *Acidobacteria* or as a separate phylum and the internal taxonomy of this division will become definitive only after the acceptance of standardized approach for assigning species to higher taxonomic ranks.

### Possible growth substrates of BY38 bacterium

Analysis of the BY38 genome revealed enzymes that can enable utilization of various carbohydrate substrates, including some polysaccharides and simple sugars. Known pathways and enzymes for the utilization of mannose, galactose, fructose, fucose, rhamnose, maltose, ribose, and arabinose are encoded (Fig. 3 and Supplemental Table S1). Consistently, the BY38 genome encodes ABC-type and major facilitator superfamily sugar transporters as well as a putative phosphotransferase system for the uptake of mannose and fructose. These sugars can be produced as a result of the hydrolysis of polysaccharides by glycosyl hydrolases of BY38 bacterium, as indicated by the presence of alpha-mannosidase, alpha- and beta-galactosidase, alpha-fucosidase, alpha-rhamnosidase, beta-glycosidase/beta-xylosidase, and alpha-N-arabinofuranosidase. The presence of N-terminal secretion signal peptides in these enzymes suggests that they are involved in extracellular hydrolysis of the corresponding polysaccharide substrates.

Intracellular metabolism of sugars is likely linked to glycolytic pathways. For example, phosphomannomutase and mannose-6-phosphate isomerase are involved in the conversion of mannose into fructose 6-phosphate. Metabolism

**Table 1** General characteristics of the genomes of *Aminicenantes*

Name	GenBank accession or reference	16S rRNA gene <sup>a</sup>	AAI with BY38	Completeness (%)	Assembly size (bp)	GC content (%)
<b>Aminicenantales</b>						
Ca. <i>Saccharicenans subterraneum</i> BY38	QUAH000000000	+	100	95	2,899,266	54.52
Aminicenantes-PK28	Robbins et al. (2016)	+	85.18	88	2,461,777	53.14
UBA10528	DMVN000000000	–	82.82	70	1,933,857	53.70
ARK-02	PNJE000000000	–	71.95	92	2,550,048	43.29
UBA1061	DCDG000000000	–	69.22	91	2,510,304	45.74
UBA2191	DCVJ000000000	–	69.72	90	2,374,313	47.40
UBA6228	DJWO000000000	–	69.61	92	2,456,980	47.19
RBG_19FT_COMBO_65_30	MEYF000000000	–	56.81	85	2,453,444	64.05
RBG_16_63_14	MEYA000000000	–	55.37	76	2,388,421	63.95
RBG_16_66_30	MEYC000000000	–	55.21	73	2,075,472	65.98
UBA4085	DFYJ000000000	–	52.90	85	3,491,420	65.29
UBA8522	DQHV000000000	–	54.29	81	3,507,216	58.43
Aminicenantes-Rifle	Sharon et al. (2015)	–	54.87	71	2,346,518	58.42
RBG_16_63_16	MEYB000000000	–	54.06	88	4,526,823	62.39
RBG_19FT_COMBO_58_17	MEYD000000000	–	53.45	78	3,573,970	58.16
4484_214	MZGL000000000	–	51.89	77	2,453,389	42.98
JdFR-80	MTOP000000000	–		93	2,914,703	44.60
Ca. <i>Aminicenans sakinawicola</i> JGI OTU-1	AWN000000000	+	50.94	84	2,913,626	39.27
Aminicenantes bacterium JGI OTU-2	AWN000000000	+	50.39	73	2,405,692	39.08
<b>Order_JdFR-78</b>						
JdFR-78	MTON000000000	–	49.55	94	2,530,072	32.60
JdFR-79	MTOO000000000	–	49.35	74	2,046,026	32.64
<b>Order_UBA2199</b>						
UBA6072	DIXU000000000	+	43.30	90	3,116,808	62.56
UBA2190	DCVK000000000	–	43.06	88	2,953,209	62.76
UBA2199	DCVB000000000	–	42.72	76	2,513,533	62.18
UBA876	DBFQ010000000	–	40.18	78	1,927,288	56.77
UBA647	DBOL010000000	–	39.56	86	2,656,053	37.77

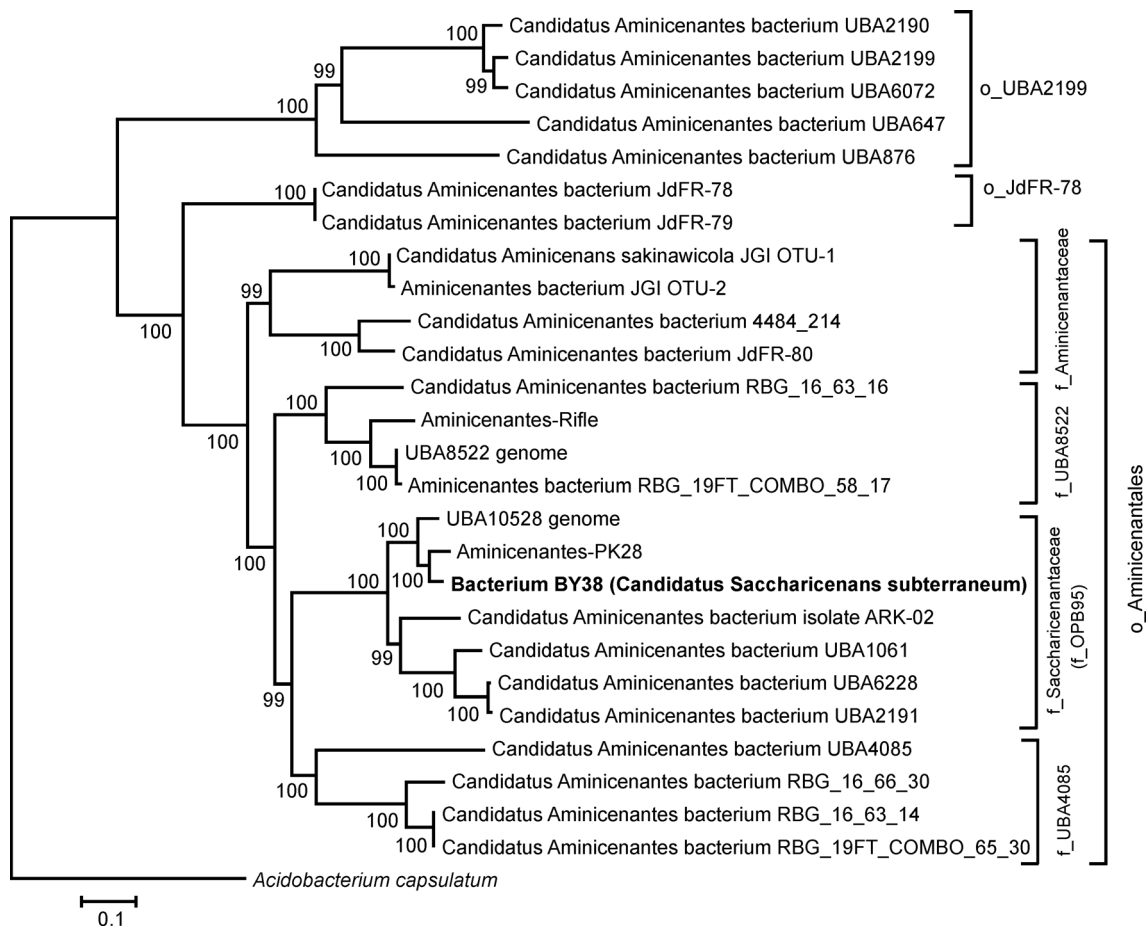
<sup>a</sup>The presence of full-length (> 1400 bp) 16S rRNA sequence

of galactose likely follows the Leloir pathway. L-arabinose isomerase and ribulokinase channel arabinose into the pentose phosphate pathway. Metabolism of L-fucose and L-rhamnose is probably mediated by the corresponding kinase, isomerase, and aldolase, yielding in both cases dihydroxyacetone phosphate and L-lactaldehyde. The latter could be oxidized to lactate by an aldehyde:ferredoxin oxidoreductase.

The utilization of starch and similar polymers could be performed by alpha-glycosidase/glucoamylase of the GH97 family and the trehalase-like glycosidase, both containing N-terminal signal peptides. The BY38 genome encodes a secreted polygalacturonase, an enzyme that cleaves the alpha-1,4 glycosidic bonds between galacturonic acid

residues in pectin, and rhamnogalacturonan hydrolase from the GH88 family that catalyzes the hydrolytic release of unsaturated glucuronic acids from oligosaccharides. However, known enzymes involved in the intracellular metabolism of galacturonate monomers were not found, which leaves open the possibility of growth on pectin. The presence of a secreted endoglucanase/cellulase of the GH5 family and beta-glucosidase indicates that BY38 bacterium could be capable of extracellular hydrolysis of cellulose substrates, although the cellulolytic microorganisms usually have a much larger set of relevant enzymes.

A complete set of enzymes that could enable utilization of chitin (Hunt et al. 2008) was found in the BY38 genome. The extracellular hydrolysis of chitin can be accomplished by a



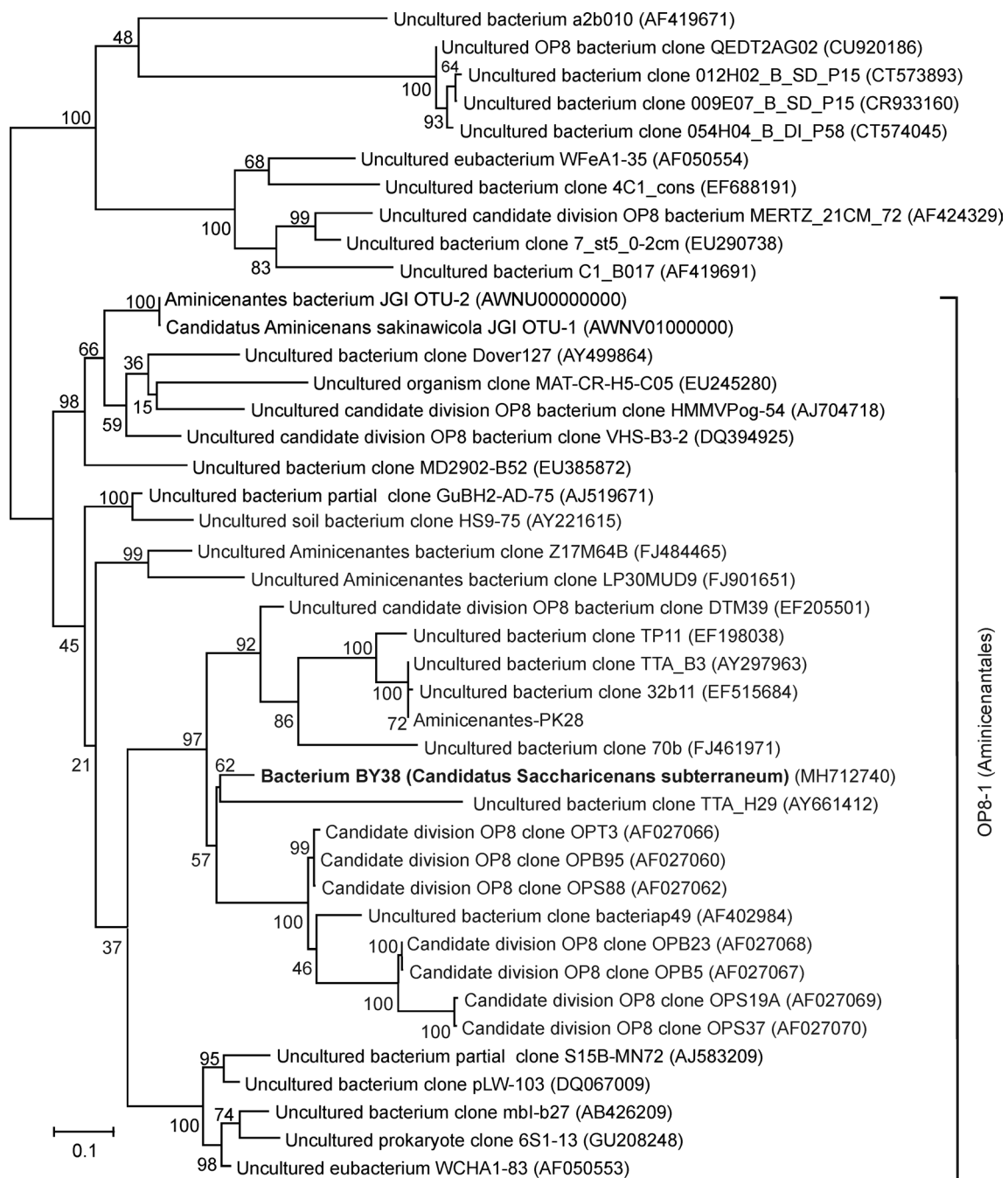
**Fig. 1** Position of BY38 in the maximum likelihood concatenated protein phylogeny. A selected part of the GTDB-Tk multiple alignment was used for tree construction in PhyML using default parameters. Aminicenantes-Rifle and Aminicenantes-PK28 refer to the *Aminicenantes* genomes described by Sharon et al. (2015) and Rob-

bins et al. (2016), respectively. The tree was inferred from the concatenation of 120 conserved bacterial marker genes. The support values for the internal nodes were estimated by approximate Bayes tests in PhyML. Taxonomy is shown according to the GTDB (*f* family, *o* order)

GH18 family endochitinase. The search for related sequences revealed similar (with 47–77% amino acid sequence identity) chitinases in genomes of several other *Aminicenantes*, including Aminicenantes-PK28 and ARK-02. The search for other chitinolytic enzymes containing N-terminal signal peptides revealed the presence of the GH20 family *N*-acetyl- $\beta$ -hexosaminidase and the GH3 family beta-hexosaminidase. These enzymes can cleave monomers of *N*-acetyl-*D*-glucosamine (GlcNAc) from the non-reducing end of chitin oligomers generated by the GH18 family endochitinase (Scigelova and Crout 1999; Hutcheson et al. 2011). All the above-mentioned chitinolytic enzymes do not contain recognizable chitin-binding domains. It is possible that the adherence of the BY38 bacterium to insoluble chitin could be mediated by type IV pili, as proposed for chitinolytic bacteria *Vibrio parahaemolyticus* and *Chitinivibrio alkalicophilus* (Frischkorn et al. 2013; Sorokin et al. 2014). Upon import into the cytoplasm, GlcNAc can be phosphorylated

by *N*-acetylglucosamine kinase NagC with the generation of GlcNAc-6-phosphate. Then *N*-acetylglucosamine-6-phosphate deacetylase NagA deacetylates GlcNAc-6-phosphate, yielding glucosamine-6-phosphate. Finally, glucosamine-6-phosphate deaminase converts glucosamine-6-phosphate into fructose-6-phosphate, an intermediate of the Embden–Meyerhof glycolysis pathway.

In addition to carbohydrates, it is likely that BY38 bacteria can use proteinaceous substrates for growth. This is indicated by the presence of several secreted peptidases of the families M16, M23, C69, and S8, as well as multiple amino acid and peptide transporters. In particular, the extracellular hydrolysis of proteins could be performed by the subtilisin-like serine protease of the S8 family. The presence of a peptidase of the M23 family, whose members can degrade bacterial cell walls, suggests that BY38 bacterium could use peptides from dead cells of other microorganisms, as proposed for other *Aminicenantes* (Robbins et al. 2016).



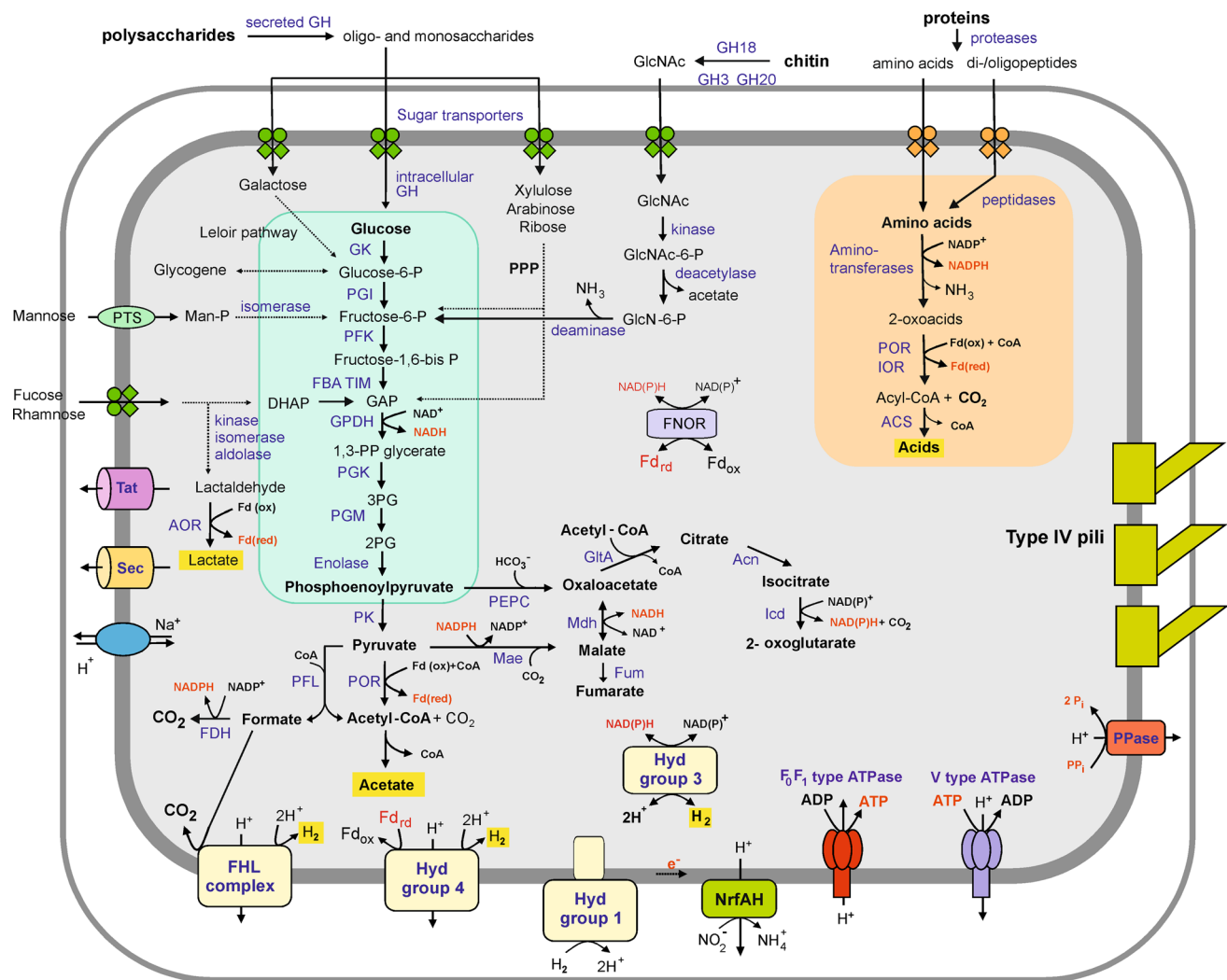
**Fig. 2** Maximum likelihood 16S rRNA gene phylogenetic tree of the candidate phylum *Aminicenantes*. The tree was computed by RAxML v. 8, using the GTRGAMMA substitution model. GenBank accession

numbers are shown after the clone names. The scale bar represents substitutions per nucleotide base. Bootstrap values are indicated at the nodes

### Predicted central metabolic pathways

The BY38 genome contains a complete set of genes encoding the enzymes of the Embden–Meyerhof glycolytic pathway and gluconeogenesis (Fig. 3 and Supplemental Table S1), including glucokinase, glucose-6-phosphate isomerase, 6-phosphofruktokinase, fructose-bisphosphate aldolase,

triosephosphate isomerase, glyceraldehyde 3-phosphate dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase, enolase, pyruvate kinase, pyruvate, phosphate dikinase, and fructose-1,6-bisphosphatase. The pentose phosphate pathway is represented only by a non-oxidative branch. The tricarboxylic acids cycle in BY38 bacterium was not closed due to lack of 2-oxoglutarate oxidoreductase and



**Fig. 3** An overview of the metabolism of BY38 *Aminicenantes*. Enzyme abbreviations: *GH* glycoside hydrolase, *GK* glukokinase, *PGI* glucose-6-phosphate isomerase, *PFK* 6-phosphofruktokinase, *FBA* fructose-bisphosphate aldolase, *TIM* triosephosphate isomerase, *GPDH* glyceraldehyde 3-phosphate dehydrogenase, *PGK* phosphoglycerate kinase, *PGM* phosphoglycerate mutase, *PK* pyruvate kinase, *POR* pyruvate ferredoxin oxidoreductase, *ACS* acetyl-CoA synthetase, *IOR* indolepyruvate ferredoxin oxidoreductase, *PEPC* phosphoenolpyruvate carboxylase, *PFL* pyruvate formate lyase, *FHL*

citrate synthase, *Acn* aconitase, *Icd* isocitrate dehydrogenase, *Fum* fumarate hydratase, *Mdh* malate dehydrogenase, *Mae* malic enzyme, *FHL* formate hydrogen lyase, *FDH* formate dehydrogenase, *Hyd* hydrogenase, *NrfAH* cytochrome c nitrite reductase, *FNOR* ferredoxin-NADP(+) reductase, *AOR* aldehyde:ferredoxin oxidoreductase, *PPase* pyrophosphatase. Other abbreviations: *PPP* pentose phosphate pathway, *PTS* phosphotransferase system, *ox/red* oxidized and reduced forms, *Pi* phosphate, *PPi* pyrophosphate, *CoA* coenzyme A

succinate dehydrogenase and is probably used for biosynthetic purposes. The key enzymes of autotrophic carbon fixation pathways, the Calvin cycle (ribulose 1,5-bisphosphate carboxylase) and the Wood–Ljungdahl pathway (carbon monoxide dehydrogenase/acetyl-CoA synthase complex), were not found.

The pyruvate produced during glycolysis could be decarboxylated to yield acetyl-coenzyme A (CoA) by pyruvate:ferredoxin oxidoreductase. Alternatively, pyruvate formate lyase could catalyze the conversion of pyruvate and CoA into formate and acetyl-CoA. Formate could

be oxidized either by cytoplasmic formate dehydrogenase or by a membrane-linked formate hydrogenlyase complex. The final step of the fermentation pathway, the oxidation of acetyl-CoA into acetate with the concomitant production of ATP, could be catalyzed by acetyl-CoA synthetase or by phosphate acetyltransferase and acetate kinase in a two-step reaction. The presence of alcohol dehydrogenases suggests that ethanol could be produced as a fermentation product.

The important role of hydrogen in the metabolism of BY38 bacterium was evidenced by the presence of [NiFe] hydrogenases from four groups. The first hydrogenase



belongs to group 3d and includes 4 subunits forming the hydrogenase (HoxYH) and the diaphorase (HoxFU) moieties. Group 3d hydrogenases are localized in the cytoplasm and, depending on the redox status of the cell, can perform fermentative NADH-dependent production of H<sub>2</sub> or the reverse reaction (Greening et al. 2016). The second hydrogenase belongs to group 4e; it is a multi-subunit energy-conserving membrane-bound enzyme that couples the oxidation of reduced ferredoxin to proton reduction. Such hydrogenases are able to translocate protons across the cytoplasmic membrane to generate a transmembrane ion gradient (Greening et al. 2016). The third hydrogenase, group 4b, is part of the formate hydrogenlyase complex, which, in addition to the hydrogenase subunits, comprises molybdenum-dependent formate dehydrogenase. Formate hydrogenlyase oxidizes the formate to CO<sub>2</sub> with the concomitant reduction of protons to H<sub>2</sub>. The complex is associated with the cytoplasmic membrane and could pump protons and generate the transmembrane protonmotive force (Kim et al. 2010). The activity of these three hydrogenases can not only drive the reoxidation of NADH and reduced ferredoxin produced during the course of fermentation, but also contribute to the generation of a transmembrane ion gradient that can be used by membrane F<sub>0</sub>F<sub>1</sub>-type ATP synthase for ATP generation.

The physiological role of the fourth hydrogenase is less clear. This enzyme belongs to group 1e of the respiratory H<sub>2</sub> uptake hydrogenases. Similar enzymes were absent in genomes of other *Aminicenantes*, except for *Aminicenantes*-PK28. The nearest homologs have been found in the genomes of *Chloroflexi*, suggesting that this hydrogenase was acquired by BY38 bacterium via lateral transfer. The hydrogenase consists of a large catalytic, small electron transfer and a third subunit linking them to the membrane. The presence of a Tat motif in the N-terminus of the small subunit indicates that the hydrogenase is localized on the outer side of the cytoplasmic membrane. Hydrogenases from group 1e are capable of oxidizing H<sub>2</sub> and transferring electrons to the quinone pool in the cytoplasmic membrane (Greening et al. 2016).

Analysis of the BY38 genome revealed the absence of the main components of an electron transport chain that are necessary for aerobic respiration, namely the proton-translocating NADH-dehydrogenase complex, the membrane-bound succinate dehydrogenase, the cytochrome *bc*<sub>1</sub> complex or alternative complex III, and the terminal cytochrome *c* oxidases. Known pathways of dissimilatory reduction of sulfate, thiosulfate, arsenate, nitrate, and sulfur compounds were also not found. The only detected potential terminal reductase for anaerobic respiration was the NrfAH-like periplasmic cytochrome *c* nitrite reductase that could enable dissimilatory reduction of nitrite to ammonia. It comprises the catalytic NrfA subunit containing seven hems and the tetraheme small subunit NrfH. The sequences of both subunits

contained N-terminal signal peptides, which indicate the orientation of the catalytic subunit toward the periplasmic space. Probably, nitrite is a terminal acceptor of electrons coming from the group 1e respiratory H<sub>2</sub> uptake hydrogenase during the course of the oxidation of hydrogen.

The genome of BY38 bacterium meets the criteria, recently suggested for description of new taxa of uncultivated prokaryotes (Konstantinidis et al. 2017), and we propose the following taxonomic names for the novel genus and species of BY38.

- Description of the novel genus *Candidatus Saccharicenans* (Sac.cha.ri.ce'nans. N.L. n. *saccharum*, sugar; L. v. *cenare*, to eat; N.L. masc. n. *Saccharicenans* a (bacterium) degrading sugars)
- Description of the novel species *Candidatus Saccharicenans subterraneum*

*Saccharicenans subterraneum* (Sub.terr.a'ne.um. L. neutr. adj. *subterraneum*, underground, subterranean).

Not cultivated. Inferred to be anaerobic, rod-shaped, obligate organotroph, obtains energy by fermentation or respiration with nitrite, and able to use various carbohydrates as growth substrates. Represented by near-complete genome (GenBank acc. no. QUAH00000000) obtained from metagenome of a deep subsurface thermal aquifer in Western Siberia, Russia.

Based on this, we propose the name *Candidatus Saccharicenantaceae* for the family. It is defined on a phylogenetic basis by comparative genome sequence analysis of *Candidatus Saccharicenans subterraneum* BY38, *Aminicenantes*-PK28 (Robbins et al. 2016), and *Candidatus Aminicenantes* bacteria ARK-02, UBA10528, UBA1061, UBA2191, and UBA6228.

### Genomic comparison of BY38 and its close relatives

Availability of near-complete MAGs of two members of *Aminicenantes* that are phylogenetically close to BY38, *Aminicenantes*-PK28 (88% completeness), and ARK-02 (92% completeness), allowed for the comparison of their genomic properties and metabolic potential. *Aminicenantes*-PK28 belongs to the genus *Candidatus Saccharicenans*, while ARK-02 represents a sister genus-level lineage within the family *Candidatus Saccharicenantaceae*. The *Aminicenantes*-PK28 genome was assembled from the formation waters of a hydraulically fractured coal bed methane production well in Australia (Robbins et al. 2016). This genome is slightly smaller in size (2.46 Mb) and coding potential, with 2,064 protein-coding genes predicted, 1,651 of which are common with BY38. The ARK-02 genome was obtained from the Arkashin Shurf hot spring in Uzon Caldera, Kamchatka, Russia. It is also shorter than the BY38 genome

(2.55 Mb) and was predicted to contain 2,519 protein-coding genes, of which 1,800 have homologs in the BY38 genome. Both *Aminicenantes*-PK28 and ARK-02 genomes contained CRISPR systems missing in BY38 and could be better protected from viral infections in their environments. Like BY38, the *Aminicenantes*-PK28 and ARK-02 bacteria lack flagellar machinery, yet contain a set of genes for type IV pili.

Comparisons of shared gene pools revealed that most of the metabolically important functions were conserved in these three genomes. All of them encode a nearly identical inventory of glycosyl hydrolases that could enable hydrolysis and utilization of chitin, starch, mannose, galactose, fructose, fucose, rhamnose, maltose, and arabinose. Interestingly, all three genomes contained genes for xylulose kinase, but lacked genes encoding xylose isomerase, the first enzyme of the isomerase pathway of xylose metabolism. Therefore, it is likely that only xylulose could be utilized. The GH5 endoglucanase gene in the *Aminicenantes*-PK28 genome was split, which does not allow a prediction of the possibility of the utilization of cellulose. The ARK-02 genome lacked close homologs of polygalacturonase and rhamnogalacturonyl hydrolase, which were found in BY38 and *Aminicenantes*-PK28, suggesting that the ARK-02 bacterium is unable to use components of pectin. The central metabolic pathways were also highly conserved in the three genomes. A notable exception was that the *Aminicenantes*-PK28 genome lacked the group 4 energy-conserving membrane-bound hydrogenase, and thus, could have more limited capabilities in generating transmembrane ion gradient.

### Environmental distribution of BY38-like *Aminicenantes*

The search for 16S rRNA sequences with > 97% identity to BY38 in the NCBI GenBank (NR database) and the Joint Genome Institute Integrated Microbial Genomes database (16S rRNA Public Isolates and 16S rRNA Public Assembled Metagenomes) revealed a total of 74 nearly full size (> 1300 bp) environmental clones (Supplemental Table S2). Most of these 16S rRNA sequences were identified worldwide in anaerobic digesters treating municipal wastewater and similar wastes (43 clones), terrestrial hot springs (15 clones), anaerobic reactors fed with organic waste from the chemical industry (10 clones), and oil-contaminated soil (3 clones). Therefore, organisms related to BY38 at the species level are distributed globally and mostly occur in anaerobic organic-rich environments within biofilms and microbial mats. Such distribution is consistent with physiological features predicted by genome analysis, anaerobic lifestyle, and the ability to utilize various carbohydrates and proteinaceous substrates.

### Ecological role of BY38 in the deep subsurface aquifer

The microbial community of the underground reservoir in which BY38-like *Aminicenantes* were detected consisted mostly of sulfate-reducing *Firmicutes* (*Ca. Desulforudis audaxviator* and *Desulfotomaculum* sp.), *Nitrospirae* (*Thermodesulfovibrio* spp.), and *Deltaproteobacteria* (*Desulfobacca* spp.), as well as uncultured members of the phyla *Chloroflexi*, *Ignavibacteriae*, *Riflebacteria*, and *Aminicenantes* (Kadnikov et al. 2018). Analysis of metabolic pathways based on the MAGs of representatives of *Chloroflexi*, *Ignavibacteriae*, and *Riflebacteria* showed that they are probably heterotrophs, capable of both fermentation of organic substances and aerobic and/or anaerobic respiration (Kadnikov et al. 2018). However, analyzed members of these phyla had rather limited capacities for the hydrolysis of complex carbohydrates and were restricted to some beta-linked polysaccharides (*Chloroflexi*) or starch (*Ignavibacteriae* and *Riflebacteria*).

The hydrolytic potential of BY38 is broader and can allow it to use a wide range of carbohydrates and proteinaceous substrates, thus determining its ecological function as a destructor of organic matter. Complex polysaccharides of plant origin probably originated from marine sediments, and were buried since the formation of the Western Siberian basin in the Mesozoic era. For example, fucose is the fundamental subunit of the seaweed polysaccharide fucoidan, rhamnose could be produced by diatom algae, and so on. Chitin, a primary component of the exoskeletons of arthropods, such as crustaceans, is abundant in marine sediments. It should be noted that microorganisms capable of degrading complex plant polysaccharides were previously isolated from the groundwater of the Western Siberian region (e.g., Mardanov et al. 2009). Carrying out fermentation of proteins and carbohydrates with the formation of hydrogen and acetate, BY38-like *Aminicenantes* provides substrates for sulfate reducers, hydrogenotrophic *Ca. Desulforudis audaxviator* and *Desulfotomaculum* sp., and acetate-consuming *Desulfobacca* spp.

This study provides the first insight into the biology of a thermophilic member of the candidate phylum *Aminicenantes* found in the deep subsurface and contributed to the understanding of the phylogenetic and metabolic diversity of *Aminicenantes*. Unlike some uncultured candidate phyla with a specialized organotrophic fermentative metabolism, such as *Saccharibacteria* (Albertsen et al. 2013) and *Atribacteria* (Nobu et al. 2016), *Aminicenantes* appears to comprise organisms with diverse metabolic capabilities, including acetate oxidation via the Wood–Ljungdahl pathway, aerobic respiration, and fermentation.

**Acknowledgements** This work was performed using the scientific equipment of the Core Research Facility ‘Bioengineering’ (Research Center of Biotechnology RAS) and supported by the Russian Science Foundation (Grant no. 14-14-01016).

## Compliance with ethical standards

**Conflict of interest** The authors confirm that this article content has no conflict of interest.

## References

- Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* 31:533–538
- Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C (2014) Binning metagenomic contigs by coverage and composition. *Nat Methods* 11:1144–1146
- Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, Brodie EL, Williams KH, Hubbard SS, Banfield JF (2016) Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* 7:13219
- Banks D, Frank Y, Kadnikov V, Karnachuk O, Watts M, Boyce A, Frengstad B (2014) Hydrochemical data report from sampling of two deep abandoned hydrocarbon exploration wells: Byelii Yar and Pabel’, Tomsk oblast’, western Siberia, Russian Federation. NGU Report, 2014.034. Geological Survey of Norway, Trondheim
- Bowers RM, Kyrpides NC, Stepanauskas R et al (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 35:725–731
- Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA 3rd, Stevens R, Vonstein V, Wattam AR, Xia F (2015) RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 5:8365
- Brown CT, Olm MR, Thomas BC, Banfield JF (2016) Measurement of bacterial replication rates in microbial communities. *Nat Biotechnol* 34:1256–1263
- Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO (2015) Anvi’o: an advanced analysis and visualization platform for ‘omics data. *Peer J* 3:e1319
- Farag IF, Davis JP, Youssef NH, Elshahed MS (2014) Global patterns of abundance, diversity and community structure of the Aminicenantes (candidate phylum OP8). *PLoS One* 9:e92139
- Federhen S (2012) The NCBI taxonomy database. *Nucleic Acids Res* 40:D136–D143
- Frischkorn KR, Stojanovski A, Paranjpye R (2013) *Vibrio parahaemolyticus* type IV pili mediate interactions with diatom-derived chitin and point to an unexplored mechanism of environmental persistence. *Environ Microbiol* 15:1416–1427
- Gies EA, Konwar KM, Beatty JT, Hallam SJ (2014) Illuminating microbial dark matter in meromictic Sakinaw Lake. *Appl Environ Microbiol* 80:6807–6818
- Greening C, Biswas A, Carere CR, Jackson CJ, Taylor MC, Stott MB, Cook GM, Morales SE (2016) Genomic and metagenomic surveys of hydrogenase distribution indicate H<sub>2</sub> is a widely utilised energy source for microbial growth and survival. *ISME J* 10:761–777
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321
- Hernsdorf AW, Amano Y, Miyakawa K, Ise K, Suzuki Y, Anantharaman K, Probst A, Burstein D, Thomas BC, Banfield JF (2017) Potential for microbial H<sub>2</sub> and metal transformations associated with novel bacteria and archaea in deep terrestrial subsurface sediments. *ISME J* 11:1915–1929
- Hugenholtz P, Pitulle C, Hershberger KL, Pace NR (1998) Novel division level bacterial diversity in a Yellowstone hot spring. *J Bacteriol* 180:366–376
- Hunt DE, Gevers D, Vahora NM, Polz MF (2008) Conservation of the chitin utilization pathway in the Vibrionaceae. *Appl Environ Microbiol* 74:44–51
- Hutcheson SW, Zhang H, Suvorov M (2011) Carbohydrase systems of Saccharophagus degradans degrading marine complex polysaccharides. *Mar Drugs* 9:645–665
- Kadnikov VV, Mardanov AV, Beletsky AV, Banks D, Pimenov NV, Frank YA, Karnachuk OV, Ravin NV (2018) A metagenomic window into the 2-km-deep terrestrial subsurface aquifer revealed multiple pathways of organic matter decomposition. *FEMS Microbiol Ecol* 94(10):fyy152. <https://doi.org/10.1093/femsec/fyy152>
- Kim YJ, Lee HS, Kim ES, Bae SS, Lim JK, Matsumi R, Lebedinsky AV, Sokolova TG, Kozhevnikova DA, Cha SS, Kim SJ, Kwon KK, Imanaka T, Atomi H, Bonch-Osmolovskaya EA, Lee JH, Kang SG (2010) Formate-driven growth coupled with H<sub>2</sub> production. *Nature* 467:352–355
- Konstantinidis KT, Rosselló-Móra R, Amann R (2017) Uncultivated microbes in need of their own taxonomy. *ISME J* 11:2399–2406
- Magnabosco C, Ryan K, Lau MC, Kuloyo O, Lollar BS, Kieft TL, van Heerden E, Onstott TC (2016) A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. *ISME J* 10:730–741
- Mandlik A, Swierczynski A, Das A, Ton-That H (2008) Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends Microbiol* 16:33–40
- Mardanov AV, Ravin NV, Svetlichnyi VA, Beletsky AV, Miroshnichenko ML, Bonch-Osmolovskaya EA, Skryabin KG (2009) Metabolic versatility and indigenous origin of the archaeon *Thermococcus sibiricus*, isolated from a siberian oil reservoir, as revealed by genome analysis. *Appl Environ Microbiol* 75:4580–4588
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J* 17:10–12
- Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform* 4:60
- Nobu MK, Dodsworth JA, Murugapiran SK, Rinke C, Gies EA, Webster G, Schwientek P, Kille P, Parkes RJ, Sass H, Jørgensen BB, Weightman AJ, Liu WT, Hallam SJ, Tsiamis G, Woyke T, Hedlund BP (2016) Phylogeny and physiology of candidate phylum ‘Atribacteria’ (OP9/JS1) inferred from cultivation-independent genomics. *ISME J* 10:273–286
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055
- Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36:996–1004
- Probst AJ, Castelle CJ, Singh A, Brown CT, Anantharaman K, Sharon I, Hug LA, Burstein D, Emerson JB, Thomas BC, Banfield JF (2016) Genomic resolution of a cold subsurface aquifer community

- provides metabolic insights for novel microbes adapted to high CO<sub>2</sub> concentrations. *Environ Microbiol* 19:459–474
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:D590–D596 (**Database issue**)
- Rinke C, Schwientek P, Sczyrba A et al (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499:431–437
- Robbins SJ, Evans PN, Parks DH, Golding SD, Tyson GW (2016) Genome-centric analysis of microbial populations enriched by hydraulic fracture fluid additives in a coal bed methane production well. *Front Microbiol* 7:731
- Rodriguez-R LM, Konstantinidis KT (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Prepr* 4:e1900v1. <https://doi.org/10.7287/peerj.preprints.1900v1>
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541
- Scigelova M, Crout DHG (1999) Microbial beta-*N*-acetylhexosaminidases and their biotechnological applications. *Enzyme Microb Technol* 25:3–14
- Sharon I, Kertesz M, Hug LA, Pushkarev D, Blauwkamp TA, Castelle CJ, Amirebrahimi M, Thomas BC, Burstein D, Tringe SG, Williams KH, Banfield JF (2015) Accurate, multi-kb reads resolve complex populations and detect rare microorganisms. *Genome Res* 25:534–543
- Sorokin DY, Gumerov VM, Rakitin AL, Beletsky AV, Sinninghe Damsté JS, Mardanov AV, Ravin NV (2014) Genome analysis of *Chitinivibrio alkaliphilus* gen. nov., sp. nov., a novel extremely haloalkaliphilic anaerobic chitinolytic bacterium from the candidate phylum Termita Group 3. *Environ Microbiol* 16:1549–1565
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313
- Wilson K (2001) Preparation of genomic DNA from bacteria. *Curr Protoc Mol Biol*. <https://doi.org/10.1002/0471142727.mb0204s56> (**chapter 2: unit 2.4**)