Systematic and Applied Microbiology 42 (2019) 67–76

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

Phylogeny and physiology of candidate phylum BRC1 inferred from the first complete metagenome-assembled genome obtained from deep subsurface aquifer

Vitaly V. Kadnikov[a], Andrey V. Mardanov[a], Alexey V. Beletsky[a], Andrey L. Rakitin[a], Yulia A. Frank[b], Olga V. Karnachuk[b], Nikolai V. Ravin[a,*]

[a] Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, 119071, Moscow, Russia
[b] Laboratory of Biochemistry and Molecular Biology, Tomsk State University, 634050, Tomsk, Russia

ARTICLE INFO

Keywords:
Candidate phylum BRC1
Uncultured bacteria
Metagenome
Deep subsurface

ABSTRACT

Candidate bacterial phylum BRC1 has been identified in a broad range of mostly organic-rich oxic and anoxic environments through molecular analysis of microbial communities. None of the members of BRC1 have been cultivated and only a few draft genome sequences have been obtained from metagenomes or as a result of single-cell sequencing. We have reconstructed complete genome of BRC1 bacterium, BY40, from metagenome of the microbial community of a deep subsurface thermal aquifer in the Tomsk Region of the Western Siberia, Russia, and used it for metabolic reconstruction and comparison with existing genomic data. Analysis of 3.3 Mb genome of BY40 bacterium revealed numerous glycoside hydrolases that could enable utilization of carbohydrates, including enzymes of chitin-degradation pathway. The bacterium lacks flagellar machinery but the twitching motility is encoded. The reconstructed central metabolism revealed pathways enabling the fermentation of organic substrates, as well as their complete oxidation through aerobic and anaerobic respiration. Phylogenetic analysis using BY40 genome supported the phylum level classification of BRC1 lineage. Based on phylogenetic and genomic analyses, the novel bacterium is proposed to be classified as *Candidatus* Sumerlaea chitinivorans, within a candidate phylum *Sumerlaeota*.

Introduction

The candidate bacterial phylum BRC1 ("Bacterial rice cluster 1") was originally revealed by phylogenetic analysis of 16S rRNA genes amplified from anoxic bulk soil of flooded rice microcosms [9]. During the next 17 years the 16S rRNA sequences assigned to this candidate phylum have been detected by molecular methods in diverse environments, including soils, activated sludge, anaerobic digesters and wastewater treatment reactors, marine and freshwater sediments, geothermal springs, and at present the SILVA database contains almost four thousand sequences classified in BRC1 [32]. The environments, where BRC1 bacteria were detected, are highly diverse, but typically they are rich in organic matter and vary in oxygen concentration. To the best of our knowledge, no microbial ecology study revealed BRC1 division as a dominant member of microbial communities suggesting that they could have a specialised metabolism like organotrophic bacteria of the uncultured candidate phyla *Aminicenantes* [35], *Saccharibacteria* [1], *Atribacteria* [28] and others.

Although the BRC1 remains yet uncultured, the first hints at the biology of BRC1 came from large-scale projects on sequencing of environmental single-cells genomes and metagenomes. At present four draft genomes of BRC1 members are available. Two single-cell genomes (AAA252-M09 and AAA257-C11) were sequenced by Rinke et al [34]; they are estimated to be 20–22% complete. Baker et al. [4] assembled approximately 43% complete composite genome of BRC1 bacterium (SM23_51) from a metagenome of estuary sediments. More recently, extensive metagenomic sequencing of DNA from groundwater collected at 160 m depth at the Horonobe Research Laboratory (Hokkaido, Japan) has yielded near-complete genome sequence of BRC1 bacterium HGW-BRC1-1, assembled in 36 contigs [15]. However, in no cases these draft genomes were used for metabolic reconstruction and analysis of the possible lifestyle of BRC1 bacteria.

Recently we used large-scale metagenome sequencing to study the microbial community of the deep subsurface thermal aquifer

* Corresponding author at: Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Leninsky prosp 33-2, Moscow, 119071, Russia.
E-mail address: nravin@biengi.ac.ru (N.V. Ravin).

https://doi.org/10.1016/j.syapm.2018.08.013
0723-2020/

located in the Western Siberia Region of the Russian Federation [19]. Microbial communities of the terrestrial deep subsurface aquatic ecosystems have been investigated in a number of studies and allowed to assembly numerous draft and even complete genomes, including representatives of uncultured prokaryotic lineages ([3,15,25,31] etc). Western Siberian basin, known as one of the largest oil reserves in the world, formed from marine sediments of the Mesozoic period. Besides oil, a number of underground water bodies were found at depths of 1–3 km in course of oil-exploration drilling. Some boreholes remained opened and provide a unique opportunity to access microbial communities of these subsurface aquifers [11,20]. We have analysed the composition of microbial community and sequenced metagenome of the 2-km-deep subsurface thermal aquifer accessible through the borehole 1-R located in the Tomsk Region, Russia [19]. Microbial community mostly consisted of sulfate-reducing *Firmicutes* and *Deltaproteobacteria*, and uncultured lineages of the phyla *Chlorofexi*, *Ignavibacteriae*, and *Aminicenantes* [19]. Members of several bacterial candidate divisions, including BRC1, were also detected.

Here we report the recovery of complete genome sequence of a member of the candidate phylum BRC1 from the metagenome of this subsurface aquifer, accurate reconstruction of its metabolism and prediction of ecological role of this uncultured bacterial lineage.

Materials and methods

Site description, sampling, field measurements, and DNA isolation

The oil-exploration borehole 1-R is located near the town Byelii Yar in Tomsk region, Western Siberia, Russia (58.4496N, 85.0279E). This oil-exploration borehole was drilled in 1961–1962 to 2563 m deep but no oil was found. The borehole passed Quaternary, Paleogene and Cretaceous sediments and at 2505 m entered Paleozoic basement, comprising sedimentary rocks up to 2534 m and then basalts [6]. The water flowing out of the borehole is expected to be originated from an aquifer system at 1997–2005 m depth interval [6].

Water samples were taken from the borehole on August 04-05, 2014. The groundwater was about 43 °C in temperature, had slightly alkaline pH (8.5) and was highly reduced (Eh from −341 to −279 mV). Ionic content of the water is dominated by sodium and chloride, and the water is likely to be derived from marine salts of relict ocean [6]. The total mineralization of the water is however only about 1.8 g/L representing only 5% of marine salinity. Therefore most of the water is derived form meteoric recharge [6,19].

Cells from 50 L of water were collected on 0.22 μm cellulose nitrate filters using a Sartorius filtration unit. The filters were frozen in liquid nitrogen and then ground and melted with TE buffer in a water bath at 37 °C. The total DNA was extracted by the CTAB/NaCl method [41].

Metagenome sequencing and assembly, contig binning, and analysis of the composite genomes

Metagenomic DNA sample was sequenced using the Illumina HiSeq2500 platform (250 nt single end mode) according to the manufacturer's instructions (Illumina Inc., USA) as described previously [19]. A total of 86.5 million high quality sequencing reads (a total of 18.5 Gbp) were obtained after primer and quality trimming with Cutadapt [27] and Sickle (https://github.com/najoshi/sickle), respectively. The assembly of the contigs was carried out using SPAdes Genome Assembler [5], specifying – meta parameter indicating metagenome assembly.

The program CONCOCT [2] was used for binning of contigs as reported earlier [19]. The completeness and contamination of the recovered metagenome-assembled genomes (MAG) were estimated using CheckM v. 1.05 [30] with lineage-specific marker genes. One of the obtained bins, BY40, comprised two large contigs (2.63 Mb and 0.65 Mb long). The search for continuations of contigs using Bandage v. 0.8.0 visualization tool [40], allowed to join them in a single linear molecule. PCR amplification and Sanger sequencing of the obtained fragment was used to close a single remaining gap and obtain complete circular genome sequence of this bacterium, designated BY40. Correctness of the assembly was verified by mapping Illumina reads back to the assembled complete genome sequence using Bowtie 2 [23], and no evidences of misassembly were found.

Gene search and annotation of the MAG assigned to candidate phylum BRC1 were performed using the RAST server 2.0 [7], followed by manual correction by searching the National Center for Biotechnology Information (NCBI) databases. Signal peptides were predicted using Signal P v.4.1 for Gram-negative bacteria (http://www.cbs.dtu.dk/services/SignalP/). The N-terminal twin-arginine translocation (Tat) signal peptides were predicted using PRED-TAT (http://www.compgen.org/tools/PRED-TAT/) and the transmembrane helices with TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). The rate of genomic DNA replication was calculated using the index of replication as implemented in iRep software [8].

Search for genomes related to BY40

In order to find genome assemblies closely related to bin BY40, we identified homologs of 122 conserved marker genes in the NCBI NR database using BLASTP search. Genome assemblies corresponding to top 100 hits for each of the marker genes (a total of 992) were selected. The average amino acid identity (AAI) between BY40 and the selected 992 genome assemblies was calculated using aai.rb script from the Enveomics Collection [36].

Phylogenetic analysis

CheckM was used to find single copy marker genes in the assembled BRC1 genome and to construct multiple alignment of concatenated single copy genes sequences, comprising ones from BRC1 MAG and all species presented in CheckM database. Four previously sequenced partial genomes of BRC1 bacteria were additionally included in this analysis: SM23_51 (LJUL00000000.1), SCGC AAA252-M09 (ASKJ00000000.1), SCGC AAA257-C11 (ASML00000000.1), and HGW-BRC1-1 (PHCB00000000.1). Selected part of the CheckM multiple alignment was used for the tree construction in PhyML [14] using default parameters. The support values for the internal nodes were estimated by approximate Bayesian tests in PhyML.

The 16S rRNA sequences were aligned using MUSCLE included in MEGA 6.0 [39]. The maximum likelihood phylogenetic tree was computed by MEGA 6.0, using Tamura-Nei substitution model and uniform rates among sites. Positions containing gaps and missing data were eliminated. Bootstrap tests were performed with 100 resamplings. 16S rRNA sequence from *Escherichia coli* was used to root the tree.

Expression of recombinant endochitinase BY40_1901 and testing its functional activity

The gene BY40_1901 was amplified from metagenomic DNA by PCR using primers BY40_1901F (5'- ATAAGATCTACGACGACGGTTTCCGCGAGCA) and BY40_1901R (5'- ATAAGCTTTAGGGTTCGCGTGAGTTGCTTTCCCA). The resulting PCR product was digested with BglII and HindIII and inserted into pQE30 (Qiagen) at BamHI and HindIII sites, yielding the plasmid pQE80-

BY40_1901. This plasmid was transformed into *Escherichia coli* strain DLT1270. The recombinant strain was grown at 37 °C in LB medium supplemented with ampicillin and induced to express recombinant enzyme by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1.0 mM at an optical density at 600 nm of approximately 0.5 and incubated further at 30 °C for 4 h.

Cells were collected by centrifugation at 3500 *g* for 15 min at 4 °C and resuspended in 50 mM phosphate buffer (pH 7.5), 0.3 M NaCl, and 5 mM imidazole. Upon sonication the cell extract was centrifuged (15,000 *g*, 4 °C, 30 min), and the recombinant proteins from the supernatant were purified by metal affinity chromatography using a Ni-NTA spin kit (Qiagen). Upon elution from the column, the proteins were dialyzed against 12.5 mM phosphate buffer (pH 6.0) at 4 °C for 3 h.

Chitinolytic activities of recombinant BY40_1901 protein were measured in a fluorimetric assay with 4-methylumbelliferyl (4-MU) derivatives using a chitinase assay kit (CS1030; Sigma). The following substrates were used: 4-MU-β-D-N, N′, N″-triacetylchitotriose, 4-MU-N, N′-diacetyl-β-D-chitobioside, and 4-MU-N-acetyl-β-D-glucosaminide. Assays were performed in triplicate at 50 °C in 100 mM citrate buffer (pH 4.5). One unit of activity was defined as the amount of protein required to release 1 nmol of 4-MU from the appropriate substrate per minute under described conditions. The temperature dependency of enzyme activity was determined with 4-MU-N, N′-diacetyl-β-D-chitobioside in 100 mM citrate buffer (pH 4.5).

Data deposits

Metagenomic reads were deposited in NCBI Sequence Read Archive under the accession number SRR710274. The annotated genome sequence of the candidate phylum BRC1 bacterium BY40 has been deposited in NCBI/GenBank database under accession number CP030759.

Results and discussion

Reconstruction of complete genome of a member of candidate phylum BRC1

In order to assemble the composite genomes of the members of the microbial community, a total of 18.5 gigabases of metagenomic data were generated and assembled into contigs, which were distributed among genome bins [19]. One bin (BY40) consisted of only two large contigs, sequenced to 37× average coverage. The relative abundance of this genotype, defined as a fraction of this bin in the whole metagenome, was about 0.8%. Upon closure of gaps the contigs were joined into a circular 3289105 bp long genome sequence. Analysis of the presence of 143 conserved single-copy marker genes using CheckM estimated the completeness of this genome to be 99.4%.

A single 23S-5S ribosomal RNA operon, a single distantly located 16S rRNA gene, and 46 transfer RNA (tRNA) genes coding for all of the 20 amino acids were identified. Annotation of the genome sequence predicted 2851 protein-coding genes (Table 1). A rather unusual feature of this genome was almost complete absence of genes related to mobile elements since only two transposases were identified. The BY40 genome contains four Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) loci with 4, 59, 81 and 103 spacer-repeat units and a set of CRISPR-associated genes of type 2 system. It is possible that this extensive CRISPR array provides efficient protection against phage and mobile elements invasions.

Table 1
General characteristics of the genomes of BY40 and HGW-BRC1-1.

Genome properties	BY40	HGW-BRC1-1
Genome size (Mb)	3.29	3.77
Contigs	1	36
Completeness	finished	99.4%
G + C content (%)	56.0	58.4
rRNA genes	3	3
tRNA genes	46	46
Protein-coding genes	2851	2947
of them functionally assigned	1843	1269

Analysis of the genome of BY40 bacterium revealed no genes encoding flagellar machinery. However, we found a set of genes necessary for generation of Type IV pili (Supplemental table S1). These pili enable twitching motility of the bacterium and its adhesion to solid surfaces [26].

Replication rate of BY40 in the subsurface aquifer

In order to determine whether the BY40 cells are actively growing in the subsurface aquifer or remain inactive, we evaluated the rate of DNA replication *in situ* using the iRep software [8]. The programm calculates an index of replication (iRep) based on the different sequencing coverage across the genome that results from higher abundance of the origin-proximal regions in actively replicating cells. The iRep values >1 indicate active DNA replication. For BY40 we obtained iRep value of 1.13, which is lower than iRep values obtained for metabolically active lineages in human guts but comparable to that observed for underground microorganisms [8]. Note that observed iRep value likely underestimates the actual rate of DNA replication *in situ* since the water has cooled from the moment of sampling until the end of the collection of cells by filtration, which probably reduced the rate of growth and replication of DNA. Nevertheless, the iRep data show that BY40 cells are likely metabolically active in the subsurface aquifer.

Phylogenetic placement of BY40

The 16S rRNA gene found in the BY40 genome has only 80% sequence identity with the nearest cultured bacteria (deltaproteobacteria *Deferrisoma camini*), indicating a phylum-level position of this bacterium. Even considering uncultured bacteria, GenBank searches (Apr 01, 2018) revealed no sequences with more than 94% identity to 16S rRNA sequence of the new bacterium further supporting the novelty of this lineage. However, many 16S rRNA sequences with 85–93% identity were described as belonging to members of the candidate phylum BRC1. A search of the 16S rRNA gene against the SILVA database [32] also classified the new bacterium as a member of the candidate phylum BRC1.

Search for relatives of BY40 on the basis of genome-to-genome distance evaluation revealed that BRC1 bacterium HGW-BRC1-1 is the closest relative of BY40 with an average AAI of 54.4%. The AAI with other genomes was below 43%. To investigate the phylogenetic position of the new bacterium in more detail, we constructed phylogenetic tree based on concatenated conservative marker genes (Fig. 1). The results show that BY40 together with four other uncultured bacteria, previously described as members of the candidate phylum BRC1, formed a monophyletic separate phylum-level lineage. These data altogether support, on a complete genome scale, the phylum level position of BRC1.

Prediction of central metabolic pathways

The BY40 genome contains a complete set of genes encoding enzymes of the Embden–Meyerhof glycolytic pathway, including

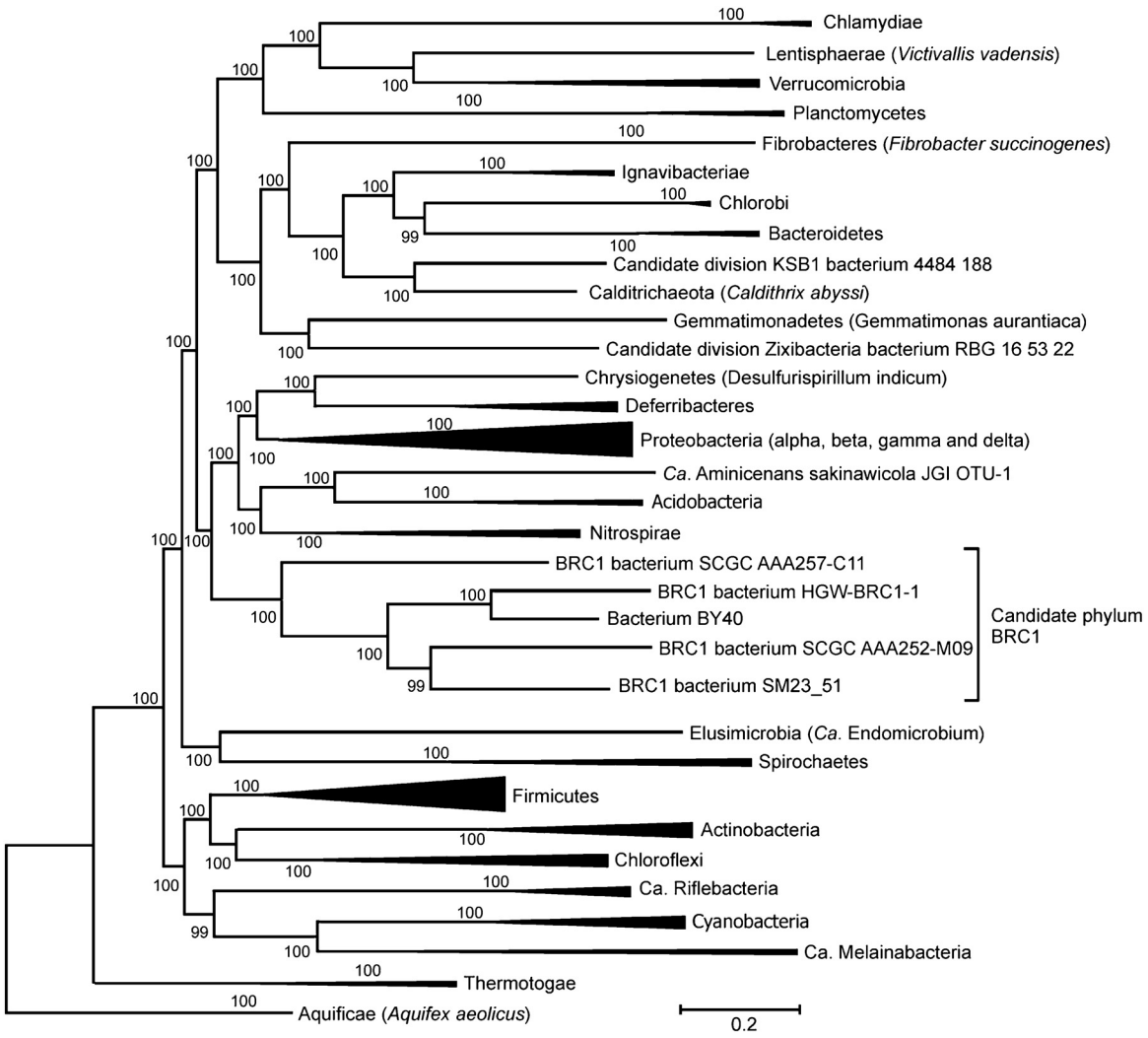

Fig. 1. Position of BY40 in the maximum likelihood concatenated protein phylogeny.
Selected part of the CheckM multiple alignment was used for the tree construction in PhyML using default parameters. The tree was inferred from the concatenation of 43 conserved marker genes. The support values for the internal nodes were estimated by approximate Bayes tests in PhyML.

glucokinase, glucose-6-phosphate isomerase, pyrophosphate-dependent fructose 6-phosphate-1-kinase, fructose-bisphosphate aldolase, triosephosphate isomerase, glyceraldehyde 3-phosphate dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase, enolase, and pyruvate, phosphate dikinase (Fig. 2, Supplemental Table S1). Only non-oxidative branch of the pentose phosphate pathway is encoded. Pyruvate generated in the glycolysis could be decarboxylated to acetyl-coenzyme A (CoA) by pyruvate:feredoxin oxidoreductase producing reduced ferredoxin. Phosphate acetyltransferase and acetate kinase could than convert acetyl-CoA to acetate via two-step reaction with the concomitant production of ATP. Another fermentation product could be ethanol, as indicated by the presence of aldehyde and alcohol dehydrogenases.

Genome analysis revealed three group 3 bidirectional [NiFe]-hydrogenases. The enzymes lacked recognizable N-terminal signal peptides and transmembrane helices that indicate their cytoplasmic location. The first belongs to group 3c heterodisulfide reductase-linked [NiFe]-hydrogenases that bifurcates electrons from H_2 to heterodisulfide and ferredoxin [13]. The second is pentameric, made of the hydrogenase (HoxYH) and the diaphorase (HoxFUE) moieties, group 3d [NiFe]-hydrogenase. These enzymes interconvert electrons between H_2 and NAD depending on redox state of the cell [13]. The last one is tetrameric group 3b [NiFe]-hydrogenase capable of coupling oxidation of NADPH to evolution of H_2. Some group 3b enzymes are bifunctional and also have sulfhydrogenase activity reducing S^0 to H_2S [13]. Production of hydrogen could provide an electron sink for fermentative pathways. Altogether, these hydrogenases are also able to interconvert electrons between ferredoxin, NAD(P)H, and H_2 depending on the availability of exogenous electron acceptors.

In addition to fermentation, BY40 bacterium was predicted to be capable of complete oxidation of organic substrates via oxidative phosphorylation (Fig. 2). In the presence of external electron acceptors, acetyl-CoA generated from pyruvate may be completely oxidized via the tricarboxylic acid cycle, which is encoded by the genome. The BRC1 genome encodes all of the major components of the electron transfer chain, including the proton-translocating NADH-dehydrogenase complex, membrane-bound succinate dehydrogenase (complex II), the quinol-oxidizing alternative Complex III [43] instead of typical cytochrome bc_1 complex, and membrane-linked terminal oxidoreductases. Membrane-bound pyrophosphatase could also contribute to translocation of protons across the cytoplasmic membrane. The resulting transmembrane proton gradient may be used for ATP generation by the membrane F_0F_1-type ATP synthase.

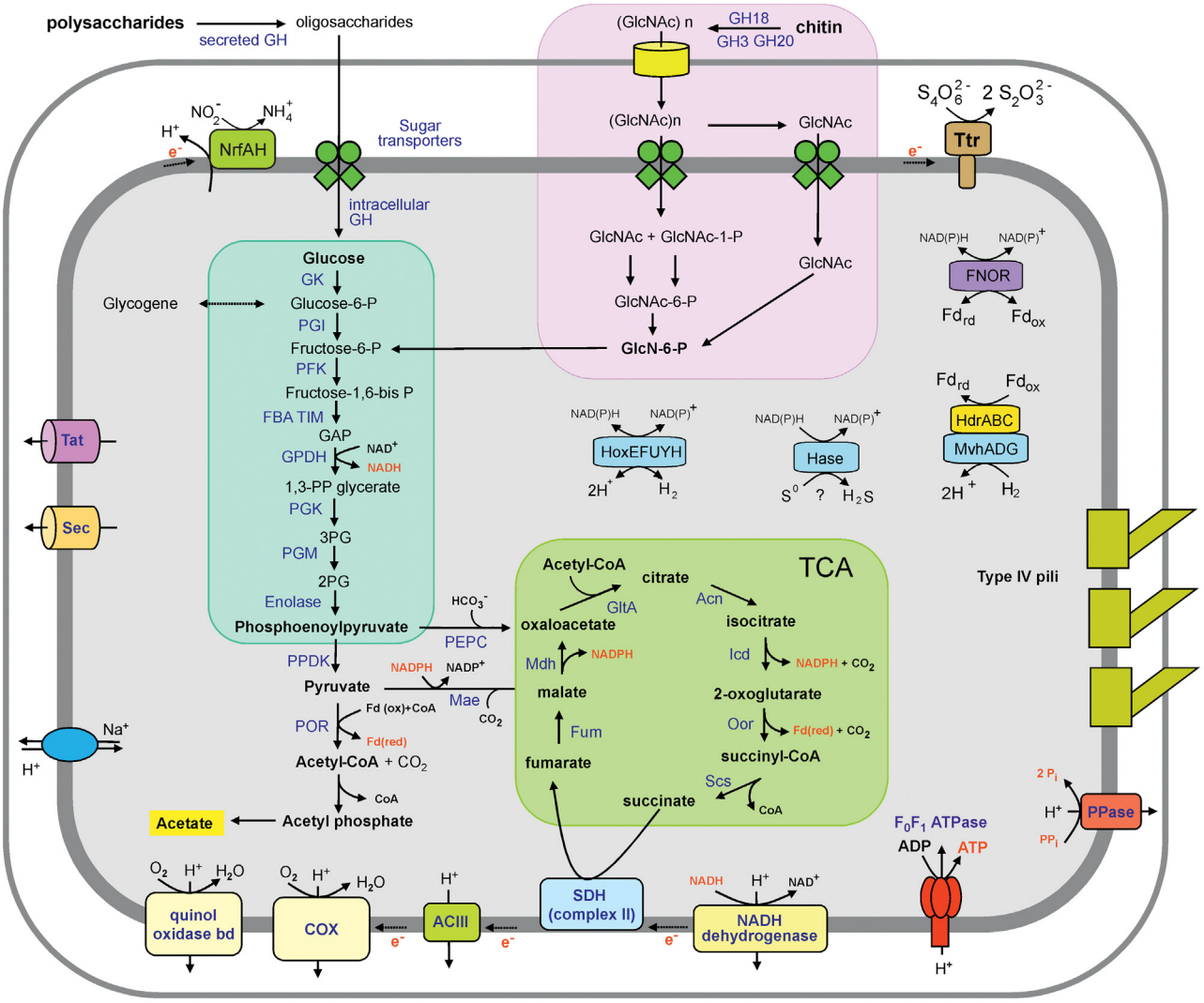

Fig. 2. An overview of the metabolism of BY40 bacterium (*Candidatus* Sumerlaea chitinivorans) reconstructed from its genome.
Enzyme abbreviations: GH, glycoside hydrolase; GK, glukokinase; PGI, glucose-6-phosphate isomerase; PFK, 6-phosphofructokinase; FBA, fructose-bisphosphate aldolase; TIM, triosephosphate isomerase; GPDH, glyceraldehyde 3-phosphate dehydrogenase; PGK, phosphoglycerate kinase; PGM, phosphoglycerate mutase; PPDK, pyruvate, phosphate dikinase; POR, pyruvate ferredoxin oxidoreductase; PEPC, phosphoenolpyruvate carboxylase;GltA, citrate synthase; Acn, aconitase; Icd, isocitrate dehydrogenase; Oor, 2-oxoglutarate oxidoreductase; Scs, succinyl-CoA synthetase; SDH, succinate dehydrogenase; Fum, fumarase; Mdh, malate dehydrogenase; Mae, malic enzyme; ACIII, alternative complex III; COX, cytochrome c oxidase; Hase, group 3b [NiFe] hydrogenase; NrfAH, cytochrome c nitrite reductase; Ttr, tetrathionate reductase; FNOR, ferredoxin-NAD(P)$^+$ reductase; PPase, pyrophosphatase. Other abbreviations: ox/red, oxidized and reduced forms; Pi, phosphate; PPi, pyrophosphate; CoA, coenzyme A.

Terminal oxygen reductases are represented by a quinol oxidase *bd* complex and proton-translocating cytochrome c oxidase (Fig. 2). The cytochrome c oxidase is encoded by a five-gene operon located downstream of a cluster of genes for alternative complex III what represents a typical genomic context of the alternative complexes [33]. The *cox* operon included gene encoding the cytochrome oxidase Cu insertion factor of SCO1/SenC family, followed by genes for subunits CoxII, CoxI, CoxIII and Cox IV. Genes encoding heme A synthase CtaA and heme O synthase CtaB (protohaem IX farnesyltransferase necessary to produce haem *o* from a haem *b* precursor) are located distantly. The ability of BY40 bacterium to grow under aerobic conditions is consistent with the presence of superoxide dismutase participating in protection against reactive oxygen species.

The search for potential terminal oxidoreductases of anaerobic respiratory pathways revealed the presence of two complexes (Fig. 2). The first is membrane-bound molybdopterin oxidoreductases of the CISM family. The catalytic subunit of this complex contains the N-terminal Tat signal peptide and shares homology with tetrathionate reductases, enzymes responsible for the anaerobic reduction of tetrathionate to thiosulfate. The presence of this enzyme can enable the use of sulfur compounds or another electron acceptors under anaerobic conditions. The second is an NrfAH-like nitrite reductase complex, encoded by a cluster of two genes that encode two multihaem c-type cytochromes homologous to the catalytic subunit NrfA and the membrane-bound subunit NrfH [10]. The pathways of dissimilatory reduction of nitrate, sulfate and iron were not detected.

Possible growth substrates

Analysis of the genome of BY40 bacterium predicted that it could be able to degrade some oligo- and polysaccharides. The utilization of starch and similar polymers may be enabled by eight alpha-amylases (Supplemental Table S1). Three of them were predicted to contain N-terminal secretion signal peptide, indicating their extracellular operation. Consistently, the BY40 genome encodes maltose/maltodextrin ABC transport system, and

enzymes for downstream maltose metabolism (alpha-glucosidase etc). Although known enzymes enabling extracellular hydrolysis of beta-linked polysaccharides such as cellulose and xylan were not identified, the presence of three signal peptide-containing glycoside hydrolases of the uncharacterized GH10L family (Pfam2638) indicates that BRC1 bacterium could degrade a wider range of polysaccharides.

Genome analysis also revealed two signal-peptide containing serine proteases, S8 family subtilisin-like protease and M8 family peptidase, which could enable extracellular hydrolysis of proteinaceous substrates.

A complete set of enzymes that could enable utilization of chitin [16] is encoded in the genome of BY40 bacterium. The extracellular hydrolysis of chitin could be performed by a GH18 family endochitinase BY40_1901, a 395-amino-acid (aa) protein predicted to carry *N*-terminal secretion signal. GenBank searches revealed the most similar proteins (45–50% amino acid sequence identity) in an unknown uncultured bacterium and in another member of the candidate phylum BRC1, bacterium HGW-BRC1-1. The search for other chitin degrading enzymes from GH families 18, 19 and 20, revealed only GH20 family protein BY40_754. This enzyme, carrying an *N*-terminal signal peptide and annotated as *N*-acetyl-β-hexosaminidase, could cleave *N*-acetyl-D-glucosamine (GlcNAc) from the non-reducing end of the soluble chitin oligomers [37], produced by extracellular endochitinase. Hydrolysis of soluble chitin oligomers could be also enabled by two GH3 family beta-hexosaminidases [17], one of which was predicted to contain an *N*-terminal secretion signal.

All above mentioned enzymes likely involved in chitin utilisation, lacked recognisable chitin-binding domains. It is possible that adherence of BY40 bacterium to chitin may be facilitated by type IV pili in a similar manner as some cellulolytic bacteria, such as *Fibrobacter succinogenes*, are adhering cellulose as substrate [18]. Among chitinolytic bacteria the role of type IV pilins for adherence to insoluble chitin was reported for *Vibrio parahaemolyticus* and *Chitinivibrio alkaliphilus* [12,38].

GlcNAc, chitobiose, and other products of hydrolysis of chitooligosaccharides may be imported into the cytoplasm by ABC-type transporters [24] and N-acetylglucosamine related transporter NagX. Chitobiose could be cleaved into GlcNAc and GlcNAc-1-P by GH94 family cellobiose/chitobiose phosphorylase [29]. Then N-acetylglucosamine-1-P-mutase converts GlcNAc-1-P into GlcNAc-6-P. GlcNAc may be phosphorylated to make GlcNAc-6-P by N-acetylglucosamine kinase NagC. Finally, *N*-acetylglucosamine-6-phosphate deacetylase NagA deacetylate GlcNAc-6-P yielding glucosamine 6-phosphate that could be converted by glucosamine-6-phosphate deaminase into fructose-6-phosphate entering the Embden-Meyerhof pathway of glycolysis.

Functional characterisation of recombinant endochitinase BY40_1901

Since the presence of complete chitinolytic pathway is a characteristic feature of the BY40 bacterium and, probably, other members of candidate division BRC1, we performed functional analysis of the predicted extracellular chitinase. The chitinase gene BY40_1901 lacking the 5′-terminal region coding for *N*-terminal secretion signal was expressed in *E. coli*. The recombinant protein was purified to homogeneity through Ni-NTA affinity chromatography. Chitinolytic activities of recombinant endochitinase were evaluated with synthetic soluble substrates. High hydrolytic activities were observed with the endochitinase substrate 4-MU-β-D-N, N′, N″-triacetylchitotriose (287 U/mg) and with 4-MU-N, N′-diacetyl-β-D-chitobioside (342 U/mg), while much lower value was detected with 4-MU-*N*-acetyl-β-D-glucosaminide (0.7 U/mg).

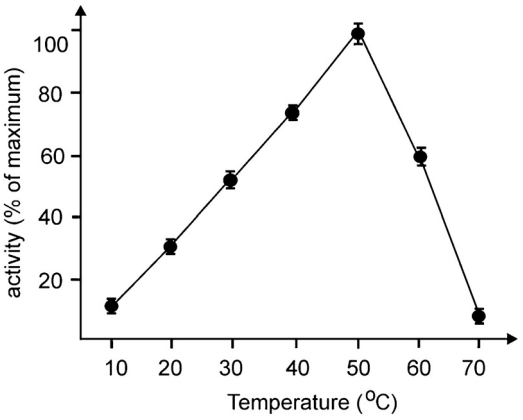

Fig. 3. Effect of temperature on the activity of the recombinant endochitinase BY40_1901.

This substrate specificity pattern is consistent with the predicted endochitinase activity of BY40_1901.

Considering the temperature dependence, the maximum activity was measured at 50 °C (Fig. 3), that is close to the temperature of the water at the borehole wellhead.

Comparative genomics of BRC1 bacteria

Availability of a near complete genome of another member of the candidate phylum BRC1, HGW-BRC1-1, allowed to compare their genome properties and metabolic potential. HGW-BRC1-1 genome was assembled from a metagenome of groundwater collected at 160 m depth in the sedimentary rocks in Japan [15]. This genome has comparable features with BY40 genome such as; size (3.77 Mb), G + C content and coding potential, with predicted 2947 protein-coding and 46 tRNA genes (Table 1). The 16S rRNA gene sequences of BY40 and HGW-BRC1-1 bacteria are 90% identical. Pairwise comparison between the BY40 and HGW-BRC1-1 genomes revealed 1883 common protein-coding genes with an average AAI of 54.4%. According to the sequence similarity thresholds proposed by in Konstantinidis et al. for uncultivated microorganisms [22], BY40 and HGW-BRC1-1 represented different families of a single order. Three other partial BRC1 genomes, SM23_51, AAA252-M09 and AAA257-C11, were phylogenetically more distant from BY40 (Fig. 1), but the low quality of these drafts (<50% completeness) and the lack of near full size 16S rRNA gene sequences in the assemblies prevented direct phylogenetic comparison.

968 and 992 genes were specific for BY40 and HGW-BRC1-1, respectively. Interestingly, most of species-specific proteins (715 of 968 in BY40 and 576 of 992 in HGW-BRC1-1) are hypothetical proteins with unknown functions, while most of metabolically-important functions were conserved in the two genomes. Like BY40, the HGW-BRC1-1 bacterium lacks flagellar machinery but contain a set of genes for Type IV pili. HGW-BRC1-1 genome contains a number of glycoside hydrolases suggesting that this bacterium can utilise various polysaccharides. Notably, complete chitinolytic pathway similar to that found in BY40 is encoded in HGW-BRC1-1 genome. The Embden-Meyerhof glycolytic pathway, the tricarboxylic acid cycle, and an aerobic respiratory chain including NADH dehydrogenase, succinate dehydrogenase, alternative complex III, and membrane-linked cytochrome *c* oxidase were found in HGW-BRC1-1 genome. The presence of ammonia-forming cytochrome *c* nitrite reductase in HGW-BRC1-1 genome could enable the use of nitrite as an anaerobic electron acceptor, as proposed for BY40. On the contrary, close homologs of membrane-

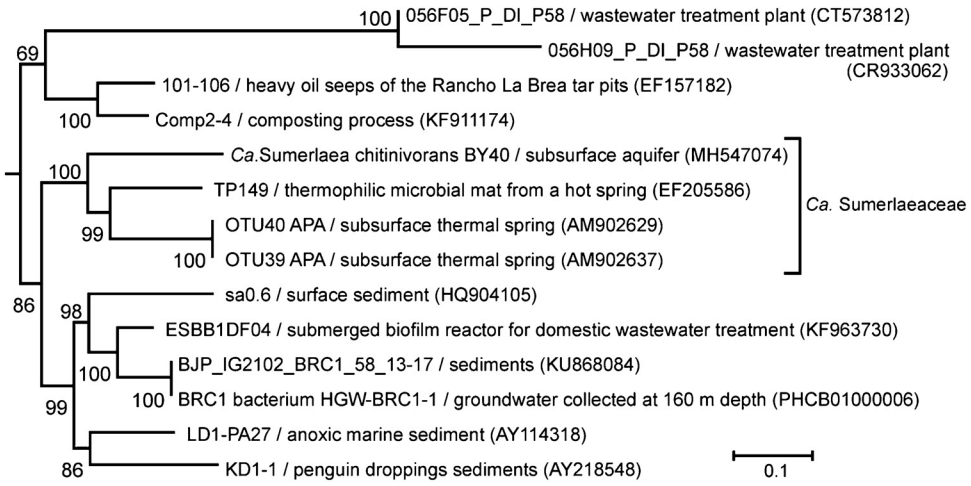

Fig. 4. Position of BY40 in the maximum likelihood 16S rRNA gene phylogeny.
The tree was computed by MEGA 6.0, using Tamura-Nei substitution model and uniform rates among sites. Isolation source and GenBank accession numbers are shown after the clone names. The scale bar represents substitutions per nucleotide base. Only BY40-proximal piece of original phylogenetic tree comprising 104 16S rRNA sequences of BRC1 members is shown (full tree is presented in Supplementary Fig. S1).

bound molybdopterin oxidoreductase found in BY40 are absent in HGW-BRC1-1 genome.

The presence of hydrogenases in HGW-BRC1-1 bacterium was reported previously [15], but the hydrogenase content of two bacteria differed significantly. Only group 3d [NiFe] NAD-reducing hydrogenase found in BY40 is present in HGW-BRC1-1, while group 3c and group 3b [NiFe] hydrogenases are missing. Instead, HGW-BRC1-1 bacterium has [FeFe]-hydrogenase of group A3 and group 1a respiratory H_2-uptake [NiFe]-hydrogenase. Probably, group A3 [FeFe]-hydrogenase, that reversibly bifurcates electrons from H_2 to ferredoxin and NAD, contributed to the maintenance of the redox state of the cell. The presence of group 1a [NiFe] hydrogenase suggested that HGW-BRC1-1 bacterium, unlike BY40, could be capable of anaerobic hydrogenotrophic respiration. The presence of a [NiFe] anaerobic uptake hydrogenase was reported also for BRC1 bacterium SCGC AAA257-C11 [13] suggesting that hydrogenotrophic respiration could be widespread in BRC1 division.

Diversity and distribution of BRC1 members

A nucleotide BLAST search revealed a number of environmental clones that were distantly related to the 16S rRNA sequence of BY40 bacterium, with sequence identities of less than 94%. The most closely related sequences were detected in subsurface thermal spring in Austria (clones OTU40/APA and OTU39/APA with 94% identity), and thermophilic microbial mat from a hot spring in Tibet (clone TP149, 93% identity). On the 16S rRNA phylogenetic tree these clones clustered with BY40 and could be assigned to the same family (Fig. 4 and Supplemental Fig. S1 for extended dataset). Several other sequences 90–92% identical to BY40 16S rRNA gene fell into a cluster that also includes HGW-BRC1-1 bacterium, – clone sa0.6 from unidentified surface sediment, clone LD1-PA27 from anoxic marine sediment, clone ESBB1DF04 from submerged biofilm reactor for domestic wastewater treatment, and clone BJP_IG2102_BRC1_58_13-17 from unidentified sediments sample (Fig. 4).

The SILVA database (accessed on Apr 01, 2018) contained a total of 3739 16S rRNA sequences assigned to the candidate phylum BRC1, of them about a thousand sequences represented near complete genes. They were identified in various environments including activated sludge, marine sediments, composting samples and anaerobic digesters, different soils, biofilms in freshwater bodies, geothermal spring mat, hypersaline microbial mat, oil fields etc (Supplemental Table S2). Nevertheless, the majority of BRC1 16S rRNA sequences were retrieved from organic-rich environments, particularly, biofilms and microbial mats. Such distribution is consistent with physiological features predicted by genome analysis.

Hydrolytic potential of BY40 could determine its ecological function as destructor of organic matter in organic-rich environments like biofilms in subsurface aquifer. Complex polysaccharides could originate from sediments of marine origin, buried since the formation of Western Siberian basin in the Mesozoic period. The capacity of anaerobic respiration and fermentation allowed this bacterium to grow in oxygen-limited niches. However, BY40 bacterium retained the aerobic respiration pathway that became advantageous upon reappearing in aerobic environment. At least some environments where BRC1 members were found, particularly soils and aquatic systems, could experience shifts between aerobic and aerobic conditions. The capacity for aerobic respiration was rather surprising for BY40 bacterium, found in a deep subsurface, generally believed to be anoxic [21]. However, analysis of the composition of gas, dissolved in thermal water flowing out of the borehole, revealed the presence of oxygen, accounting for ~7% of exsolved gas. The oxygen could enter the aquifer with meteoric recharge water [19]. Therefore, catabolic versatility and adaptation to both aerobic and anaerobic conditions of BY40 represents a successful strategy of adaptation to variable environmental conditions and colonization of novel habitats. The same traits could enable global distribution of the whole BRC1 lineage.

The genome of BY40 bacterium meets the criteria, recently suggested for description of new taxa of uncultivated Bacteria and Archaea [22], and we propose the following taxonomic names for the novel genus and species of BY40 (Table 2):

- Description of the novel genus *Candidatus* Sumerlaea (Su.mer.lae'a. N.L. fem. n. *Sumerlaea*, genus named for Sumerla, the underground goddess in Slavic mythology).
- Description of the novel species *Candidatus* Sumerlaea chitinivorans (chi.ti.ni.vo'rans. N.L. neut. n. *chitinum* chitin; L.v. *voro* to devour; N.L. part. adj. *chitinivorans* chitin devouring).

Not cultivated. Inferred to be facultatively anaerobic, obligate organotroph, obtains energy by fermentation, aerobic or anaerobic respiration, and able to use starch and chitin as a growth substrates. Represented by complete genome (acc. no. CP030759) obtained

Table 2
A digital protologue for *Candidatus* Sumerlaea chitinivorans and the corresponding higher taxa.

TAXONUMBER	*CA00037*
SPECIES NAME	*Candidatus* Sumerlaea chitinivorans
GENUS NAME	*Candidatus* Sumerlaea
SPECIFIC EPITHET	chitinivorans
GENUS ETYMOLOGY	*Candidatus* Sumerlaea (Su.mer.lae'a N.L. fem. n. Sumerlaea, genus named for Sumerla, an underground goddess in Slavic mythology)
TYPE SPECIES OF THE GENUS	*Candidatus* Sumerlaea chitinivorans
GENUS STATUS	gen. nov.
SPECIES STATUS	sp. nov.
SPECIES ETYMOLOGY	*Candidatus* Sumerlaea chitinivorans (chi.ti.ni.vo'rans. N.L. neut. n. *chitinum* chitin; L.v. *voro* to devour; N.L. part. adj. *chitinivorans* chitin devouring)
FAMILY NAME	*Candidatus* Sumerlaeaceae
FAMILY STATUS	fam. nov.
ORDER NAME	*Candidatus* Sumerlaeales
ORDER STATUS	ord. nov.
CLASS NAME	*Candidatus* Sumerlaeia
CLASS STATUS	classis nov.
AUTHORS	Kadnikov VV, Mardanov AV, Beletsky AV, Rakitin AL, Frank YA, Karnachuk OV, Ravin NV
TITLE	Phylogeny and physiology of candidate phylum BRC1 inferred from the first complete metagenome-assembled genome obtained from deep subsurface aquifer
JOURNAL	Systematic and Applied Microbiology
CORRESPONDING AUTHOR	Nikolai V. Ravin
E-MAIL OF THE CORRESPONDING AUTHOR	nravin@biengi.ac.ru
SUBMITTER	
E-MAIL OF THE SUBMITTER	nravin@mail.ru
DESIGNATION OF THE TYPE MAG	BY40
16S rRNA GENE ACCESSION NUMBER	MH547074
METAGENOME ACCESSION NUMBER	MAG
MAG/SAG ACCESSION NUMBER [RefSeq]	CP030759
GENOME STATUS	complete
GENOME SIZE	3,289,105 bp
GC mol %	56.0
COUNTRY OF ORIGIN	Russia
REGION OF ORIGIN	Tomsk region
SOURCE OF SAMPLE	groundwater
SAMPLING DATE	2014-08-04
GEOGRAPHIC LOCATION	Byelii Yar
LATITUDE	58.4496N
LONGITUDE	85.0279E
DEPTH	2000
TEMPERATURE OF THE SAMPLE [In Celsius degrees]	43
pH OF THE SAMPLE	8.5
RELATIONSHIP TO O2	aerotolerant
ENERGY METABOLISM	chemoorganotroph
DNA EXTRACTION METHOD	CTAB/NaCl method. Wilson K. (2001) Preparation of genomic DNA from bacteria. Curr. Protoc. Mol. Biol. Chapter 2: Unit 2.4. doi: 10.1002/0471142727.mb0204s56.
ASSEMBLY	1 sample
SEQUENCING TECHNOLOGY	Illumina HiSeq2500
BINNING SOFTWARE USED	CONCOCT
ASSEMBLY SOFTWARE USED	SPAdes Genome Assembler
HABITAT	deep subsurface terrestrial aquifer
BIOTIC RELATIONSHIP	free-living
MISCELLANEOUS, EXTRAORDINARY FEATURES RELEVANT FOR THE DESCRIPTION	Analysis of 3.3 Mb genome of BY40 bacterium revealed numerous glycoside hydrolases that could enable utilization of carbohydrates, including enzymes of chitin-degradation pathway. The bacterium lacks flagellar machinery but the twitching motility is encoded. The reconstructed central metabolism revealed pathways enabling the fermentation of organic substrates, as well as their complete oxidation through aerobic and anaerobic respiration. Phylogenetic analysis showed that BY40 bacterium belongs to the uncultured candidate division BRC1. Based on phylogenetic and genomic analyses, the novel bacterium was proposed to be classified as *Candidatus* Sumerlaea chitinivorans, within a candidate phylum Sumerlaeota.

from metagenome of a deep subsurface thermal aquifer in Western Siberia, Russia.

Based on this, we propose the following names for the class, order, and family:

- *Candidatus* Sumerlaeia classis nov.
- *Candidatus* Sumerlaeales ord. nov.
- *Candidatus* Sumerlaeaceae fam. nov.

Description of the novel family *Candidatus* Sumerlaeaceae (Su.mer.lae.a.ce'ae. N.L. fem. n. *Sumerlaea* a Candidatus genus name; *–aceae* ending to denote a family; N.L. fem. pl. n. *Sumerlaeaceae* the Sumerlaea family). The family *Candidatus* Sumerlaeaceae is defined on a phylogenetic basis by comparative 16S rRNA sequence analysis of *Candidatus* Sumerlaea chitinivorans BY40 and uncultured representatives detected in various environments. The type genus is *Candidatus* Sumerlaea.

Description of the novel order *Candidatus* Sumerlaeales (Su.mer.lae.a'les. N.L. fem. n. *Sumerlaea* a Candidatus genus name; *–ales* ending to denote an order; N.L. fem. pl. n. *Sumerlaeales* the Sumerlaea order). The description is the same as that for the family *Candidatus* Sumerlaeaceae. The type genus is *Candidatus* Sumerlaea.

Description of the novel class *Candidatus* Sumerlaeia (Su.mer.lae'i.a. N.L. fem. n. *Sumerlaea* a Candidatus genus name; *–ia* ending to denote a class; N.L. pl. n. *Sumerlaeia* the Sumerlaea

class). The description is the same as that for the family *Candidatus* Sumerlaeaceae. The type order is *Candidatus* Sumerlaeales.

Assuming the results of phylogenetic analysis, and because of the appearance of a representative with known complete genome sequence, the candidate phylum BRC1 is proposed to be named as *Candidatus* Sumerlaeota phyl. nov., following recommendations suggested in [42].

Description of the novel phylum *Candidatus* Sumerlaeota (Su.mer.lae.o'ta. N.L. fem. n. *Sumerlaea* a Candidatus genus name; *–ota* ending to denote a phylum; N.L. pl. n. *Sumerlaeota* the Sumerlaea phylum). The phylum *Candidatus* Sumerlaeota is defined on a phylogenetic basis by comparative 16S rRNA and genome sequence analysis of *Candidatus* Sumerlaea chitinivorans BY40, Candidate division BRC1 bacterium HGW-BRC1-1, BRC1 bacterium SCGC AAA252-M09, BRC1 bacterium SCGC AAA257-C1, candidate division BRC1 bacterium SM23_51, and multiple uncultured representatives, inclusive of the formerly identified candidate division BRC1. The type class is *Candidatus* Sumerlaeia.

Declarations of interest

None.

Acknowledgments

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 14-14-01016).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.syapm.2018.08.013.

References

[1] Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., Nielsen, P.H. (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotechnol. 31 (6), 533–538.
[2] Alneberg, J., Bjarnason, B.S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., Quince, C. (2014) Binning metagenomic contigs by coverage and composition. Nat. Methods 11 (11), 1144–1146.
[3] Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas, B.C., Singh, A., Wilkins, M.J., Karaoz, U., Brodie, E.L., Williams, K.H., Hubbard, S.S., Banfield, J.F. (2016) Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat. Commun. 7, 13219.
[4] Baker, B.J., Lazar, C.S., Teske, A.P., Dick, G.J. (2015) Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. Microbiome 3 (1), 14.
[5] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19 (5), 455–477.
[6] 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 Parabel', Tomsk oblast', western Siberia, Russian Federation.NGU Report, 2014.034, Geological Survey of Norway. Trondheim, 2014.
[7] Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., Olson, R., Overbeek, B., Parrello, G.D., Pusch, M., Shukla, J.A., Thomason, R., III, Stevens, V., Vonstein, R., Wattam, A.R., 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.
[8] Brown, C.T., Olm, M.R., Thomas, B.C., Banfield, J.F. (2016) Measurement of bacterial replication rates in microbial communities. Nat. Biotechnol. 34, 1256–1263.
[9] Derakshani, M., Lukow, T., Liesack, W. (2001) Novel bacterial lineages at the (sub) division level as detected by signature nucleotide-targeted recovery of 16S rRNA genes from bulk soil and rice roots of flooded rice microcosms. Appl. Environ. Microbiol. 67 (2), 623–631.
[10] Einsle, O. (2011) Structure and function of formate-dependent cytochrome c nitrite reductase, NrfA. Methods Enzymol. 496, 399–422.
[11] Frank, Y.A., Kadnikov, V.V., Gavrilov, S.N., Banks, D., Gerasimchuk, A.L., Podosokorskaya, O.A., Merkel, A.Y., Chernyh, N.A., Mardanov, A.V., Ravin, N.V., Karnachuk, O.V., Bonch-Osmolovskaya, E.A. (2016) Stable and variable parts of microbial community in siberian deep subsurface thermal aquifer system revealed in a long-term monitoring study. Front. Microbiol. 7, 2101.
[12] Frischkorn, K.R., 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.
[13] Greening, C., Biswas, A., Carere, C.R., Jackson, C.J., Taylor, M.C., Stott, M.B., Cook, G.M., Morales, S.E. (2016) Genomic and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised energy source for microbial growth and survival. ISME J. 10 (3), 761–777.
[14] Guindon, S., Dufayard, J.F., 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 (3), 307–321.
[15] Hernsdorf, A.W., Amano, Y., Miyakawa, K., Ise, K., Suzuki, Y., Anantharaman, K., Probst, A., Burstein, D., Thomas, B.C., Banfield, J.F. (2017) Potential for microbial H_2 and metal transformations associated with novel bacteria and archaea in deep terrestrial subsurface sediments. ISME J. 11 (8), 1915–1929.
[16] Hunt, D.E., Gevers, D., Vahora, N.M., Polz, M.F. (2008) Conservation of the chitin utilization pathway in the Vibrionaceae. Appl. Environ. Microbiol. 74, 44–51.
[17] Hutcheson, S.W., Zhang, H., Suvorov, M. (2011) Carbohydrase systems of Saccharophagus degradans degrading marine complex polysaccharides. Marine Drugs 9 (4), 645–665.
[18] Jun, H.-S., Qi, M., Gong, J., Egbosimba, E.E., Forsberg, C.W. (2007) Outer membrane proteins of Fibrobacter succinogenes with potential roles in adhesion to cellulose and in cellulose digestion. J. Bacteriol. 189, 6806–6815.
[19] Kadnikov, V.V., Mardanov, A.V., Beletsky, A.V., Banks, D., Pimenov, N.V., Frank, Y.A., Karnachuk, O.V., Ravin, N.V. (2018) A metagenomic window into the 2-km-deep terrestrial subsurface aquifer revealed multiple pathways of organic matter decomposition. FEMS Microbiol. Ecol. 94, fiy152.
[20] Kadnikov, V.V., Frank, Y.A., Mardanov, A.V., Beletsky, A.V., Ivasenko, D.A., Pimenov, N.V., Karnachuk, O.V., Ravin, N.V. (2017) Uncultured bacteria and methanogenic archaea predominate in the microbial community of Western Siberian deep subsurface aquifer. Microbiology (Moscow) 86 (3), 412–415, http://dx.doi.org/10.1134/S0026261717030079.
[21] Kieft, T.L. (2016) Microbiology of the deep continental biosphere. In: Hurst, C.J. (Ed.), Their World: A Diversity of Microbial Environments; Advances in Environmental Microbiology, Cham: Springer International Publishing, pp. 225–249.
[22] Konstantinidis, K.T., Rosselló-Móra, R., Amann, R. (2017) Uncultivated microbes in need of their own taxonomy. ISME J. 11 (11), 2399–2406.
[23] Langmead, B., Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. Nat. Methods 9 (4), 357–359.
[24] Li, X., Roseman, S. (2004) The chitinolytic cascade in Vibrios is regulated by chitin oligosaccharides and a two-component chitin catabolic sensor/kinase. Proc. Natl. Acad. Sci. U. S.A. 101 (2), 627–631.
[25] Magnabosco, C., Ryan, K., Lau, M.C., Kuloyo, O., Lollar, B.S., Kieft, T.L., van Heerden, E., Onstott, T.C. (2016) A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. ISME J. 10 (3), 730–741.
[26] 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 (1), 33–40.
[27] Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17 (1), 10–12.
[28] Nobu, M.K., Dodsworth, J.A., Murugapiran, S.K., Rinke, C., Gies, E.A., Webster, G., Schwientek, P., Kille, P., Parkes, R.J., Sass, H., Jørgensen, B.B., Weightman, A.J., Liu, W.T., Hallam, S.J., Tsiamis, G., Woyke, T., Hedlund, B.P. (2016) Phylogeny and physiology of candidate phylum '*Atribacteria*'(OP9/JS1) inferred from cultivation-independent genomics. ISME J. 10 (2), 273–286.
[29] Park, J.K., Keyhani, N.O., Roseman, S. (2000) Chitin catabolism in the marine bacterium Vibrio furnissii Identification, molecular cloning, and characterization of a N, N′-diacetylchitobiose phosphorylase. J. Biol. Chem. 275, 33077–33083.
[30] Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25 (7), 1043–1055.
[31] Probst, A.J., Castelle, C.J., Singh, A., Brown, C.T., Anantharaman, K., Sharon, I., Hug, L.A., Burstein, D., Emerson, J.B., Thomas, B.C., Banfield, J.F. (2016) Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO_2 concentrations. Environ. Microbiol. 19, 459–474.
[32] Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41 (Database issue), D590–D596.
[33] Refojo, P.N., Sousa, F.L., Teixeira, M., Pereira, M.M. (2010) The alternative complex III: a different architecture using known building modules. Biochim. Biophys. Acta 1797 (12), 1869–1876.
[34] Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., Dodsworth, J.A., Hedlund, B.P., Tsiamis, G., Sievert, S.M., Liu, W.T., Eisen, J.A., Hallam, S.J., Kyrpides, N.C., Stepanauskas, R., Rubin, E.M., Hugenholtz, P., Woyke, T. (2013) Insights into the

phylogeny and coding potential of microbial dark matter. Nature 499 (7459), 431–437.
[35] Robbins, S.J., Evans, P.N., Parks, D.H., Golding, S.D., Tyson, G.W. (2016) Genome-centric analysis of microbial populations enriched by hydraulic fracture fluid additives in a coal bed methane production well. Front. Microbiol. 7, 731.
[36] Rodriguez-R, L.M., Konstantinidis, K.T. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. Peer J. Preprints 4, http://dx.doi.org/10.7287/peerj.preprints.1900v1, e1900v1.
[37] Scigelova, M., Crout, D.H.G. (1999) Microbial beta-N-acetylhexosaminidases and their biotechnological applications. Enzyme Microb. Technol. 25, 3–14.
[38] Sorokin, D.Y., Gumerov, V.M., Rakitin, A.L., Beletsky, A.V., Sinninghe Damsté, J.S., Mardanov, A.V., Ravin, N.V. (2014) Genome analysis of Chitinivibrio alkaliphilus gen. nov., sp. nov., a novel extremely haloalkaliphilic anaerobic chitinolytic bacterium from the candidate phylum Termite Group 3. Environ. Microbiol. 16 (6), 1549–1565.
[39] Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30 (12), 2725–2729.
[40] Wick, R.R., Schultz, M.B., Zobel, J., Holt, K.E. (2015) Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31 (20), 3350–3352.
[41] Wilson, K. (2001) Preparation of genomic DNA from bacteria. Curr. Protoc. Mol. Biol., http://dx.doi.org/10.1002/0471142727.mb0204s56, Chapter 2: Unit 2.4.
[42] Whitman, W.B., Oren, A., Chuvochina, M., da Costa, M.S., Garrity, G.M., Rainey, F.A., Rossello-Mora, R., Schink, B., Sutcliffe, I., Trujillo, M.E., Ventura, S. (2018) Proposal of the suffix *–ota* to denote phyla: Addendum to 'Proposal to include the rank of phylum in the International Code of Nomenclature of Prokaryotes'. Int. J. Syst. Evol. Microbiol. 68, 967–969.
[43] Yanyushin, M.F., del Rosario, M.C., Brune, D.C., Blankenship, R.E. (2005) New class of bacterial membrane oxidoreductases. Biochemistry 44 (30), 10037–10045.