Systematic and Applied Microbiology 42 (2019) 85–93

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

Candidatus Krumholzibacterium zodletonense gen. nov., sp nov, the first representative of the candidate phylum Krumholzibacteriota phyl. nov. recovered from an anoxic sulfidic spring using genome resolved metagenomics

Noha H. Youssef[a,*], Ibrahim F. Farag[a], C. Ryan Hahn[a], Hasitha Premathilake[b], Emily Fry[a], Matthew Hart[a], Krystal Huffaker[a], Edward Bird[a], Jimmre Hambright[a], Wouter D. Hoff[a], Mostafa S. Elshahed[a]

[a] Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK 74074, United States
[b] Department of Animal and Food Sciences, Oklahoma State University, Stillwater, OK 74074, United States

ARTICLE INFO

Keywords:
Candidatus Krumholzibacterium
zodletonense
Genomic analysis
Uncultured bacteria

ABSTRACT

The accumulation of genomes of uncultured organisms has highlighted the need for devising a taxonomic and nomenclature scheme to validate names and prevent redundancies. We here report on the recovery and analysis of four phylogenetically related genomes recovered from an anoxic sulfide and sulfur-rich spring (Zodletone spring) in southwestern Oklahoma. Phylogenetic analysis based on 120 single copy markers attested to their position as a novel distinct bacterial phylum. Genomic analysis suggests Gram-negative flagellated organisms that possess type IV pili. The organisms are predicted to be rod-shaped, slow-growers, with an anoxic, heterotrophic, and fermentative lifestyle. Predicted substrate utilization pattern includes multiple amino acids, dipeptides, tripeptides, and oligpopeptides; as well as few sugars. Predicted auxotrophies include proline, vitamin B6, lipoic acid, biotin, and vitamin B12. Assessment of the putative global distribution pattern of this novel lineage suggests its preference to anoxic marine, terrestrial, hydrocarbon-impacted, and freshwater habitats. We propose the candidatus name *Krumholzibacterium zodletonense* gen. nov, sp. nov. for Zgenome0171^T, with the genome serving as the type material for the novel family Krumholzibacteriaceae fam. nov., order Krumholzibacteriales ord. nov., class Krumholzibacteria class nov., and phylum Krumholzibacteriota phyl. nov. The type material genome assembly is deposited in GenBank under accession number QTKG01000000.

Introduction

The last decade witnessed rapid experimental and computational advances that enabled direct recovery of microbial genomes from a wide range of environments [5,33,34,39]. A fraction of the recovered genomes using such approaches often belongs to yet uncultured microbial lineages. Such genomes often belong to lineages previously encountered in amplicon-based 16S rRNA gene diversity surveys [3,44,46], a considerable fraction often represents completely novel lineages [5,40]. As such, the global phylum level diversity, especially within the domain Bacteria, appears to be much broader than previously implied using both culture-based and amplicon-based diversity surveys [45].

The rapid accumulation of genomes directly recovered from environmental samples has also triggered a timely debate on: 1. The most appropriate procedures for data and metadata deposition and availability [4], a crucial issue to enable detailed assessment of reported phylogenetic and metabolic attributes by the broader community and preventing duplicate reporting (e.g. [13,42]), 2. Minimal quality standards required for reporting novel single cell genomes (SAGs) and Metagenome assembled genomes (MAGs) [4], 3. Approaches for circumscribing boundaries for various taxonomic ranks based on genome sequence data [5,18,19,27] to ensure compatibility in taxonomic rank assignments between cultured and uncultured organisms, 4. Required and recommended analyses for describing an uncultured lineage based on genomic data [18,36], and 5. A framework for SAGs and MAGs nomenclature procedures [18,25]. While a broad consensus has been building around some issues, e.g. quality thresholds and data/metadata deposition, others have been an area of active discussions, e.g. taxon boundaries, and

* Corresponding author.
E-mail address: noha@Okstate.edu (N.H. Youssef).

https://doi.org/10.1016/j.syapm.2018.11.002
0723-2020/

laws and involvement of International Committee of Systematics of Prokaryotes (ICSP) in naming uncultured organisms [25].

We here describe the phylogenetic position, putative structural features and metabolic capabilities of multiple genomic assemblies recovered from Zodletone spring, a sulfide and sulfur-rich spring in southwestern Oklahoma. We propose the name Candidatus Krumholzibacterium zodletonense gen. nov, sp. nov. to accommodate Zgenome017^T, the type material (Tax ID number 2291675, Biosample accession number SAMN09745294, GenBank WGS accession number QTKG01000000). This also serves as the type material for the novel candidate family Krumholzibacteriaceae fam. nov., order Krumholzibacteriales ord. nov., class Krumholzibacteria class nov., and phylum Krumholzibacteriota phyl. nov.

Materials and methods

Sampling, DNA extraction and sequencing

Sediment samples were obtained from the source sediments of Zodletone spring, an anaerobic sulfide and sulfur-rich spring in the Anadarko Basin of Western Oklahoma (N34.99562° W98.68895°). The spring geochemistry has been described in detail in prior publications [6,38]. Samples were obtained using sterile spatulas and deposited into sterile 50 ml polypropylene plastic tubes. Samples were transferred to the laboratory on ice, where they were immediately processed. DNA was extracted from the samples using the DNeasy PowerSoil kit (Qiagen, Valencia, CA, USA). Sequencing was conducted using the services of a commercial provider (Novogene, Beijing, China) using two lanes of the Illumina HiSeq 2500 system. A total of 281 Gbp of raw data were obtained from the source sediment. This single metagenome was used for subsequent assembly and binning as explained below.

Assembly, binning, and quality control

Reads were assembled and binned into MAGs (metagenome-assembled genomes) using a pipeline that combines MegaHit [22] for assembling reads into contigs, MaxBin [43] for binning contigs into genomic assemblies, and CheckM [26] for estimation of genome completeness, strain heterogeneity, and contamination. Genomic assemblies showing contamination levels higher than 5%, and/or strain heterogeneity more than 10% were further refined based on the taxonomic affiliations of the binned contigs, as well as the GC content, tetranucleotide frequency, and coverage levels using RefineM [21]. Low quality bins (>5% contamination) were cleaned by removal of the identified outlier contigs, and the % completeness and contamination were again re-checked using CheckM to ensure that the final genomic assemblies analyzed were of high quality. Overall, we obtained 87 high-quality draft, 196 medium-quality draft, and 42 low-quality draft genomes from the source (according to the nomenclature system proposed in Ref. [4]).

Results of initial taxonomic classification of genomes demonstrated that Zodletone spring source sediments are highly diverse, harboring multiple novel high rank (phylum, class, and order) taxonomic diversities. Collectively, representatives of 46 bacterial and 8 archaeal phyla were identified, 32 of which belong to uncultured bacterial and archaeal phyla. The overall diversity of the sample will be the subject of forthcoming research publications. This manuscript describes 4 MAGs that were obtained as part of the above effort.

Phylogenetic analysis

The overall taxonomic scheme was based on the Genome Taxonomy Database using GTDB-Tk (https://github.com/Ecogenomics/GtdbTk). Phylogenetic placement was conducted using a concatenated alignment of 120 single-copy phylogenetic marker genes [27] obtained using the software GTDB-Tk. Concatenated alignments were used to construct maximum likelihood trees in FastTree [29]. The monophyly of branches was tested using various tree-building algorithms and varying the number and identity of taxa included. Average amino acid identity (AAI) was estimated based on the AAI calculator (http://enve-omics.ce.gatech.edu/). Shared gene content (SGC) was calculated using CompareM (https://github.com/dparks1134/CompareM). To assign an appropriate taxonomic rank for the target lineage, we utilized conservative taxonomic cutoffs for AAI indices (<46% for phylum, >46% for class), shared gene content (<24% for phylum, >24% for class), and AAI score (<70% for phylum, >70% for class) (these cutoffs were the most conservative deduced from the figures in the supplementary document for Ref. [27]).

Phylogeny based on 16S rRNA gene was conducted by querying the nt NCBI database (July 2018) using 16S rRNA gene recovered in assembled bins. A conservative threshold of 90% nucleotide identity and 80% subject alignment length was employed to identify potential hits. Metgenomic bins 16S rRNA gene sequences were aligned to those from potential hits as well as references from closely related phyla using the Silva SINA aligner [30]. Alignment was used to construct a maximum likelihood phylogenetic tree using FastTree [29].

Genomic analysis and metabolic reconstruction

iRep [14] was used to predict the replication rate. iRep calculates the ratio of sequencing coverage at the origin compared to the terminus of replication to measure replication rates. Since iRep calculates average coverage values using a sliding window of 5 Kbp, it does not require sequencing coverage of Ori and Ter sites, which makes it ideal for use with less than complete genomic assemblies. Several genomic features and COG category distribution were used to predict the organisms putative growth preferences and lifestyle as described before [47]. Transfer RNAs were predicted using tRNAscan-SE [23]. Psort [48] was used for prediction of subcellular protein localization. Structural features, e.g. cell wall structure, presence or absence of flagella, pili, fimbriae, cell division and expected cell shape etc. were examined by querying for relevant genes.

Detailed metabolic reconstruction of relevant pathways was performed using both KEGG [15] and Metacyc [7] databases. Proteases, peptidases, and protease inhibitors were identified using Blastp against the Merops database [31]. Transporters were identified using the transporter classification database (TCDB) [37].

Ecological distribution

Relative abundance of the lineages in Zodletone spring sediments was evaluated as the number of reads belonging to this lineage as a percentage of total reads in the 281.0 Gbp raw sequence data as estimated using Bowtie2 [20].

To investigate the ecological distribution and occurrence of the target lineage, we queried the nt NCBI database (July 2018) using the gene for 16S rRNA as described above. The environments from which significant hit sequences were obtained were assigned into one of six major habitats (marine, freshwater, terrestrial, bioremediation, host-associated, and engineered) following the IMG/M database habitat classification scheme [24]. The marine, freshwater, and terrestrial habitats were further subclassified into 15 different sub-habitats [24], and the global distribution of the target lineage in various habitats was assessed.

Table 1
General genomic features of the four genomes assembled from Zodletone spring sediment as well as closely related genomes from Suncor tailing pond [13] and Guyamas basin sediment [40].

	Zgenome0171	Zgenome0218	Zgenome0207_2	Zgenome0274_1	UBA4771	UBA4783	UBA2172	NATK01
Binning source	Zodletone sediment				Suncor tailing pond			Guaymas Basin sediment
Binned size, Mb	2.69	2.79	1.17	1.18	2.18	2.48	2.1	1.53
% completeness	90.1	70.33	46.2	46.2	75.01	89.91	73.29	68.7
% contamination	4.46	2.2	0	0	3.4	1.2	0	2.2
Estimated genome size, Mb	2.99	3.98	2.54	2.56	2.91	2.51	2.87	2.23
% coding bases	92	91.12	90.27	90.22	85.6	91.7	82.7	90.2
% GC content	49.18	50.88	55.52	55.5	66.8	66.6	66.7	50.9
Average gene length, bp	1036	1118	1029	1027	969	1060	891	920
Total number of genes	2443	2343	1042	1053	2177	2200	2282	1523
Total number of tRNA	51	41	13	13	22	34	33	22
Total number of protein-coding genes	2392	2302	1029	1040	2155	2166	2249	1505
Number of protein-coding genes with COG categories	1511	1368	598	602	1501	1613	1469	1173
WGS accession	QTKG01	QTKF01	QTKE01	QTKD01	DHHN01	DHHB01	DCWC01	NATK01
Digital protologue number	CA00045	CA00046	CA00047	CA00048	NA*	NA*	NA*	NA*

*NA: not available.

Comparative genomics to other closely related phyla

PCA was conducted to identify differences in genomic features and COG distributions between Zodletone lineages and representatives from closely related phyla. Genome size, number of genes, GC content, coding density, COG categories distribution (percentage of proteins belonging to each COG category), protein sub-localization (percentage of proteins destined to the cytoplasm, cytoplasmic membrane, periplasm, outer membrane, and extracellular milieu) were used to construct the PCA using the prcomp function in the labdsv package of R [35]. A biplot was constructed using the biplot function in R, where genomes are represented as points and variables are represented as arrows pointing in the direction of maximal abundance.

Additional genomes

Closely related genomes were identified through comparison of 16 ribosomal protein sequences (large subunit ribosomal proteins L2, L3, L4, L5, L6, L14, L15, L16, L18, L22, L24, and small subunit ribosomal proteins S3, S8, S10, S17, and S19) [12] to the NCBI WGS database using tblastn with 70% sequence identity cutoff.

Accession numbers

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the Bioproject accession number PRJNA483775, Biosample accession numbers SAMN09745294, SAMN09745295, SAMN09745296, and SAMN09745297, and genome accession numbers QTKG00000000, QTKF00000000, QTKE00000000, and QTKD00000000, for Zgenome0171^T, Zgenome0218, Zgenome0207-2, and Zgenome0274-1, respectively. The version described in this paper is QTKG01000000, QTKF01000000, QTKE01000000, and QTKD01000000.

Results and discussion

Genome assemblies from Zodletone spring

Four genomes (Zgenome0171^T, Zgenome0218, Zgenome0274-1, and Zgenome0207-2) were recovered from Zodletone spring sediments (Table 1). Two-way Average Nucleotide Identity (ANI) calculation showed a value of 75.68% for Zgenome0171^T, and Zgenome0218 and a value of 100% for Zgenome0274-1, and Zgenome0207-2 but was lower than the detection level when comparing Zgenome0171^T, and Zgenome0218 to Zgenome0274-1, and Zgenome0207-2 using the default alignment options of 700 bp minimum alignment length, 50 minimum alignments, and 70% minimum identity (using 1000 bp window size and 200 bp step size). Accordingly, we opted for AAI to describe the similarity between the four MAGs. The four genomes had a high level of amino acid identity (AAI) (78.19 ± 21.15%) (Table 2). Zgenome0171^T, Zgenome0218 exhibited AAI of 72.5%, while Zgenome0274-1, and Zgenome0207-2 were more similar with an AAI of 100. However, Zgenome0171^T, and Zgenome0218 on one hand, and Zgenome0274-1, and Zgenome0207-2 on the other exhibited lower AAI of 57.83%, suggesting that, while all 4 MAGs belong to the same genus, they might represent members of two species. We stress here that, due to the low completion levels of Zgenome0274-1, and Zgenome0207-2, this is just a speculation. Obtaining more genomes belonging to the same genus would certainly improve this assessment. One high-quality (90.10% completion, 4.46% contamination) assembly (Zgenome0171^T) encompassing a small subunit rRNA gene sequence was chosen as the type material. These four genomes were obtained as part of a larger effort to assemble genomes of yet-uncultured bacterial and archaeal phyla from Zodletone Spring source sediments. A brief description of the diversity of the spring at the time of sampling is provided in the methods section.

Phylogenetic analysis

Concatenated protein trees using 120 single copy marker proteins consistently placed strains Zgenome0171^T, Zgenome0218, Zgenome0274-1, and Zgenome 0207-2 as a distinct monophyletic lineage. The closest relatives of Zodletone genomes were three genomic assemblies constructed by Parks et al. [27] from publicly available metagenomes from Suncor tailing pond (UBA2172 (GCA_002327475.1), UBA4771 (GCA_002403295.1), and UBA4873 (GCA_002403075.1)); and one genomic assembly, ex4484_7, from Guyamas basin sediments [10] (GCA_002085285.1) (Fig. 1a). The three genomes described by Parks et al. [27] were previously designated as representatives of the UBP1 phylum for which no Latin name was given, presumably since they did not meet the proposed quality threshold proposed by the same research group for Latin names designation [9]. This Zodletone-UBP1-ex4484_7 clade exhibited a distant relationship with candidate phyla Zixibacteria [8], TA06 [2], Eisenbacteria [1], Edwardsbacteria (UBP2) [1], UBP14 [27], and "Latescibacteria" [34] (Fig. 1a). We used multiple rank discrimination criteria (AAI, shared gene content) (Tables 2–3), to assess the putative taxonomic rank of Zodletone genomes in rela-

Table 2
Amino acid identities (AAI) and shared gene content (SGC) of Krumholzibacteriota genomes compared in this study. Values were calculated based on the total number of proteins using the AAI calculator at (http://enve-omics.ce.gatech.edu/).

	Zgenome0171		Zgenome0218		Zgenome0274_1		Zgenome0207_2		NATK01		UBA2172		UBA4771		UBA4783	
	AAI	SGC	AAI	SGC	AAI	SGC	AAI	SGC	AAI	SGC	AAI	SGC	AAI	SGC	AAI	SGC
Zgenome0171	100	100														
Zgenome0218	72.5	34.4	100	100												
Zgenome0274_1	57.3	31.1	58.3	38.05	100	100										
Zgenome0207_2	57.3	31.2	58.4	38.2	100	100	100	100								
NATK01	54.6	25.9	53.4	25.3	50.5	27.5	50.4	27.6	100	100						
UBA2172	54.7	26.9	54.1	27.9	54.2	25.1	54.3	25.05	56.9	26	100	100				
UBA4771	53.8	27.9	53.2	28.4	53	25.3	53.1	25.1	55.4	26.4	98.7	38	100	100		
UBA4783	54	31.2	53.7	31.2	54	25.8	54.1	25.7	56.3	28.3	85.5	37.3	85.3	38.6	100	100

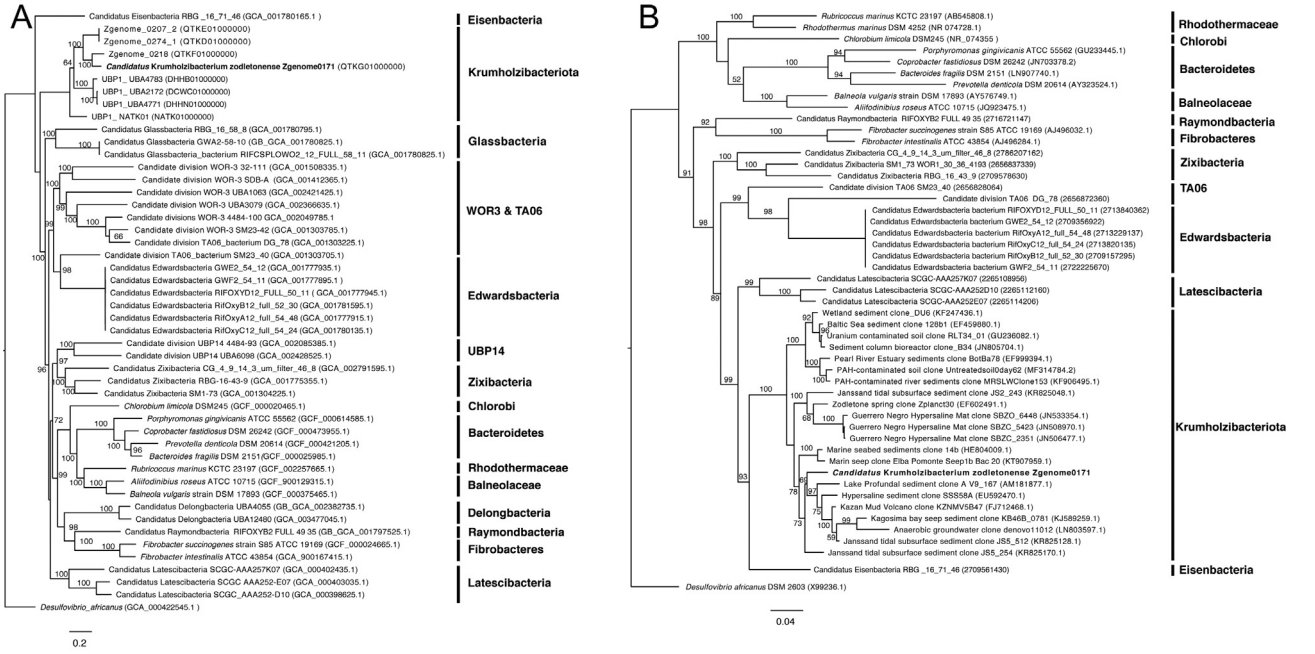

Fig. 1. Maximum likelihood phylogenetic trees based on the concatenated protein alignment of 120 single-copy markers (A) as well as 16S rRNA gene (B) highlighting the phylogenetic position of Krumholzibacteriota genomes. Reference taxa in both trees are either type strains of cultured microorganisms, genomes of relevant uncultured bacterial phyla recovered using single cell genomics or genome resolved metagenomics, 16S rRNA amplicons recovered in culture independent 16S rRNA gene diversity surveys (only in B). Some taxa are represented in concatenated protein tree but not in 16S rRNA gene tree because of the unavailability of 16S rRNA gene in their publicly available genome assemblies (e.g. seven of the eight Krumholzibacteriota genomes, as well as members of the candidate phyla Glassbacteria, Delongbacteria, WOR3, and UBP14). The concatenated alignment used to construct the protein tree in A was generated using the GTdb-Tk, while the 16S rRNA tree in B was generated using SILVA-aligned sequences. Both trees were obtained using FastTree. Bootstrap values (from 100 replicates) are shown for nodes with more than 50 bootstrap support.

tion to the 3 UBP1 genomes [27] and the single Guyamas basin sediment genome ex4484_7 [10]. Our results suggest that Zodletone genomes, the three UBP1 Suncor tailing pond genomes, and the Guyamas basin sediment genome represent three distinct lineages within this novel phylum (Table 2). However, due to the incompleteness of genomes, a definitive phylogenetic rank could not be confidently assigned to describe the relationship between these three lineages (orders versus classes within the novel phylum). We believe that obtaining more complete genomes in the future should facilitate this.

Prevalence in Zodletone spring, and global ecological distribution patterns

A small fraction of raw reads from Zodletone spring dataset mapped to Zgenome0171^{T} scaffolds (3.1M/971M reads, 0.32%), indicating that this lineage represented a minor fraction of the community at the time of sampling. Members also appear to be slow growers, with a replication index of 1.23, indicating that less than one quarter of the cells belonging to this lineage were actively replicating at the time of sampling [14].

Table 3
Averages ± standard deviation of AAI, and shared gene content (SGC) of the phyla compared. Cells are shaded in different colors to highlight AAI above 46 and SGC above 24 (red shading, denoting within phylum differences), and AAI below 46 and SGC below 24 (green shading, denoting across phyla differences).

	Eisenbacteria		Edwardsbacteria		Krumholzbacterota		WOR3		TA06	
	AAI	SGC	AAI	SGC	AAI	SGC	AAI	SGC	AAI	SGC
Eisenbacteria	79.3±18.4	41.6±9.2								
Edwardsbacteria	40.7±0.6	19±1	99.4±1	45.6±7.5						
Krumholzbacterota	40.9±13	17.9±2.4	40.3±0.96	19.5±2.8	69.4±20.7	34±10.5				
WOR3	37.8±0.8	16.9±1.5	40.4±0.7	21.4±2.2	38.5±0.9	18.9±2	73.7±19.6	34.2±12		
TA06	40.6±2.5	18.3±1.1	43.1±0.6	21.3±0.6	41.6±0.9	18.7±3.7	42.4±0.9	20.7±1.9	99.4±0.7	44.3±6.4

We utilized the 16S rRNA gene sequence recovered from Zgenome0171^T assembly to identify prior detection, if any, of this novel lineage in amplicon-based studies. We identified 76 sequences in GenBank nt and SILVA database that are >90% similar to Zgenome0171^T 16S rRNA gene. Indeed, these sequences were monophyletic to Zgenome0171^T sequence in 16S rRNA gene-based tree (Fig. 1b). It is worth noting that several of the amplicon sequences associated with Zgenome071^T are currently labeled as belonging to the candidate phylum Latescibacteria in the current version of the SILVA database (Release 132). However, as the curators of the SILVA database acknowledge, the "Latescibacteria" sequences in SILVA are an umbrella term of possibly as much as 25 phyla based on the high intersequence diversity within this group and lack of genome sequence data within this part of the tree of life. The proposed novel phylum hence represents an additional anchoring taxon through which 16S rRNA gene sequences could be assigned in the future. Remarkably, Zgenome0171^T 16S rRNA gene exhibited 93.6% similarity to a PCR amplicon from Zodletone spring sediment previously obtained using Planctomycetes-biased primers in a prior study in 2007 [11] (GenBank accession number EF602491.1). This strongly indicates the endemic nature of members of this lineage in Zodletone spring. Sequences affiliated with this lineage (n = 76) belonged to 36 distinct marine, terrestrial, hydrocarbon-impacted, or freshwater habitats (Table S1). While seemingly disparate the majority of these habitats appear to be anoxic or hypoxic, suggesting a preference from anaerobiosis. Such preference is indeed seconded by the locations from which the UBP1 genomes were identified (an oil sand tailing pond) and the ex4484_7 genome (Guyamas Basin sediments), as well as metabolic reconstruction analysis described below.

General genomic features, and structural features

The four recovered genomes possess a relatively high GC content (52.8 ± 3.24%), are relatively medium sized (estimates of 3.01 ± 0.67 Mb), with a few crispers (n = 0–1), low percentage of non-coding regions (9.1 ± 0.84%), and ~1053 ± 44 bp average gene length (Table 1). In comparison, the three genomes from Suncor tailing pond had a significantly higher GC content (66.7 ± 0.1) (Student t-test p-value = 0.0008).

Genomic analysis of all genomes indicates a Gram-negative cell wall, based on the identification of genes encoding LPS biosynthesis and the absence of evidence for peptidoglycan pentaglycine bridge synthesis (Table 4). As well, genes encoding type IV pili production were also identified in all 8 genomes (Table 4). Interestingly, flagellar assembly genes were identified in the four Zodletone genomes and in the ex4484_7 genome from Guyamas basin sediment but were not identified in genomes from Suncor tailing pond (UBP1 genomes) (Table 3). The identification of the rod-shape determining proteins MreBCD and RodA [28], and the absence of peptidoglycan carboxypeptidase, previously linked to maintaining a helical shape, strongly suggest straight rod-shaped organism (Table 4). No homologues of Pfam03319 or Pfam00936 were detected, suggesting the absence of bacterial microcompartment (BMC), previously identified in other phyla within the FCB, e.g. Latescibacteria [46]. Similarly, evidences for encapsulin nanocompartment (Pfam04454) [41] were absent, as were evidences for magnetosome biogenesis (absence of homologues for mamAB-like operon or any of the mad genes [17]).

Physiological preferences, metabolic abilities and predicted substrate utilization pattern

Metabolic analysis of all genomes suggested a heterotrophic lifestyle, with no phototrophic or carbon dioxide fixation mechanisms identified. The organisms appeared to be anaerobic as evident by the lack of respiratory cytochrome C oxidase (complex IV) components, the absence of the oxidative branch of the PPP, and the identification of several oxidative stress enzymes (Table 4). Zodletone genomes encoded rubrerythrin, and rubredoxin, both known to be implicated in oxidative stress protection in strict anaerobes as they do not produce O_2 during their catalytic cycle [32] (Table 4), as well as catalase and peroxidase. UBP1 genomes assembled from the Suncor tailing pond encoded superoxide dismutase Fe/Mn family enzyme, and all genomes encode alkylhydroperoxide reductase. Other defense mechanisms include the CRISPR/Cas system identified only in the Suncor tailing pond UBP1 genomes, and type I restriction endonucleases identified only in Zodletone genomes and the ex4484_7 genome from Guyamas basin sediment.

Genomic analysis of the biosynthetic capabilities showed that Zodletone genomes encoded for de novo biosynthesis of at least 12 amino acids, but were apparently auxotrophic (missing all enzymes in the biosynthetic pathway) to proline. Biosynthetic machineries for the other seven amino acid were partially encoded in Zodletone genomes. In comparison, Suncor tailing pond genomes and Guyamas basin genome appeared to lack key enzymes in several of these biosynthetic pathways making their organisms apparently auxotrophic for 9–10 amino acids (Table 4). Analysis of Zodletone genomes also suggests fewer cofactor auxotrophies in comparison to genomes from Suncor tailing pond and Guyamas basin. This observed level of interdependency is common in multiple anaerobic slow growing organisms in eutrophic habitats [16], where large amounts of nutrients and precursors from lysed cells and biomass are conducive to uptake rather than de novo synthesis of energetically expensive cofactors. All genomes encode a complete gluconeogenic pathway with minor differences (Table 4). These include two different routes of phosphofructokinase reversal in Suncor tailing pond genomes (the genomes encode both a bifunctional fructose 1,6-bisphosphate aldolase/phosphatase as well as its monofunctional counterparts), as opposed to just the bifunctional enzyme in Guyamas basin genome and just the monofunctional enzymes in Zodletone genomes. Similarly, Zodletone genomes and Suncor tailing pond genomes encode two routes for pyruvate kinase reversal (via pyruvate phosphate dikinase, and the sequential action of pyruvate carboxylase and PEP carboxykinase).

Catabolically, a complete glycolytic pathway was identified, but no pyruvate dehydrogenase complex for pyruvate oxidative decarboxylation to acetyl CoA was identified in any of the genomes. Instead, a pyruvate:ferredoxin oxidoreductase as well as pyruvate-formate lyase (and its activating enzyme) are presumably utilized for pyruvate to acetate conversion. An incomplete TCA cycle, putatively used for biosynthetic purposes rather for acetyl-CoA oxidation to CO_2, was identified. Multiple enzymes were identified for pyruvate reduction to fermentation end products for electron disposal. Fermentation end products included acetate, formate, and ethanol (not in the Guyamas basin genome) (Table 4). Interestingly, while all genomes encode a partial NADH dehydrogenase complex I (4–5 subunits encoded), and a complete succinate dehydrogenase complex II, only the Suncor tailing pond genomes encode what appears to act as a terminal oxidase complex. These genomes encode *nrfAH,* cytochrome C nitrite reductase (NH3-forming) [EC 1.7.2.2], which suggests the possibility of anaerobic nitrite reduction to ammonia.

Putative substrate utilization patterns were deduced by the identification of enzymes channeling various substrates to central metabolic intermediates, as well as the identification of putative transporters for such substrates. Substrates theoretically supporting growth included the sugars glucose and mannose; several amino acids through their conversion to pyruvate directly or indirectly via one of the TCA cycle intermediates. Putative hydrolysis of dipeptides, oligopeptides, and complex proteins via extracellular

Table 4
Features deduced from genomic analysis of Krumholzibacteriota genomes assembled from Zodletone spring sediment (Zod), Suncor tailing pond (UBP1), and Guyamas basin (NATK)[a].

	Zod	UBP1	NATK
Structural features			
Cell wall			
LPS biosynthesis	√	√	√
Peptidoglycan with DAP and no pentaglycine bridge	√	√	√
Cell membrane glycerophospholipid			
Phosphatidyl glycerol	√	√	√
Cardiolipin	√	x	x
Flagellar motility	√	x	√
Type IV pilus assembly	√	√	√
Chemotaxis	√	√	√
Cell shape			
Rod-shape determining RodA/mreBCD	√	√	√
Defense mechanisms			
CRISPR Cas system	x	√	x
Type I restriction endonucleases	√	x	√
Oxidative stress			
Superoxide dismutase Fe/Mn family	x	√	x
Catalase	√	x	x
Peroxidase	√	x	x
Rubrerythrin	√	x	x
Rubredoxin	√	x	x
Alkylhydroperoxide reductase	√	√	√
Biosynthesis			
Gluconeogenesis			
Bifunctional fructose 1,6-bisphosphate aldolase/phosphatase [EC:4.1.2.13 3.1.3.11]	x	√	√
Monofunctional fructose 1,6-bisphosphate aldolase class II	√	√	x
Monofunctional fructose 1,6-bisphosphatase class I	√	√	x
Monofunctional fructose 1,6-bisphosphatase class II	√	√	x
Reversal of pyruvate kinase			
Via pyruvate phosphate dikinase (EC 2.7.9.1)	√	√	√
Via pyruvate carboxylase and PEP carboxykinase (ATP) [EC:4.1.1.49]	√	√	x
Amino acids			
Asp from oxaloacetate	√	√	√
Asn from Asp	√	√	√
Glu from alpha-ketoglutarate	√	√	√
Gln from Glu	√	√	√
Cys from Ser	√	x	x
Ser from Gly	√	√	√
Thr from Gly	√	√	√
Gly from Ser	√	√	√
Met from Cys	√	x	x
Lys (Diaminopimelate intermediates)	√	x	x
Arg from Glu	√	x	x
Val	Partial	Partial	Partial
Leu	Partial	Partial	Partial
Ile	Partial	Partial	Partial
Tyr	Partial	x	x
Phe	Partial	x	x
Trp	Partial	x	x
Ala	√	x	√
His	Partial	x	x
Pro	x	x	x
Cofactor biosynthesis			
Thiamine	√	x	x
Riboflavin	√	√	x
Pyridoxal-phosphate	Partial	Partial	x
NAD/NADP	√	√	Partial
Pantothenate from valine and aspartate	√	√	Partial
Coenzyme-A	√	√	√
Acyl-carrier protein	√	√	√
Biotin biosynthesis from pimelate	x	x	x
Biotin import then ligation to enzymes	√	√	√
Lipoic acid biosynthesis from octanoyl-ACP	x	x	x
Lipoic acid salvage	√	x	x
Folate biosynthesis from GTP	√	√	Partial
Molybdenum cofactor from GTP	√	x	x
Heme biosynthesis from Glu	Partial	x	x
Vitamin B12	x	x	x
MEP/DOXP pathway for terpenoid backbone biosynthesis	√	Partial	Partial
Menaquinone biosynthesis from terpenoids	√	√	Partial

Table 4 (*Continued*)

	Zod	UBP1	NATK
Catabolism			
Sugar catabolism			
Glucose	√	√	√
Mannose	√	√	√
Galactose	x	√	√
Amino acid catabolism			
Ala	√	√	√
Asp	√	√	√
Asn	√	x	x
Glu	√	√	√
Gln	√	x	x
His	√	√	
Met	√	√	√
Cys	√	x	x
Ser	√	x	√
Thr	√	√	√
Gly	√	√	x
Products of metabolism			
Fermentation			
Acetate production from pyruvate (EC 6.2.1.1)	√	√	√
Ethanol production from acetyl-CoA	√	√	x
Formate production from pyruvate			
Via pyruvate formate lyase (EC 2.3.1.54)	√	√	√
Via pyruvate:fereedoxin oxidoreductase (EC 1.2.7.1) (EC 1.2.7.11)	√	√	√
Respiration (anaerobic)			
NADH dehydrogenase (complex I)	Partial	Partial	Partial
Succinate dehydrogenase (complex II)	√	√	√
Nitrite reductase (cytochrome; ammonia-forming) (EC 1.7.2.2)	x	√	x

[a] Information in this table is based on genomic analysis of incomplete genomes and care should be taken on interpreting the results on auxotrophies or partial presence of certain pathways as these could be due to the incompleteness of the genomes. However, a √ denotes that a full pathway or a complete set of genes were identified in the genomes.

peptidases coupled to the import of the resulting amino acids and short peptides suggest that the organism could depend on peptides and amino acids rather than sugars for energy production. However, we could not rule out that these enzymes could be utilized for satisfying auxotrophies rather than energy production. Comparatively, Zodletone genomes encode a more elaborate amino-acid degrading machinery, while Suncor tailing pond genomes and the Guyamas basin genome encode galactose degradation (Table 4).

Comparative genomics to other closely related phyla

Comparative genomics using general genomic features placed Zodletone genomes (red hexagons in Fig. 2) close to Suncor tailing pond genomes (black hexagons in Fig. 2) and Guyamas basin genome (blue hexagon in Fig. 2) and away from representatives of Eisenbacteria, Edwardsbacteria, Latescibacteria, TA06, and WOR-3 (phyla closely related to them in the protein tree, Fig. 1b), further justifying their placement in a new phylum. Their position in the PCA biplot was mainly due to their relatively higher GC content and the enrichment in COGs belonging to cell motility, signal transduction, intracellular trafficking and secretion, transcription, and amino acid metabolism categories, which fits well with their flagellar motility, and predicted dependence mainly on amino acids metabolism for energy production.

We propose the creation of a new bacterial phylum to accommodate the assemblies Zgenome0171^T, Zgenome0218, Zgenome0274-1, and Zgenome 0207-2 obtained from Zodletone spring, as well as the three UBP1 genomes from Suncor tailing pond and ex4484_7 genome from Guyamas basin sediment. The designation is justified by its unique phylogenetic position in the bacterial tree of life based on concatenated sequences. For the type material (Zgenome0171^T), we propose the genus name Krumholzibacterium to honor the contribution of Professor Lee Krumholz to the fields of diversity, genetics, physiology of microorganisms in anaerobic habitats, including Zodletone spring, and the species name *zodletonense* to reflect the sampling site from which the genomes were recovered. The genome assembly Zgenome0171^T serves as the type material and is deposited in GenBank under Biosample accession number SAMN09745294 and WGS accession number QTKG00000000.

Description of candidatus *Krumholzibacterium gen. nov.*

Etymology. Krumholzibacterium. Krumholz: named after Professor Lee Krumholz to recognize his contribution to the fields of diversity, genetics, and physiology of anaerobes in multiple habitats, including Zodletone spring.

Genomic analysis predicts an anaerobic, Gram-negative, motile (flagellated) chemoorganoheterotrophic rod. Genomic analysis identified auxotrophies for proline, vitamin B6, lipoic acid, biotin, and vitamin B12. Substrates supporting growth based on genomic analysis include glucose, mannose, aspartate, alanine, asparagine, glutamate, glutamine, cysteine, serine, threonine, glycine, methionine, histidine, dipeptides, tripeptides, and oligopeptides. Fermentation end products include acetate, formate, and ethanol. The genome assembly Zgenome0171^T serves as the type material and is deposited in GenBank under Biosample accession number SAMN09745294 and WGS accession number QTKG00000000.

Description of candidatus *Krumholzibacterium zodletonense sp. nov.*

Etymology. Krumholzibacterium. Krumholz: named after Professor Lee Krumholz to recognize his contribution to the fields of diversity, genetics, and physiology of anaerobes in multiple habitats, including Zodletone spring. zodletonense: pertaining to the

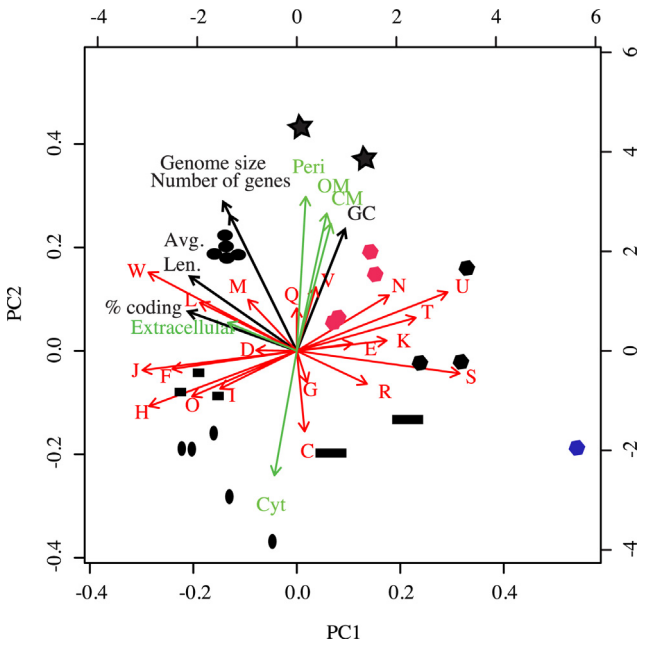

Fig. 2. Principal component analysis biplot constructed using several general genomic features (black arrows; include genome size in Mb (Genome size), total number of protein-coding genes (Number of genes), average gene length (Avg. Len.), percentage of coding region (%coding), and the GC content (GC)), sub-localization of proteins (green arrows; Periplasmic (Peri), Outer membrane (OM), cytoplasmic membrane (CM), Extracellular, and cytoplasmic (Cyt)), and COG category classification (red arrows; showing one-letter COG abbreviation). The first two axes explained ~90% of variance. Genomes clustered by their respective phyla as follows: Eisenbacteria (stars), Edwardsbacteria (circles), candidate phylum TA06 (ovals), candidate phylum WOR-3 (squares), Latescibacteria (rectangles), and Krumholzibacteriota genomes from Zodletone spring (red hexagons) and previously recognized UBP1 genomes from Suncor tailing pond (black hexagons), and Guyamas basin (blue hexagon). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sampling site (Zodletone spring in southwestern Oklahoma, USA) from which the type species genome was recovered.

Exhibits the following properties in addition to those given in the genus description. Appears to be a slow-growing member of the rare biosphere. Possesses catalase and peroxidase-encoding genes. Lacks CRISPR-Cas system. The genome assembly Zgenome0171^T serves as the type material and is deposited in GenBank under Biosample accession number SAMN09745294 and WGS accession number QTKG00000000.

Description of candidatus *Krumholzibacteriaceae fam. nov.*

Krumholzibacteriaceae (Krum.holz.bac.te.ri.a.ce′ae. N.L. masc. n.). Krumholzibacterium type genus of the family; L. suff. – aceae ending to donate a family; N.L. fem. pl. n. Krumholzibacteriaceae the family of the genus Krumholzibacterium.

The description is the same as for the genus Krumholzibacterium. The genome assembly Zgenome0171^T serves as the type material and is deposited in GenBank under Biosample accession number SAMN09745294 and WGS accession number QTKG00000000.

Description of candidatus *Krumholzibacteriales ord. nov.*

Krumholzibacteriales (Krum.holz.bac.te.ri.a′les. N.L. fem. n. Krumholzibacterium type genus of the order; – ales ending to donate an order; N.L. fem. pl. n. Krumholzibacteriales the order of the genus Krumholzibacterium).

The description is the same as for the genus Krumholzibacterium. Contains the family Krumholzibacteriaceae. The genome assembly Zgenome0171^T serves as the type material and is deposited in GenBank under Biosample accession number SAMN09745294 and WGS accession number QTKG00000000.

Type genus: Krumholzibacterium.

Description of candidatus *Krumholzibacteria classis nov.*

Krumholzibacteria (Krum.holz.bac.te′ri.a. N.L. fem. n. Krumholzibacterium type genus of the type order of the class; suff. – ia ending to donate a class; N.L. neut. pl. n. Krumholzibacteria the class of the order Krumholzibacteriales).

The description is the same as for the genus Krumholzibacterium. Contains the order Krumholzibacteriales. The genome assembly Zgenome0171^T serves as the type material and is deposited in GenBank under Biosample accession number SAMN09745294 and WGS accession number QTKG00000000.

Type order: Krumholzibacteriales.

Description of candidatus *Krumholzibacteriota phyl. nov.*

Krumholzibacteriota (Krum.holz.bac.te ro′ta. N.L. fem. pl. n. Krumholzibacterium type genus of the type class of the phylum; L. suff. – ota ending to denote phylum; N.L. neut. pl. n. Krumkolzbacteriota the phylum of the class Krumholzibacteria).

The *candidatus* phylum Krumholzibacteriota is defined by eight genomic assemblies (Zgenome0171^T, Zgenome0218, Zgenome0207_2, Zgenome0274_1, UBA2172, UBA4771, UBA4783, and ex4484_7) recovered using genome-resolved metagenomics from three distinct habitats (Zodletone spring sediment, Suncor tailing pond, and Guyamas basin), as well as 16S rRNA gene sequences from uncultured representatives in a wide range of terrestrial, marine, and aquatic habitats that are predominantly anoxic or hypoxic. The genome assembly Zgenome0171^T serves as the type material and is deposited in GenBank under Biosample accession number SAMN09745294 and WGS accession number QTKG00000000. Genomic analysis predicted Gram-negative cell walls, and rod-shaped cells with heterotrophic lifestyle, and the ability to degrade glucose, mannose, alanine, aspartate, glutamate, and methionine and production of acetate and formate as fermentation end products.

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.2018.11.002.

References

[1] 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.

[2] 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, 14.

[3] Becraft, E.D., Woyke, T., Jarett, J., Ivanova, N., Godoy-Vitorino, F., Poulton, N., Brown, J.M., Brown, J., Lau, M.C.Y., Onstott, T., Eisen, J.A., Moser, D., Stepanauskas, R. (2017) Rokubacteria: genomic giants among theuncultured bacterial phyla. Front. Micorobiol. 8, 2264.

[4] Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A., Tringe, S.G., Ivanova, N.N., Copeland, A., Clum, A., Becraft, E.D., Malmstrom, R.R., Birren, B., Podar, M., Bork, P., Weinstock, G.M., Garrity, G.M., Dodsworth, J.A., Yooseph, S., Sutton, G., Glöckner, F.O., Gilbert, J.A., Nelson, W.C., Hallam, S.J., Jungbluth, S.P., Ettema, T.J.G., Tighe, S., Konstantinidis, K.T., Liu, W.-T., Baker, B.J., Rattei, T., Eisen, J.A., Hedlund, B., McMahon, K.D., Fierer, N., Knight, R., Finn, R., Cochrane,

G., Karsch-Mizrachi, I., Tyson, G.W., Rinke, C., The Genome Standards Consortium, Lapidus, A., Meyer, F., Yilmaz, P., Parks, D.H., Murat Eren, A., Schriml, L., Banfield, J.F., Hugenholtz, P., Woyke, T. (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat. Biotechnol. 35, 725.

[5] Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A., Wilkins, M.J., Wrighton, K.C., Williams, K.H., Banfield, J.F. (2015) Unusual biology across a group comprising more than 15% of domain Bacteria. Nature 523, 208–211.

[6] 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.

[7] Buhring, S.I., Sievert, S.M., Jonkers, H.M., Ertefai, T., Elshahed, M.S., Krumholz, L.R., Hinrichs, K.-U. (2011) Insights into chemotaxonomic composition and carbon cycling of phototrophic communities in an artesian sulfur-rich spring (Zodletone, Oklahoma, USA), a possible analog for ancient microbial mat systems. Geobiology 9, 166–179.

[8] Caspi, R., Billington, R., Fulcher, C.A., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Midford, P.E., Ong, Q., Ong, W.K., Paley, S., Subhraveti, P., Karp, P.D. (2018) The MetaCyc database of metabolic pathways and enzymes. Nucleic Acids Res. 46, D633–D639.

[9] Castelle, C.J., Hug, L.A., Wrighton, K.C., Thomas, B.C., Williams, K.H., Wu, D., Tringe, S.G., Singer, S.W., Eisen, J.A., Banfield, J.F. (2013) Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. Nat. Commun. 4, 2120.

[10] Chuvochina, M., Rinke, C., Parks, D.H., Rappé, M.S., Tyson, G.W., Yilmaz, P., Whitman, W.B., Hugenholtz, P. (2019) The importance of designating type material for uncultured taxa. Syst. Appl. Microbiol. 42 (1), 15–21 https://doi.org/10.1016/j.syapm.2018.07.003.

[11] Dombrowski, N., Seitz, K.W., Teske, A.P., Baker, B.J. (2017) Genomic insights into potential interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments. Microbiome 5, 106.

[12] Elshahed, M.S., Youssef, N.H., Luo, Q., Najar, F.Z., Roe, B.A., Sisk, T.M., Bühring, S.I., Hinrichs, K.-U., Krumholz, L.R. (2007) Phylogenetic and metabolic diversity of planctomycetes from anaerobic, sulfide- and sulfur-rich Zodletone Spring, Oklahoma. App. Environ. Microbiol. 73, 4707–4716.

[13] Hug, L.A., Castelle, C.J., Wrighton, K.C., Thomas, B.C., Sharon, I., Frischkorn, K.R., Williams, K.H., Tringe, S.G., Banfield, J.F. (2013) Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. Microbiome 1, 22.

[14] Hug, L.A., Thomas, B.C., Sharon, I., Brown, C.T., Sharma, R., Hettich, R.L., Wilkins, M.J., Williams, K.H., Singh, A., Banfield, J.F. (2015) Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. Environ. Microbiol. 18, 159–173.

[15] Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 42, D199–D205.

[16] Kantor, R.S., Wrighton, K.C., Handley, K.M., Sharon, I., Hug, L.A., Castelle, C.J., Thomas, B.C., Banfield, J.F. (2013) Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. mBio, 4.

[17] Kolinko, S., Richter, M., Glöckner, F.-O., Brachmann, A., Schüler, D. (2015) Single-cell genomics of uncultivated deep-branching magnetotactic bacteria reveals a conserved set of magnetosome genes. Environ. Microbiol. 18, 21–37.

[18] Konstantinidis, K.T., Rossello-Mora, R., Amann, R. (2017) Uncultivated microbes in need of their own taxonomy. ISME J. 11, 2399–2406.

[19] Konstantinidis, K.T., Tiedje, J.M. (2005) Towards a genome-based taxonomy for prokaryotes. J. Bacteriol. 187, 6258–6264.

[20] Langmead, B., Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357.

[21] Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25.

[22] Li, D., Liu, C.M., Luo, R., Sadakane, K., Lam, T.W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674–1676.

[23] Lowe, T.M., Chan, P.P. (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 44, W54–W57.

[24] Markowitz, V.M., Chen, M.A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Pagani, I., Tringe, S., Huntemann, M., Billis, K., Varghese, N., Tennessen, K., Mavromatis, K., Pati, A., Ivanova, N.N., Kyrpides, N.C. (2013) IMG/M 4 version of the integrated metagenome comparative analysis system. Nucleic Acids Res. 42, D568–D573.

[25] Oren, A. (2017) A plea for linguistic accuracy – also for Candidatus taxa. Int. J. Syst. Evol. Microbiol. 67, 1085–1094.

[26] 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, 1043–1055.

[27] Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.A., Woodcroft, B.J., Evans, P.N., Hugenholtz, P., Tyson, G.W. (2017) Recovery of nearly 8000 metagenome-assembled genomes substantially expands the tree of life. Nat. Microbiol. 2, 1533–1542.

[28] Pichoff, S., Lutkenhaus, J. (2007) Overview of cell shape: cytoskeletons shape cells. Curr. Opin. Microbiol. 10, 601–605.

[29] Price, M.N., Dehal, P.S., Arkin, A.P. (2010) FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490.

[30] Pruesse, E., Peplies, J., Glöckner, F.O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28, 1823–1829.

[31] Rawlings, N.D., Waller, M., Barrett, A.J., Bateman, A. (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 42, D503–D509.

[32] Riebe, O., Fischer, R.J., Wampler, D.A., Kurtz, D.M., Jr., Bahl, H. (2009) Pathway for H_2O_2 and O_2 detoxification in Clostridium acetobutylicum. Microbiology 155, 16–24.

[33] Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., Dmitrieff, E., Malmstrom, R., Stepanauskas, R., Woyke, T. (2014) Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. Nat. Protocols 9, 1038–1048.

[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, 431–437.

[35] Roberts, D.W., Package 'labdsv': Ordination and Multivariate Analysis for Ecology, in, 2016.

[36] Rosselló-Móra, R., Whitman, W.B. (2019) Dialogue on the nomenclature and classification of prokaryotes. Syst. Appl. Microbiol. 42 (1), 5–14.

[37] Saier, M.H., Jr., Reddy, V.S., Tsu, B.V., Ahmed, M.S., Li, C., Moreno-Hagelsieb, G. (2016) The Transporter Classification Database (TCDB): recent advances. Nucleic Acids Res. 44, D372–D379.

[38] Senko, J.M., Campbell, B.S., Henricksen, J.R., Elshahed, M.S., Dewers, T.A., Krumholz, L.R. (2004) Barite deposition mediated by photootrophic sulfide-oxidizing bacteria. Geochim. Cosmochim. Acta 68, 773–780.

[39] Sharon, I., Banfield, J.F. (2013) Genomes from metagenomics. Science 342, 1057–1058.

[40] Siegl, A., Kamke, J., Hochmuth, T., Piel, J., Richter, M., Liang, C., Dandekar, T., Hentschel, U. (2011) Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. ISME J. 5, 61–70.

[41] Sutter, M., Boehringer, D., Gutmann, S., Gunther, S., Prangishvili, D., Loessner, M.J., Stetter, K.O., Weber-Ban, E., Ban, N. (2008) Structural basis of enzyme encapsulation into a bacterial nanocompartment. Nat. Struct. Mol. Biol. 15, 939–947.

[42] Wang, Z., Guo, F., Liu, L., Zhang, T. (2014) Evidence of carbon fixation pathway in a bacterium from candidate phylum SBR1093 revealed with genomic analysis. PLoS One 9, e109571.

[43] Wu, Y.W., Simmons, B.A., Singer, S.W. (2016) MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32, 605–607.

[44] Youssef, N.H., Blainey, P.C., Quake, S.R., Elshahed, M.S. (2011) Partial genome assembly for a candidate division OP11 single cell from an anoxic spring (Zodletone Spring, Oklahoma). Appl. Environ. Microbiol. 77, 7804–7814.

[45] Youssef, N.H., Couger, M.B., McCully, A.L., Guerrero-Criado, A.E., Elshahed, M.S. (2015) Assessing the global phylum level diversity within the bacterial domain: a review. J. Adv. Res. 3, 269–282.

[46] Youssef, N.H., Farag, I.F., Rinke, C., Hallam, S.J., Woyke, T., Elshahed, M.S. (2015) In silico analysis of the metabolic potential and niche specialization of candidate phylum "Latescibacteria" (WS3). PLoS One 10, e0127499.

[47] Youssef, N.H., Rinke, C., Stepanauskas, R., Farag, I., Woyke, T., Elshahed, M.S. (2015) Insights into the metabolism, lifestyle and putative evolutionary history of the novel archaeal phylum 'Diapherotrites'. ISME J. 9, 447–460.

[48] Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J., Brinkman, F.S. (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26, 1608–1615.