Genome biology of a novel lineage of planctomycetes widespread in anoxic aquatic environments

Stefan Spring 1, Boyke Bunk 2, Cathrin Spröer 3, Manfred Rohde 4 and Hans-Peter Klern 1

1 Department Microorganisms, Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany.
2 Department Bioinformatics, Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany.
3 Department Central Services, Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany.
4 Central Facility for Microscopy, Helmholtz Centre for Infection Research, Braunschweig, Germany.

Introduction

The phylum Planctomycetes comprises ubiquitous bacteria with a usually spherical cell shape that have an ecological relevance in many soil, freshwater and marine environments (Fuerst and Sagulenko, 2011; Boedeker et al., 2017). Taxonomically this phylum is currently divided into the two classes Planctomycetes and Phycisphaerae. Most of the available and characterized strains of the Planctomycetes are aerobic or facultatively anaerobic chemoheterotrophs assigned to the order Planctomycetales within the class Planctomycetia, while only a few isolates could be affiliated with the recently proposed orders Phycisphaerales (Fukunaga et al., 2009; Yoon et al., 2014) and Tepidisphaerales (Kovaleva et al., 2015) both belonging to the Phycisphaerae class. The only strictly anaerobic planctomycetes known to date are fastidious chemotrophic bacteria specialized on the anaerobic oxidation of ammonium to dinitrogen, a process designated as anammox (Jetten et al., 2001). The known anammox planctomycetes are not yet available as pure cultures and form a monophyletic group that has been tentatively classified at the order level into the ‘Candidatus Brocadiales’ within the class Planctomycetia. However, this lineage rather corresponds to a separate class or even phylum according to current phylogenetic standards proposed by Yarza et al. (2014).

In recent years, numerous cultivation independent molecular studies indicated a significant abundance of planctomycetes in permanently anoxic aquatic environments that do not belong to the anammox clade, but can be affiliated with the Phycisphaerae class (e.g., Eshahed et al., 2007; Harrison et al., 2009; Baker et al., 2015; Robbins et al., 2016). Among these studies a report dealing with Caspian Sea sediments (Mahmoudi et al., 2015) is of special interest, because it documents not only a high abundance of Phycisphaerae sequences in suboxic sediments (up to 11% of total), but also an increase of their frequency with depth of the anoxic sediment. Therefore, this finding implies a high incidence of bacteria with a
strictly anaerobic phenotype among representatives of this phylogenetic group. In the course of a study analysing the microbial diversity in a lithifying cyanobacterial mat it was possible to confirm this assumption by obtaining the first strictly anaerobic isolate of the Phycisphaerae class (Spring et al., 2015). By using selective enrichment methods four additional anaerobic strains belonging to this clade could be retrieved from hypersaline sediments of solar saltlains in Portugal and Spain. In the present study, the novel isolates and phylogenetically related metagenome-assembled genomes (MAGs) were analysed. Based on the obtained results it was possible to gain detailed insights into the metabolic specialization and evolution of this abundant and widespread group of anaerobic planctomycetes.

Results and discussion

Phylogeny and biogeography of anaerobic bacteria affiliated with the Phycisphaerae class

In total five different anaerobic strains could be obtained from various geographic locations. Strains SM-PulAB-D3 and SM-Chi-D1 were retrieved from anoxic sediments of the saline Es Trenc in Mallorca, while strains ST-PulAB-D4 and ST-NAGAB-D1 were isolated from sediments of an evaporation pond in Tavira, Portugal. In addition, the strain L21-RPul-D3 has been isolated previously from a cyanobacterial mat of a hypersaline lake on the Kiritimati Atoll. A phylogenetic tree based on a comprehensive set of environmental 16S rRNA gene sequences shows the placement of the novel anaerobic isolates within a distinct lineage of the Phycisphaerae class that is only distantly related to the order Phycisphaerales and so far devoid of any cultured representatives (Fig. 1). Based on the classification used in the SILVA 128 SSU Ref NR99 database the novel isolates were placed within the tentatively named clades ML-A-10 (ST-PulAB-D4, SM-PulAB-D3, L21-RPul-D3 and SM-Chi-D1) and 0622D293 (ST-NAGAB-D1), both comprising mainly sequences of uncultivated bacteria retrieved from hypersaline sediments and mats. In order to gain deeper insights into the evolution and metabolic plasticity of this group of the Phycisphaerae class genome sequences of the retrieved anaerobic strains were determined and additional reference genomes were obtained from NCBI (GenBank) and DOE JGI (IMG/M) databases (Table 1). At the time of writing complete genome sequences were available of the type strain of Phycisphaerae mikurensis and the novel isolates ST-PulAB-D4, L21-RPul-D3, SM-Chi-D1 and ST-NAGAB-D1, while the genome of strain SM-PulAB-D3 remains in the draft state. All other genomes represent MAGs varying in their completeness from 49.3% to 100%. Several MAGs that encode 16S rRNA genes of sufficient length formed a coherent group with the novel isolates and were included in the phylogenetic tree shown in Fig. 1. The origins and phylogenetic affiliations of these reconstructed genome sequences were as follows: The MAGs GW2C_45_44 and GWF2_41_51 were obtained from groundwater samples of an aquifer system (Anantharaman et al., 2016) and assigned to the S15A-MN16 clade; bin0 was retrieved from a coal bed methane production well (Phycisphaerae-PK28, Robbins et al., 2016) and affiliated with the KCLumb-38–53 clade; Planc_01 originated from an anaerobic digester supplied with cellulose (VanWonterghem et al., 2016) and was placed within the MSBL9 clade. On the other hand, the MAGs EBPR_Bin_263 from sewage and SM23_33 from estuary sediment were only distantly related to this coherent group and assigned to the family Phycisphaeraceae and an unclassified order-level lineage within the Phycisphaerae class respectively (data not shown).

In addition, RNA polymerase beta subunit (RpoB) protein sequences were analysed, because several MAGs lacked 16S rRNA genes, but encoded almost complete RpoB sequences, which could be used as an alternative phylogenetic marker for the reconstruction of evolutionary relationships (Mollet et al., 1997; Bondoso et al., 2013). The placement of the analysed RpoB protein sequences within the Phycisphaerae class is illustrated in Supporting Information Fig. S1. Several MAG sequences retrieved from anoxic estuary sediments (SG8_4, SM1_79; Baker et al., 2015) and subsurface aquifer sediments (RBG_13_62_9, RBG_13_46_10, RBG_13_50_24; Anantharaman et al., 2016) could be assigned along with Planc_01 to the MSBL9 clade, while the RpoB sequence of SM23_30 was only distantly related to the novel isolates and represented a separate order-level lineage of the Phycisphaerae class.

Sequences of clades comprising the novel isolates and most MAGs as well as the clades MSB-3A7 and SHA-43 formed a stable coherent group supported by high bootstrap values above 95% in trees based on 16S rRNA gene sequences. In order to elucidate a possible classification of this lineage and related phylogenetic groups in higher taxonomic ranks identity values of 16S rRNA gene sequence comprising major branches within the Phycisphaerae class were determined (Supporting Information Table S1). The minimum and median sequence identity values of the lineage represented by the novel anaerobic isolates were 75.89% and 89.90% respectively, and therefore close to the recommended threshold values for the definition of an order in the taxonomy of bacteria and archaea (Yarza et al., 2014). The corresponding values for the remaining major lineages within the Phycisphaerae as delineated in Fig. 1 were in a comparable range and ranged from 70.98% to 80.93% for minimum and 84.90% to 90% for median sequence identity values. Consequently, it seems justified to establish a novel order to accommodate the first
Fig. 1. Phylogenetic tree of the *Phycisphaerae* class deduced from 16S rRNA gene sequences showing the placement of the newly isolated anaerobic strains within a novel order-level lineage. The tree topology was reconstructed under the maximum-likelihood criterion and rooted using the 16S rRNA gene sequence of *Candidatus Kuenenia stuttgartiensis* (AMCG01002665, not shown). Polygons represent clades of several sequences. Sequences extracted from metagenome-assembled genomes are indicated by the abbreviation MAG. Accession numbers are given in parentheses (IMG numbers refer to the respective ID of the 16S rRNA gene in the IMG/M database). Support of a distinct branching by bootstrap analyses is indicated by symbols. Black dots at a distinct node indicate that bootstrap values of 95% or above (percentages of 1000 resamplings) were obtained with three different reconstruction methods, while grey dots indicate that values of 95% or above were obtained with only two reconstruction methods. White dots indicate that bootstrap values of 75% or above were obtained with at least one reconstruction method. In such cases the values of 75% or above are given from left to right for the maximum-likelihood, neighbor-joining and maximum parsimony method. Sequences assigned to the proposed order *Sedimentisphaerales* are highlighted with different colours that should indicate the type of habitat in which a sequence or isolate was detected. Green indicates a preference for groundwater, freshwater sediments or sewage, blue points to marine or brackish environments and red to hypersaline sediments or microbial mats. Scale bar, 0.10 changes per nucleotide position.
cultured strict anaerobes of the *Phycisphaerae* class. Based on the source of isolation of most strains we propose the name *Sedimentisphaerales*. It should be noted that a survey of the environmental distribution of 16S rRNA gene sequences in most other order-level groups of the class *Phycisphaerae* revealed a general preference for marine sediments and hypersaline mats as well, with the notable exception of the *Tepidisphaerales*, which are most common in soil (Supporting Information Table S1). Nevertheless, some differences in the habitat preference of representatives of major lineages were revealed in a more detailed analysis presented in Fig. 2. It appears that members of the order *Phycisphaerales* are more versatile in the adaptation to ecological niches found in wastewater, soil and animals compared to the *Sedimentisphaerales*, which occupy mostly marine and hypersaline sediments.

The ecological relevance of members of the *Sedimentisphaerales* has been determined for several populations represented by MAGs and isolates. Strain ST-PulAB-D4<sup>T</sup> was isolated from the fourth step of a decimal dilution series starting with an inoculum of around 0.1 g anoxic mud, so that its prevalence in the examined hypersaline sediment can be estimated to be at least 10<sup>5</sup> active cells per cm<sup>3</sup>. Furthermore, bacteria represented by the MAG PlanC_01 reached a relative abundance of 1.79% in an anaerobic digester under certain conditions (Vanwonterghem et al., 2016).

### Table 1. General characteristics of genomes used in the phylogenetic analyses of the class *Phycisphaerae*.

| Designation | Accession number<sup>a</sup> | Size (Mb) | Completeness (%) | DNA G + C (mol%) |
|-------------|-------------------------------|-----------|------------------|-----------------|
| ST-PulAB-D4<sup>T</sup> | CP021023 | 3.19 | 100.0 | 46.31 |
| SM-PulAB-D3 | NHPQ00000000 | 3.06 | 98.9/100.0<sup>b</sup> | 46.42 |
| L21-RPul-D3<sup>T</sup> | CP019633 | 2.95 | 100.0 | 46.20 |
| SM-Chi-D1 | CP019646 | 3.86 | 100.0 | 47.04 |
| ST-NAGAB-D1 | CP019646 | 4.25 | 100.0 | 51.98 |
| MAG bin0<sup>c</sup> | 2593339136 | 2.91 | 95.5 | 55.23 |
| MAG SM1_79 | LJVF00000000 | 2.95 | 49.3 | 50.29 |
| MAG RBG_13_50_24 | MHYE00000000 | 2.34 | 86.0 | 49.68 |
| MAG SG8_4 | LJTR00000000 | 4.46 | 71.9 | 54.85 |
| MAG RBG_13_46_10 | MHYD00000000 | 1.93 | 74.4 | 43.78 |
| MAG PlanC_01 | LVBB00000000 | 5.65 | 100.0 | 62.78 |
| MAG RBG_13_62_9 | MHYG00000000 | 4.17 | 81.4 | 59.82 |
| MAG GWF2_41_51 | MHXZ00000000 | 4.83 | 90.7 | 41.04 |
| MAG GWC2_45_44 | MHXU00000000 | 2.93 | 81.4 | 44.90 |
| MAG SM23_30 | LJUC00000000 | 4.01 | 71.2 | 51.01 |
| MAG SM23_33 | LJUF00000000 | 3.43 | 75.3 | 67.83 |
| MAG EBPR_Bin_263 | 2619618906 | 3.65 | 85.7 | 67.43 |
| *Phycisphaera mikurensis* NBRC 102666 | AP012338 | 3.88 | 100.0 | 73.22 |
| MAG RBG_13_62_9 | MHYG00000000 | 4.17 | 81.4 | 59.82 |
| MAG GWF2_41_51 | MHXZ00000000 | 4.83 | 90.7 | 41.04 |
| MAG GWC2_45_44 | MHXU00000000 | 2.93 | 81.4 | 44.90 |
| MAG SM23_30 | LJUC00000000 | 4.01 | 71.2 | 51.01 |
| MAG SM23_33 | LJUF00000000 | 3.43 | 75.3 | 67.83 |
| MAG EBPR_Bin_263 | 2619618906 | 3.65 | 85.7 | 67.43 |
| *Phycisphaera mikurensis* NBRC 102666 | AP012339 | 3.88 | 100.0 | 73.22 |

<sup>a.</sup> All accession numbers are from NCBI GenBank, except 2593339136 and 2619618906, which are from the JGI IMG/M database.

<sup>b.</sup> Genome completeness was estimated using the software CheckM (Parks et al., 2015). The value of 98.9% was identical to the result obtained with the complete genome of strain ST-PulAB-D4<sup>T</sup>.

<sup>c.</sup> The prefix MAG indicates metagenome-assembled genomes.

References of the used genome sequences are as follows: ST-PulAB-D4<sup>T</sup>, SM-PulAB-D3, L21-RPul-D3<sup>T</sup>, SM-Chi-D1, and ST-NAGAB-D1 (this study); GWC2_45_44, GWF2_41_51, RBG_13_62_9, RBG_13_46_10, and RBG_13_50_24 (Anantharaman et al., 2016); SG8_4, SM1_79, SM23_30 and SM23_33 (Baker et al., 2015); PlanC_01 (Vanwonterghem et al., 2016); bin0 (Robbins et al., 2016), *Phycisphaera mikurensis* NBRC 102666<sup>T</sup> and EBPR_Bin_263 (unpublished).

**Fig. 2.** Habitat preference of members of major lineages of the class *Phycisphaerae*. The bar chart illustrates the relative proportions of environmental sources of 16S rRNA gene sequences according to the SILVA 128 SSU Ref NR 99 data set. The number of analysed sequences of a distinct group is given in Supporting Information Table S1.
2016), while bacteria represented by the MAG bin0 became even dominating (9% relative abundance) in a coal bed methane production well after injection of polysaccharide-containing hydraulic fracture fluids (Robbins et al., 2016). Based on these findings it can be concluded that representatives of this lineage can reach a high relative abundance in anoxic sites of aquatic environments, where they could play a significant role in the anaerobic degradation of polymers.

**General features of genomes**

Some general characteristics of genomes of members of the proposed order *Sedimentisphaerales* and other representatives of the *Phycisphaerae* class are summarized in Tables 1 and 2. The size of genomes among members of the novel lineage ranged from 1.93 to 5.65 Mb, the DNA G + C content from 41.04 to 62.78 mol% and the total number of genes from 2003 to 4564. Most genomes of this group encode one or two sets of ribosomal RNA genes, which were not organized in a single *rrn* operon, but had the 16S rRNA gene unlinked from the 23S-5S rRNA genes, which is a typical trait found in members of the *Planctomycetes* (Liesack and Stackebrandt, 1989). Among the novel isolates a small genome size correlated with a low DNA G + C content, which could indicate that genome evolution in this phylogenetic group is partly driven by streamlining (Giovannoni et al., 2014). As most environments inhabited by these bacteria are not characterized by nutrient limitation, the observed moderate streamlining is probably based on specialization caused by competition of several populations for the same environmental niche.

The number of detected transposase genes within genomes can be used as a proxy for genome plasticity (including genetic rearrangements or horizontal gene transfer events) and relative evolution rates in bacteria (Li et al., 2014). As shown in Table 2, the determined average prevalence of transposase genes seems to correlate with the habitat and is significantly higher in strains isolated from hypersaline sediments and mats (2.26% of all genes) than in other members of the novel order-level lineage (0.30%) or the *Phycisphaeraeales* (0.35%) inhabiting marine or freshwater environments. A survey of the complete genomes of the novel strains isolated from hypersaline sediments revealed that a high abundance of transposase genes correlated with a large number of putative genomic islands, which supports the previous assumption that a high load of transposase genes reflects a high plasticity of the genome. The location and arrangement of transposase genes and putative genomic islands within complete genomes of the novel isolates are illustrated in the circular DNA plots shown in Fig. 3.

Integrity and stability of microbial genomes are in most cases protected against foreign DNA by sequence directed defense mechanisms. Arrays of clustered, regularly interspaced short palindromic repeats (CRISPR), which provide functions involved in adaptive immunity against infection with foreign DNA (Marraffini and Sontheimer, 2010), could be detected in most genomes of the *Sedimentisphaerales* (Table 2). The number of CRISPR loci showed however some variation and ranged from 0 to 6, which could reflect a variable burden of viral attacks depending on the occupied ecological niches. Remains of potential prophage regions could be detected in all complete isolate genomes, but the identified sites (1 – 5 per genome) contained only a few phage genes and had a low completeness score (below 70 according to PHASTER). PacBio SMRT sequencing data obtained for the novel isolates from hypersaline environments enabled also the study of active restriction-modification (R-M) systems in this group. The results are available in REBASE (http://
Fig. 3. Genome structures and origins of replication in cultured strains representing the proposed order Sedimentisphaerales. Results of the prediction of oriC regions are shown for the following strains: (A) SM-Chi-D1, (B) ST-PulAB-D4T, (C) ST-NAGAB-D1, (D) L21-RPul-D3T. The first nucleotide of the dnaA_1 gene was used as starting point of the genome sequence. The left panel illustrates predictions of the oriC using the software tool Ori-Finder. For detection of DnaA boxes the sequence motif found in Flavobacteriaceae (tgttccacg) was considered with no more than one mismatch. Based on the determined GC disparity curve the suggested regions of termination and oriC are denoted by red and green arrows respectively. Numbers indicate the position of sites identified at the chromosome in kb. In the middle coverage plots along the genome sequences obtained by PacBio sequencing are shown. The green and red arrows mark the deduced sites of origin and termination of replication respectively. On the right circular plots of the genome sequences are shown. Circles denote from the outside to inside: position of transposase genes (blue), position of genes associated with the canonical oriC (green) and priA (red), position of dnaA genes (red), sites of predicted genomic islands (turquoise), G+C content (grey, above average; black, below average), GC skew (olive green, above average; purple, below average). The purple diamonds and green dots on the G+C graphs indicate the maximum of DnaA box clusters and the tentative origins of replication deduced from the coverage plots respectively.
rebase.neb.com/rebase/rebase.html; Roberts et al., 2015) and indicate a high variability of R-M systems among the cultured isolates. In the closely related strains ST-PulAB-D4T and L21-RPul-D3T nine distinct sequence motifs for methylation could be recognized, which displayed N4-methylcytosine (m4C) and N6-methyladenine (m6A) base modifications. Both strains seem to rely mainly on type II restriction endonucleases and a type IV methylation-dependent restriction system. Strain ST-NAGAB-D1 encodes in addition to type II and type IV systems several type I restriction enzyme complexes and revealed eight distinct methylated sequence motifs. Strain SM-Chi-D1 showed the most extensive methylation pattern with 19 distinct identified sequence motifs of which four were unique for this strain. It encoded multiple type II restriction enzymes, but also R-M systems of type I, III and IV. Although, there are several possible functions of R-M systems in bacteria (Vasu and Nagaraja, 2013) the observed extensive DNA methylation pattern in strain SM-Chi-D1 is likely used as cellular defence against phages and compensates for a lacking CRISPR system, which is functional in all other genome-sequenced isolates from hypersaline sediments and mats.

**Prediction and characterization of DNA replication origins**

An important feature in the organization of bacterial chromosomes is the location of the origin of replication (oriC), which represents the site where the replication machinery is assembled and then precedes bidirectionally around the circular chromosome until both replication forks meet at a termination region at the opposite site. In many bacterial chromosomes an asymmetry in nucleotide composition between the leading and lagging strand of replication can be observed, which is commonly referred to as GC skew. In these cases, the asymmetry changes its polarity at the origin and terminus of replication, with a minimum of the GC disparity curve usually found at the oriC. In addition, the arrangement of particular genes within the chromosome can indicate a certain location of the oriC. The gene of the replication initiator protein DnaA is usually found in close proximity of the oriC and some genes essential for DNA replication and growth, like dnaN (DNA polymerase III subunit beta), gyrB (DNA gyrase subunit B), recF (DNA replication and repair protein), rnap (rRNA operon) and rpmH (50S ribosomal protein L34) are frequently found in the same region (Mackiewicz et al., 2004). Furthermore, the active replication of DNA in growing cells leads to an increase of sequences close to the origin, which stimulates the arrangement of highly expressed genes like ribosomal proteins and rRNA operons near the origin of replication (Rocha, 2004). In order to detect potential oriC sites we used local minima of the GC disparity curves and the presence of DnaA box clusters (short repetitive nonamer sequences binding DnaA) typically found in the oriC region. The in-silico predictions made by the Ori-Finder program could be examined with PacBio SMRT sequencing data that allowed the detection of peaks of the sequence copy number at distinct regions of the genomes. A similar method based on next-generation sequencing data was applied previously to identify putative origins of replication in DNA extracted from growing cells of *Escherichia coli* (Skovgaard et al., 2011; Maduike et al., 2014).

First, we tried to validate this combined approach with two different genome-sequenced reference strains, which were also originating from a hypersaline environment. In Supporting Information Fig. S2A, it is shown that the genome of the spirochete *Salinispira pacifica* L21-RPul-D2T (acc. no. CP006939) appears to have a canonical oriC region. In this strain the cluster of DnaA boxes correlates with the minimum of the cumulative GC skew and in the same region the genes dnaA, dnaN, recF and rpmH could be found. Accordingly, the PacBio coverage plot indicates a maximum within the region of the predicted replication origin and a minimum at the presumed termination site. In the genome of *Salinivirga cyanobacterivorans* L21-Spi-D4T (acc. no. CP013118), a strain representing a novel family within the *Bacteroidetes* phylum, the oriC was also correctly predicted, although the presence of multiple copies of the dnaA gene, of which none was located close to the oriC, represented a deviation of the typical organization (Supporting Information Fig. S2B).

In contrast, in silico predictions of the oriC in genomes of the novel strains belonging to the *Phycisphaerae* class was ambiguous. This is exemplary elucidated with the genome of strain SM-Chi-D1. The minimum of the cumulative GC skew and the location of a cluster of DnaA boxes pointed to a position of the oriC around 3.18 Mb (Fig. 3A). However, at this site only dnaN (SMSP2_02492), but no dnaA gene was detected. Interestingly, the PacBio coverage plot revealed a tentative origin of replication at around 0.69 Mb distance from the predicted oriC close to the location of one copy of the dnaA gene (SMSP2_00001). Within this region the priA gene (SMSP2_00004) was also detected, which is involved in the DnaA/oriC-independent replication of the *E. coli* chromosome at oriK sites. These sites are thought to represent regions prone to the formation of RNA-DNA hybrids (R-loops) that may stimulate initiation of replication independent of the DnaA protein. In *E. coli*, initiation at oriK sites depends on unfavourable growth conditions and therefore could be used to promote adaptation to environmental stress by higher mutation rates (Masai and Arai, 1996; Maduike et al., 2014). Currently, it is not clear, if in strain SM-Chi-D1 initiation of replication at a putative oriK site was induced by suboptimal cultivation conditions or if a non-canonical oriC site
was translocated to a different location by rearrangements of the genome. However, the retained pattern of strand asymmetry indicates that either oriK-dependent replication occurs rarely in nature or that the oriC was only recently translocated, so that there was not enough time to change the nucleotide composition of both DNA strands by replication-associated mutational pressure.

In genomes of the remaining strains recognition of a typical oriC site was even more difficult, which is unveiled by an increasing distance of the minimum of the GC disparity curves from the locations of the detected DnaA box clusters in the genomes of strains ST-PulAB-D4T (Fig. 3B), ST-NAGAB-D1 (Fig. 3C) and L21-RPul-D3T (Fig. 3D). In addition, the local minima of the GC skew plots were either distended (ST-PulAB-D4T) or occurred at several sites of the chromosome (ST-NAGAB-D1 and L21-RPul-D3T), which reflect a weak or fragmentary nucleotide asymmetry in the cumulative GC skew diagrams shown in the circular DNA plots. In all of these strains the tentative origins of replication deduced from the PacBio coverage plots were located in some distance to detected DnaA box clusters or the position of essential replication genes usually associated with the oriC, which suggests a non-canonical initiation of replication. To the best of our knowledge, we could show for the first time that DNA replication in a group of environmental relevant bacteria is largely independent of the initiator protein DnaA. Previously, it was noted that in silico prediction of the oriC site in the planctomycete Pirellula was hardly possible (Mackiewicz et al., 2004). Therefore, we hypothesize that initiation of DNA replication in most planctomycetes is regularly initiated at alternative sites of the chromosome thereby explaining the degeneration of a canonical oriC site. This genetic peculiarity may be especially beneficial in planctomycetes bound to hypersaline habitats because it could accelerate evolution by an increased mutation rate resulting in rapid adaptation under conditions of environmental stress.

General survey of gene abundance patterns

The gene contents of isolate genomes and MAGs were analysed to detailed insights into the metabolic potential and possible niche adaptation of representatives of the Sedimentisphaerales. A general overview of the variation of abundances of genes represented by COGs (clusters of orthologous groups) allocated to distinct functional categories in members of the novel lineage compared to the related orders Physicisphaerales and Planctomycetales is given in Fig. 4, while some essential results are detailed below.

Large shifts in the abundance of genes in category C (energy production and conversion) probably reflect an adaptation of members of the Sedimentisphaerales to anoxic environmental niches. Especially genes encoding low-potential electron-carrier proteins, like ferredoxin (COG1141), electron transfer flavoprotein (COG2086), desulfoferrodoxin (COG2033), rubredoxin (COG1773) and rubreythrin (COG1592), which play a role in fermentation reactions and oxygen detoxification were found almost exclusively in genomes of members of the novel order. In addition, genes representing a Na+-translocating RNF complex for the reoxidation of reduced ferredoxin seems to be restricted to members of this group. On the other hand, COGs representing subunits of terminal heme/copper-type cytochrome/quinol oxidases or NADH:ubiquinone oxidoreductases were absent, thereby indicating the lack of a functional respiratory electron transport chain.

An analysis of the distribution of COGs associated with the transport and metabolism of carbon compounds (categories E-I) indicates a specialization of members of this group on the utilization of carbohydrates, while genes involved in the metabolism of lipids and amino acids or peptides were generally depleted. The increase in the diversity and abundance of COGs related to the metabolism of carbohydrates (category G) is mainly caused by enzymes that play a role in the degradation of recalcitrant polysaccharides, for instance α-L-fucosidase (COG3669) or α-L-arabinofuranosidase (COG3534). In contrast, the
depletion of genes affiliated with the functional category I (lipid transport and metabolism) may be caused by a simple structure of cell membranes with a limited requirement for the production of fatty acid compounds, which is for example illustrated by a lack of COG1398 (fatty-acid desaturase) or COG1562 (phytoene/squalene synthetase).

Most functional categories of COGs associated with general cellular functions (categories J-M) tend to be depleted in members of the Sedimentisphaerales, which could point to a simpler organization and regulation of cellular functions compared to members of related phylogenetic lineages. This is confirmed by the pronounced contraction of COGs associated with the functional categories T (signal transduction mechanisms) and N (cell motility). The combination with an expansion of COGs of category W (extracellular structures) indicates the preference of a sessile lifestyle in stable environments (e.g., biofilms) over a planktonic phenotype that actively seeks optimal conditions for growth. Based on the obtained data it can be concluded that the reduction of the diversity and abundance of genes of known functional categories in members of the Sedimentisphaerales reflects a general tendency for metabolic specialization that already became apparent in the analyses of the general genome features of strains from hypersaline habitats.

**Distribution of genes with specific functions**

In addition to the general gene content analyses based on COGs, the distribution of genes with a distinct metabolic function was analysed. This survey was mainly focussed on complete genomes of the novel isolates, which were compared with the available MAGs assigned to the Sedimentisphaerales. In this way special adaptations of strains to hypersaline environments should become apparent. Some of the obtained data are shown in Table 2, while the most significant results are summarized below and illustrated in Supporting Information Fig. S3. The analysis of the abundance of transporter genes revealed that the families ABC (ATP-binding cassette superfamily) and SSS (solute:sodium symporter) were among the most prevalent types in this lineage (Supporting Information Fig. S3A). Strain SM-Chi-D1 seems to be the most versatile within this group with regard to nutrient-scavenging capabilities due to a surpassing number of transporters for organic compounds. The prevalence of SSS transporters in genomes of Sedimentisphaerales was mainly due to putative Na⁺:proline symporter, which may indicate a requirement for proline, especially in strain SM-Chi-D1. The amino acid proline is known as an effective osmoprotectant (Hayat et al., 2012; Zaprasis et al., 2015) and therefore could be used either as carbon source or osmolyte against osmotic stress by members of this lineage. The gene encoding proline dehydrogenase, a key enzyme for proline degradation in bacteria, was lacking in all genomes except MAG SM1_79, so that it seems more likely that proline or its derivate proline betaine (Amin et al., 1995) is accumulated in these strains as compatible solute to protect against desiccation or oxidative stress. Multidrug efflux pumps of the MFS- or RND-type were quite abundant in all members of this lineage, but especially in genomes of the novel isolates from hypersaline habitats. This could indicate a high resistance against antimicrobial compounds, for instance produced from antagonistic microorganisms in densely populated niches like biofilms. Interestingly, TRAP-type C4-dicarboxylate transport systems could be only detected in genomes of the Sedimentisphaerales, but not in analysed genomes of the related orders Phycisphaerales and Planctomycetales, which could indicate a possible role in the anaerobic metabolism of these bacteria. On the other hand, genes encoding sulfate permease (SulP) or ABC-type sulfate transporter were not identified, which may indicate that reduced sulfur compounds are the preferred sulfur source in strains representing this lineage. In support of this finding complete pathways for the assimilation of sulfate were lacking in most genomes, except strain ST-NAGAB-D1 and the MAG bin0 (Robbins et al., 2016). The exclusive use of reduced sulfur compounds probably reflects an adaptation to anoxic environmental niches, where these compounds are usually available in significant amounts.

A survey of peptidase genes within available genomes of this group revealed a prevalence of serine-type peptidases, besides some metallopeptidases and cysteine-type peptidases of moderate abundance (Supporting Information Fig. S3B). In general, only a few secreted or membrane-bound peptidases could be identified, which could play a role in the utilization of proteinaceous substrates. Members of the S09X subfamily, representing the most abundant peptidase-like enzymes in this group with an average of 10.6 genes per genome, were mainly annotated as carboxyl esterases that could be involved in the degradation of recalcitrant polysaccharides by removing protective carboxyl groups. A reason for the depletion of these enzymes in the obtained cultured strains could be that in hypersaline environments complex carbohydrates are modified preferentially with sulfate instead of carboxyl groups (Spring et al., 2016). Another feature of genomes of the novel isolates was the observed lack of serine peptidases of the S12 family, which was caused by the absence of putative class C beta-lactamases that were on the other hand prevalent in MAGs of uncultured strains. Representatives of metallopeptidases (especially M26 family) were in general more frequently found in the novel isolates, but in most cases the function of these enzymes remained unknown.

The distribution of carbohydrate-active enzymes in genomes of the novel isolates and other members of the Sedimentisphaerales is shown in Supporting Information Fig. S3C. In general, a prevalence of glycoside hydrolases...
targeting oligosaccharides, especially of families GH2 (e.g., β-galactosidase, β-glucuronidase), GH0 (enzymes not yet assigned to a family) and GH43 (e.g., β-xyllosidase, α-L-arabinofuranosidase) was revealed. Several of the identified glycoside hydrolase families represented enzymes required for the utilization of complex carbohydrates, like α-N-acetylgalactosaminidase (GH109), α-L-fucosidase (GH29) and α-L-rhamnosidase (GH78). On the other hand, glycosyltransferases and polysaccharide lyases involved in the depolymerisation of polysaccharides were less abundant. In the genomes of strains bound to marine or hypersaline environments a high abundance of sulfatases was observed. The newly isolated strains obtained from hypersaline habitats contained an average of 62.8 sulfatase genes per genome. The average count of sulfatase genes in MAGs retrieved from marine environments was 42.5, but from freshwater environments only 9.7. It is assumed that most sulfatases are expressed in saccharolytic microorganisms to remove sulfate groups from complex carbohydrates in order to make them more accessible for further enzymatic degradation and utilization (Wegner et al., 2013; Spring et al., 2016). The result of our survey on the presence of carbohydrate-active enzymes within this lineage suggests that its members could play a secondary role in the degradation of complex polysaccharides by utilizing partly degraded fragments produced by other bacteria under anaerobic conditions. Possible exceptions may be bacteria represented by the MAGs Plancl_01 and bin0, which were retrieved from environments supplemented with polysaccharides and encode several enzymes for the degradation of cellulose and galactomannans respectively.

Based on the presence of genes with a key function in the central metabolism (Supporting Information Table S2) the main pathways operative in members of the Sedimentisphaerales could be reconstructed (see Supporting Information, supplementary results). In summary, results of the gene content analyses and pathway reconstructions indicate that most members of this lineage thrive in anoxic aquatic habitats, have a sessile lifestyle and are specialized on the fermentation of sugars, enzymatically released from recalcitrant oligosaccharides. This allows a clear distinction to related phylogenetic lineages like Physiscphaerales and Planctomyctales, comprising mainly aerobic or facultative anaerobic bacteria that are adapted to environmental niches exposed to oxygen and are often found in association with live plant biomass, especially algae (Lage and Bondoso, 2014).

Factors driving evolution and speciation

Based on recently proposed ecological models we hypothesize that the observed general characteristics of genomes of members of the Sedimentisphaerales are driven by two main factors: First, as outlined in the study of Bentkowski et al. (2017) the type of habitat could have an effect on genome size and metabolic versatility. In this case ecological niches characterized by a stable supply of nutrients (e.g., hypersaline mats) would promote a specialization of metabolism and the reduction of metabolic maintenance costs, eventually leading to smaller streamlined genomes around three Mbp or less (e.g., strain L21-RPul-D3T). In contrast, variable environments like waste water may stimulate the evolution of larger genomes conferring a high efficiency of resource uptake and degradation capabilities (e.g., MAG Plancl_01). Second, environmental stress factors, like changing salinities, varying concentrations of oxygen, periodic desiccation, particle-attached growth or viral attacks can influence genome stability. It is assumed that in very dynamic or extreme habitats higher relative evolution rates, frequent genome rearrangements and horizontal gene transfers are favourable (Li et al., 2014; Vigil-Stenman et al., 2017), which could explain the high abundance of transposases and genomic islands in strains retrieved from hypersaline sediments and mats compared to other members of this group inhabiting moderate freshwater or marine environments.

In addition to general factors driving evolution, mechanisms of speciation can play a major role in the emerging diversity of phylogenetic lineages. A prerequisite for the detailed analyses of speciation is the reliable definition of taxonomic units (i.e., species), which has become feasible by using taxogenomic methods. An approach widely used to estimate the evolutionary distance among closely related isolates is based on the determination of the shared gene content. In Supporting Information Fig. S4, the obtained results are depicted as Venn diagrams, which illustrate the overlap of protein-coding genes among three strains related at strain or species level (16S rRNA gene identity values above 98.6%; Kim et al., 2014) and three distantly related strains (16S rRNA gene identity values below 90%). The closely related strains ST-PulAB-D4T, SM-PulAB-D3 and L21-RPul-D3T share 1987 homologous proteins corresponding to an average value of 85.4% of the total protein-coding genes of each genome (Supporting Information Fig. S4A). The largest number of unique genes is found in strain L21-RPul-D3T (211), which might reflect an adaptation to cyanobacterial mats as habitat, while both strains ST-PulAB-D4T and SM-PulAB-D3 were retrieved from sediments. In contrast, a comparison of the distantly related strains ST-PulAB-D4T, SM-Chi-D1 and ST-NAGAB-D1 revealed only 1170 shared genes, representing 33.5% to 49.3% of the total number of proteins encoded in each genome (Supporting Information Fig. S4B). Based on a recent study the percentage of conserved proteins (POCP) among two genomes can be used for genus demarcation (Qin et al., 2014). According to this proposal POCP values below 50% in combination with 16S rRNA
gene identity values below 93% would clearly indicate separate genera. Consequently, the strains ST-PulAB-D4\(^T\), SM-Chi-D1 and ST-NAGAB-D1 should be assigned to different genera.

Despite some evolutionary constraints on the relative positions of genes in genomes their order will change due to genetic rearrangements, which accumulate over the course of evolution. Therefore, the extent of synteny among genomes of closely related strains can be used to reveal their evolutionary relationships. In Supporting Information Fig. S5, synteny plots are shown that are based on the comparison of the genome of strain ST-PulAB-D4\(^T\) with the genomes of strains SM-PulAB-D3, L21-RPul-D3\(^T\), SM-Chi-D1 and ST-NAGAB-D1. Apparently, the highest synteny is found between the genomes of ST-PulAB-D4\(^T\) and SM-PulAB-D3, which represent the most closely related strains and share a 16S rRNA gene identity value of 99.8%. The strains L21-RPul-D3\(^T\) and ST-PulAB-D4\(^T\) are also closely related (98.8% 16S rRNA gene identity value), but synteny between their genomes is restricted to some distinct regions thereby indicating a continued separate evolution resulting in genetic isolation. Consistently, no recognizable synteny is evident between ST-PulAB-D4\(^T\) and the genomes of strains SM-Chi-D1 and ST-NAGAB-D1, which share only 88.2% and 86.3% 16S rRNA gene identity values with the reference strain ST-PulAB-D4\(^T\) respectively.

Currently, the most accurate genome-based tools for the definition of genera and species rely on the determination of average amino acid identity (AAI) and average nucleotide identity (ANI) values. Threshold values of 65% AAI for the demarcation of genera and 95% ANI/AAI for the definition of species are widely accepted (Konstantinidis et al., 2017). A matrix of the determined ANI and AAI values among genomes of isolates is shown in Supporting Information Fig. S6. The obtained data indicate that strains ST-PulAB-D4\(^T\) and SM-PulAB-D3 belong to the same species, while strain L21-RPul-D3\(^T\) represents a different species of the same genus. Dendrograms based on these data are in line with the reconstructed phylogenetic trees and suggest the classification of the novel strains into four species and three different genera.

Based on the obtained data it can be deduced that species of this lineage can show a disjunct distribution, even if adapted to anoxic hypersaline conditions impeding their dispersal. On the other hand, distantly related strains could be retrieved from the same sediment sample at the same time (SM-PulAB-D3 and SM-Chi-D1), indicating a prolonged genetic isolation by niche separation. Consequently, we hypothesize that halophilic members of the Sedimentisphaerales belong to a global community of extremophilic microorganisms that evolves by sympatric speciation and is dispersed by yet unknown mechanisms.

### Descriptions of novel species and classification into higher taxa

Phenotypic traits of the novel isolates were determined and correlated with the revealed genotypic characteristics. Besides several common characteristics a considerable variability of phenotypes was revealed, which is in good agreement with the large phylogenetic distance between some of the isolates. The optimal growth conditions of these strains reflected the habitats from which they were isolated and were at temperatures between 30 and 40°C, pH values around 7.5 and salinities around 10% (w/v) NaCl. The cell morphology of all strains was spherical and the average size ranged from 0.5 to 1.2 \(\mu\)m. In Fig. 5, some morphological characteristics of strain L21-RPul-D3\(^T\) are shown, which were also typically of the remaining strains. Cells had a Gram-negative type of cell wall, were non-motile and multiplied by binary fission (Fig. 5A, B, D). Intracellular membranes were not detected. With the exception of strain SM-Chi-D1, all strains produced extracellular polymeric substances (EPS), which led to an increase of medium viscosity during growth, especially with polysaccharides as carbon source (Fig. 5C and D). On the other hand, strain SM-Chi-D1 had the tendency to form aggregates, especially in media containing sulfide as reductant. The pattern of fermentation products observed upon growth with glucose indicated a mixed acid fermentation. Ethanol was a major end product in all strains except strain SM-Chi-D1, which produced mainly lactate.

More details about the phenotypic characterization of the novel isolates, especially chemotaxonomic features, are reported in the Supporting Information (Supplementary Results, Table S3 and Fig. S7). Based on the combination of genotypic and phenotypic data the novel genus Sedimentisphaera comprising two novel species represented by the type strains ST-PulAB-D4\(^T\) and L21-RPul-D3\(^T\) is proposed. Main phenotypic characteristics that distinguish both type strains include a significant tolerance against cytosine and the utilization of arabinose, cellobiose and xylose restricted to strain L21-RPul-D3\(^T\), while only strain ST-PulAB-D4\(^T\) is able to grow at 45°C and pH 8.5. In Supporting Information Table S4, characteristic traits of the novel type strains are summarized and compared with representatives of related species within the Phycisphaerae class. Detailed formal descriptions and a proposed classification of cultivated members of this clade into high rank taxa will be presented below. Descriptions of the remaining strains SM-Chi-D1 and ST-NAGAB-D1 are in preparation and will be presented elsewhere.

Description of *Sedimentisphaera gen. nov.*

*Sedimentisphaera* (Se.d.i.men.ti.spha'e.ra. N.L. masc. n. *sedimentum*, a settling, sediment; L. fem. n. *sphaera*, a globe, sphere; N.L. fem. n. *Sedimentisphaera*, a spherical bacterium from sediment).
Free-living, unpigmented, non-motile and non-spore-forming coccoid cells that divide by binary fission. Intracellular membranes are not formed. Gram-negative type of cell wall with a peptidoglycan layer. Dominating cellular fatty acids are \textit{anteiso-C_{15}:0} and \textit{anteiso-C_{13}:0}. Unsaturated cellular fatty acids, respiratory lipoquinones or cytochromes are not present. The polar lipid composition is dominated by phosphatidylglycerol and several distinct glycolipids. Tests for catalase and oxidase are negative. Nitrate is not reduced. Sulfate is not assimilated. Obligately anaerobic, moderately halophilic, mesophilic and neutrophilic. Vitamins required for growth. Strictly fermentative metabolism with sugars as preferred substrates. Medium viscosity increases during growth due to the production of extracellular polymeric substances. Oxygen can be tolerated up to 0.5% (v/v) in the gas atmosphere but does not stimulate growth. Susceptible to the antibiotics chloramphenicol and tetracycline, but resistant to rifampicin.

The type species is \textit{Sedimentisphaera salicampi}.

Description of \textit{Sedimentisphaera salicampi} sp. nov. \textit{Sedimentisphaera salicampi} (sa.li.cam’pi. L. n. sal salt; L. n. campus field; N.L. gen. n. salicampi of a salt field).

Shows the following characteristics in addition to those given for the genus. Most cells have a diameter ranging from 0.7 to 1.2 \textmu m. Optimal conditions for growth are 37°C, pH 7.5 to 8.0 and a salinity of 10% (w/v) NaCl; temperatures from 20 to 45°C, pH values from 6.0 to 8.5 and salinities from 60 to 160 g l\(^{-1}\) NaCl are tolerated. The vitamins biotin and B\(_{12}\) are required for growth in defined mineral media containing a single carbon source. The following carbon sources support fermentative growth: dextran, \textit{d}-fructose, \textit{d}-galactose, \textit{d}-glucose, pullulan, \textit{d}-lactose, \textit{d}-maltose, \textit{d}-mannose, \textit{d}-melibiose, starch and sucrose.

The following compounds were tested, but not utilized: acetate, N-acetylglucosamine, agar, alginate, l-arabinose, benzoate, butanol, butyrate, carrageenan, Casamino acids, \textit{d}-cellobiose, cellulose, chitin, citrate, ethanol, furfuran, \textit{l}-fucose, fumarate, glycerol, lactate, malate,
d-mannitol, methanol, propionate, propanol, pyruvate, L-rhamnose, D-ribose, succinate, trehalose, Tryptone, D-xylene and yeast extract. The non-gaseous end products resulting from D-glucose fermentation are ethanol, acetate, formate and lactate. Susceptible to chloramphenicol, cycloserine and tetracycline (each at 20 mg l\(^{-1}\)). Ampicillin, carbenicillin, penicillin G and rifampicin are tolerated at least in concentrations up to 100 mg l\(^{-1}\), while kanamycin A and gentamicin are tolerated up to 1000 mg l\(^{-1}\). Unidentified polar lipids and phospholipids are present in addition to phosphatidylglycerol and glycolipids. In addition to the dominating cellular fatty acids listed in the description of the genus, significant amounts (>5% of total amount) of iso-C\(_{14:0}\), iso-C\(_{16:0}\), C\(_{20:0}\), C\(_{16:0}\) and C\(_{18:0}\) are present. The DNA G+C content of the type strain is 46.3 mol%.

The type strain is ST-PuLAB-D4\(^{T}\) (= DSM 101733\(^{T}\) = JCM 31929\(^{T}\) = KCTC 15597\(^{T}\)), isolated from the anoxic sediment of an evaporation pond of a solar saltern in Tavira, Portugal.

Description of *Sedimentisphaera cyanobacteriorum* sp. nov.

*Sedimentisphaera cyanobacteriorum* (cy.a.no.bac. te.ri.o’-rum. N.L. neut. gen. pl. cyanobacteriorum of cyanobacteria).

Shows the following characteristics in addition to those given for the genus. Most cells have a diameter ranging from 0.7 to 1.2 \(\mu\)m. Optimal conditions for growth are 35°C, pH 7.5 and a salinity of 12% (w/v) NaCl; temperatures from 18 to 40°C, pH values from 6.5 to 8.0 and salinities from 70 to 180 g l\(^{-1}\) NaCl are tolerated. The vitamins biotin and B\(_{12}\) are required for growth in defined mineral media containing a single carbon source. The following carbon sources support fermentative growth: L-arabinose, D-cellobiose, dextran, D-fructose, D-galactose, D-glucose, pullulan, D-lactose, D-maltose, D-mannose, D-melibiose, starch, sucrose, and D-xylene.

The following compounds were tested, but not utilized: acetate, N-acetylglycosamine, agar, alginate, benzoate, butanol, butyrate, carrageenan, Casamino acids, cellulose, chitin, citrate, ethanol, fucoidan, L-fucose, fumarate, glycerol, lactate, malate, p-mannitol, methanol, propionate, propanol, pyruvate, L-rhamnose, D-ribose, succinate, trehalose, Tryptone, and yeast extract. The non-gaseous end products resulting from D-glucose fermentation are ethanol, acetate, formate and lactate. Susceptible to chloramphenicol and tetracycline (each at 20 mg l\(^{-1}\)). Ampicillin, carbenicillin, cycloserine, penicillin G and rifampicin are tolerated at least in concentrations up to 100 mg l\(^{-1}\), while kanamycin A and gentamicin are tolerated up to 1000 mg l\(^{-1}\). In addition to the dominating cellular fatty acids listed in the description of the genus, significant amounts (>5% of total amount) of iso-C\(_{14:0}\), iso-C\(_{16:0}\), C\(_{16:0}\) and iso-C\(_{12:0}\) are present. The DNA G+C content of the type strain is 47.4 mol%.

The type strain is L21-RPul-D3\(^{T}\) (= DSM 27091\(^{T}\) = JCM 19196\(^{T}\) = KCTC 15598\(^{T}\)), isolated from the suboxic zone of a lithifying cyanobacterial mat in the hypersaline Lake 21 at the Kiritimati Atoll, Republic of Kiribati.

Description of *Sedimentisphaeraceae* fam. nov.

*Sedimentisphaeraceae* (Se.di.men.tisphae.rae’. N.L. fem. dim. n. Sedimentisphaera, type genus of the family; suff. –aceae, ending to denote a family; N.L. fem. pl. n. Sedimentisphaeraceae, the *Sedimentisphaera* family).

Gram-negative, free-living, unpigmented, non-motile and non-spore-forming cocoid cells that divide by binary fission. Intracellular membranes are not formed. Resistant against rifampicin. The family encompasses mainly bacteria found in anoxic hypersaline environments.

The affiliation of novel species to this family depends on the phylogenetic position, which should be determined on the basis of comparative sequence analyses of 16S rRNA genes. The 16S rRNA gene sequence identity values of newly described strains affiliated with this family to the type strain of the type species *S. salicampi* should be around 86% or above, which represents the threshold recommended for the definition of families (Yarza et al., 2014).

The type genus of the family is *Sedimentisphaera*.

Description of *Sedimentisphaerales* ord. nov.

*Sedimentisphaerales* (Se.di.men.tisphae.ales. N.L. fem. dim. n. Sedimentisphaera, type genus of the order; suff. –ales, ending to denote an order; N.L. fem. pl. n. Sedimentisphaerales, the *Sedimentisphaera* order).

Gram-negative, non-spore-forming cocoid cells that divide by binary fission. Representatives are mainly restricted to anoxic aquatic environments.

The affiliation of novel species to this order depends on the phylogenetic position, which should be determined on the basis of comparative sequence analyses of 16S rRNA genes and/or RpoB proteins.

The type genus of the order is *Sedimentisphaera*.

**Experimental procedures**

Sources, enrichment and isolation of anaerobic *Phycisphaerae* strains

Enrichment and isolation of strain L21-RPul-D3\(^{T}\) from the suboxic zone of a lithifying cyanobacterial mat was described previously (Spring et al., 2015). Anoxic sediment samples for the enrichment of further strains were taken in May 2014 and May 2015 from evaporation ponds of commercial solar salt-erns in Spain (Salinas de Es Trenc, Campos, Mallorca) and Portugal (Rui Simeão Tavira Sal, Tavira) respectively. The salinities of pond sediments from Es Trenc and Tavira were estimated with a portable refractometer (Kern; Balingen-Frommern, Germany) to be around 10% and 25% respectively. Sediment samples were transported to the DSMZ for further processing in Schott glass bottles closed with butyl

© 2018 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*, 20, 2438–2455
rubber stoppers to retain anoxic conditions. Anoxic media used for enrichment and cultivation were prepared according to the anaerobic cultivation technique of Hungate (Hungate, 1950) with the modifications introduced by Bryant (Bryant, 1972). The used enrichment media were based on the defined mineral solution described previously containing either 60.0 or 120.0 g l\(^{-1}\) NaCl (Ben Hania et al., 2017). Enrichment cultures were incubated anaerobically in the dark at 35°C.

Strain ST-PulAB-D4\(^{T}\) was enriched from Tavira saline sediment using a decimal dilution series prepared in medium with 120.0 g l\(^{-1}\) NaCl containing 0.5 g l\(^{-1}\) pullulan as substrate and the antibiotics rifampicin and carbenicillin at a concentration of 0.1 g l\(^{-1}\), each. After an incubation time of three months the fourth tube of the dilution series showed a positive growth response and was used for purification. Strain SM-PulAB-D3 was enriched from Es Trenc saline sediment using a decimal dilution series prepared in medium with 60.0 g l\(^{-1}\) NaCl containing 0.5 g l\(^{-1}\) pullulan as substrate and the antibiotics rifampicin and carbenicillin at a concentration of 0.1 g l\(^{-1}\), each. After an incubation time of seven weeks the third tube of the dilution series displayed growth of Gram-negative cocci. Strain SM-Chi-D1 was also enriched from Es Trenc sediment using a medium with 60.0 g l\(^{-1}\) NaCl and 0.5 g l\(^{-1}\) chitin as substrate. Growth of Gram-negative cocci was observed in this culture after an incubation time of 6 weeks. Strain ST-NAGAB-D1 was enriched from Tavira saline sediment using a medium with 120.0 g l\(^{-1}\) NaCl containing 1.0 g l\(^{-1}\) N-acetylglucosamine as substrate and the antibiotics rifampicin and carbenicillin at a concentration of 0.1 g l\(^{-1}\), each. Growth of Gram-negative cocci was observed after an incubation time of three weeks.

The obtained enrichment cultures were further purified by several successive dilution-to-extinction series in basal mineral medium without antibiotics and 1.0 g l\(^{-1}\) glucose as carbon source. For the purification of strain SM-Chi-D1 it was necessary to supplement the medium with 0.5 g l\(^{-1}\) yeast extract in addition to glucose. The purity of isolates was examined by phase contrast microscopy, streaking cultures on aerobic and anaerobic agar plates and direct sequencing of PCR-amplified 16S rRNA genes.

For routine cultivation of strains ST-PulAB-D4\(^{T}\), SM-PulAB-D3, L21-RPul-D3\(^{T}\) and ST-NAGAB-D1 the DSMZ medium 1527 (http://www.dsmz.de/microorganisms/medium/pdf/DSMZ_Medium1527.pdf) was used, while strain SM-Chi-D1 was cultured in DSMZ medium 1526a (http://www.dsmz.de/microorganisms/medium/pdf/DSMZ_Medium1526a.pdf) supplemented with 0.5 g l\(^{-1}\) yeast extract.

**Phylogeny and biogeography**

The 16S rRNA gene sequence of strain L21-RPul-D3\(^{T}\) was determined previously (Spring et al., 2015) and deposited in the GenBank/EMBL/DDBJ databases under the accession number KC665951. The 16S rRNA gene sequences of the strains SM-PulAB-D3, SM-Chi-D1, ST-PulAB-D4\(^{T}\) and ST-NAGAB-D1 were determined in this study and deposited in the GenBank/EMBL/DDBJ databases under the accession numbers KP729375, KX151396, KU672518 and KY471010 respectively. The newly determined 16S rRNA gene sequences were added to the alignment of the SILVA database (Quast et al., 2013; SSU Ref NR 99 release 128) using the integrated aligner of the ARB software package (Ludwig et al., 2004). Based on the curated guide tree included in the SSU Ref NR 99 database reference sequences representing all delineated clades of the *Phycisphaerae* class and several sequences assigned to main lineages of the class *Planctomyceta* were selected resulting in a comprehensive data set for phylogenetic analyses. RpoB protein sequences were extracted from genome sequences of the newly isolated strains or obtained from the GenBank or Uniprot protein databases in case of reference strains. In addition, 16S rRNA gene sequences and RpoB protein sequences were extracted from MAGs representing bacteria phylogenetically related with the newly isolated strains. RpoB protein sequences were aligned using the ClustalW algorithm implemented in the ARB package.

Phylogenetic trees based on aligned data sets of 16S rRNA gene or RpoB protein sequences were reconstructed using programs implemented in the ARB software package. When the ARB neighbor-joining program was used, phylogenetic distances were calculated with the corrections of Jukes-Cantor for nucleic acids and Kimura for proteins. Maximum likelihood trees were reconstructed using RAxML (version 7.7.2) with the GTR+Gamma model for DNA and PROTCAT-LG for proteins under the rapid bootstrap analysis algorithm. The maximum parsimony program of PHYLIP included in the ARB package was used with default settings for nucleotide (dna pars) or amino acid (protpars) sequences. The robustness of the tree topologies was evaluated by performing 1000 rounds of bootstrap replicates. The phylogenetic diversity within a group of sequences was deduced with the similarity option of the ARB distance matrix program. Sequences of low quality (labelled in red in the SILVA 128 SSU Ref NR99 guide tree) were excluded from the calculation. The environmental distribution of sequences representing members of the class *Phycisphaerae* was deduced from the sources of isolation given in the SILVA 128 SSU Ref NR 99 data set.

**Genome sequencing**

Genomic DNA was isolated from stationary cultures of the newly isolated strains using the Qiagen Genomic Tip 100/G (Qiagen, Hilden, Germany) according to the protocol provided by the manufacturer. For strains ST-PulAB-D4\(^{T}\), L21-RPul-D3\(^{T}\), SM-Chi-D1 and ST-NAGAB-D1 complete genome sequences were determined using a combination of two genomic libraries of which one was prepared for sequencing with the PacBio RSII (Pacific Biosciences, Menlo Park, CA, USA) platform. The SMRTbell\(^{TM}\) template library was prepared and sequenced according to the instructions from Pacific Biosciences following the Procedure & Checklist ‘Greater than 10 kb Template Preparation and Sequencing’. For all strains a single SMRT cell was sequenced using a single 240-min movie with exception of strain ST-NAGAB-D1, where two SMRT cells were used and L21-RPul-D3\(^{T}\), where five SMRT cells were sequenced applying the older C2 chemistry. Paired-end libraries for hybrid error correction were generated and sequenced on the Illumina MiSeq platform (Illumina, San Diego, CA, USA) with 150 cycles. Hereby, inserts of an average size of 500 bp were achieved and about 6 million reads generated per run. Draft long read genome assemblies were generated.
using the ‘RS_HGAP_Assembly.3’ protocol included in SMRTPortal version 2.3.0 applying default parameters. L21-RPul-D3\(^3\) has been assembled in SMRTPortal 2.0.1 using the protocol described previously (Ben Hania et al., 2015). For each of the genomes one chromosomal contig could be obtained, which was trimmed, circularized and adjusted to dnaA as first gene. Total coverages of ~125x have been calculated within the long-read assembly process with exception of L21-RPul-D3\(^3\), where 220x was calculated due to the higher throughput needed for complete genome assembly. DNA base modifications analyses were performed by ‘RS_Modification_and_Motif_Analysis’ protocol with default settings. Hybrid error correction was performed for each of the genomes by a mapping of Illumina short read data onto the draft circular chromosomes using BWA (Li and Durbin, 2009) followed by automated variant calling using VarScan 2 (Koboldt et al., 2009). Quality check of the final consensus sequences regarding overall coverage as well as SNPs was performed using IGV (Thorvaldsdóttir et al., 2013). For strain SM-PulAB-D3 only a draft genome was determined based on Illumina HiSeq sequencing resulting in 63 contigs after short-read assembly with Velvet 1.2.10 (Zerbino and Birney, 2008) and a genome coverage of 811x. The quality of draft genomes was checked using the software CheckM (Parks et al., 2015).

Genome annotation was primarily done using PROKKA version 1.8 (Seemann, 2014). Annotated genomes were then compared with results provided by the RAST server (Overbeek et al., 2014). In cases where automatic annotation by RAST and PROKKA led to aberrant results, the function prediction of PROKKA was checked and eventually corrected manually by BLASTP searches for similar proteins in the UniProtKB database (http://www.uniprot.org/blast/). Additional gene prediction analyses and functional annotation was performed within the Integrated Microbial Genomes - Expert Review (IMG/M-ER) platform (Markowitz et al., 2009).

The annotated genomes were deposited in NCBI GenBank under the following accession numbers: CP021023 (ST-PuLAB-D4\(^4\)), NHPQ00000000 (SM-PuLAB-D3), CP018633 (L21-RPul-D3\(^3\)), CP018646 (SM-Chi-D1) and CP019791 (ST-NAGAB-D1).

**Comparative genome analyses**

The software Ori finder (Gao and Zhang, 2008) was used to determine the GC disparity curves and location of dnaA boxes of complete genome sequences. Putative genomic islands were detected with IslandViewer 4.0 (Bertelli et al., 2017) and tentative prophage regions were identified using the web server PHASTER (Arndt et al., 2016). The program DNAPlotter (Carver et al., 2009) was used for the illustration of the localization of genes of interest, putative genomic islands, GC skew plots and GC plots.

General gene abundance patterns were determined by comparison gene contents of members of the proposed order Sedimentisphaerales with members of the related orders Planctomycetales and Cytophagales. Genomes used in the analyses are depicted in the phylogenetic tree shown in Supporting Information Fig. S1. Abundance profiles of genes allocated to distinct COGs were derived by applying the respective function in the Compare Genomes menu of the IMG/M-ER server (Chen et al., 2017) using the output types Matrix and Gene count. Based on the obtained results average values of gene counts of a distinct COG in each phylogenetic group were calculated. For a detailed analysis of the prevalence of genes with a distinct function within genomes of the newly isolated strains in comparison to MAGs of other members of the Sedimentisphaerales the search functions of several specific databases were applied. The database TransportDB 2.0 (Elbourne et al., 2017) was used to explore the abundance profiles of membrane transporter genes, MEROPS (Rawlings et al., 2016) for the analyses of peptidase genes and CAZy (Lombard et al., 2014) for the detection of genes of carbohydrate-active enzymes. Database searches were based on sets of protein sequences generated from the respective genome sequences using the RAST Prokaryotic Genome Annotation Server (Overbeek et al., 2014). Venn diagrams showing the shared gene content among genomes were drawn based on calculations made with the analysis option ‘Phylogenetic Profiler for Single Genes’ available from the IMG/M-ER server. Synteny plots of genomes of the newly isolated strains were generated by using the IMG/ MER Dotplot Synteny Viewer that employs Mummer to generate a dotplot between sets of two genomes. Average nucleotide identity (ANI) and average amino acid identity (AAI) values were determined with the OrthoANI calculator (Lee et al., 2016) and the AAI Matrix Calculator (Rodriguez-r and Konstantinidis, 2016) respectively.

**Phenotypic characterization**

The Gram reaction was determined with heat fixed liquid cultures stained with BD Diffco kit reagents. The shape and size of live cells was observed with an AxioScope.A1 phase contrast microscope (Carl Zeiss, Jena, Germany). Samples for electron microscopy were prepared according to the protocols described elsewhere (Wittmann et al., 2014). The pH and temperature ranges for growth were determined using DSMZ medium 1527 (ST-PuLAB-D4\(^4\), SM-PuLAB-D3, L21-RPul-D3\(^3\), ST-NAGAB-D1) or DSMZ medium 1526a supplemented with 0.5 g l\(^{-1}\) yeast extract (SM-Chi-D1) as described previously (Ben Hania et al., 2017). The salinity range for growth was determined by directly weighing NaCl in Hungate-type tubes before dispensing medium. Substrates were tested at a final concentration of 2.0 g l\(^{-1}\) (complex compounds), 0.5 g l\(^{-1}\) (polysaccharides) or 1.0 g l\(^{-1}\) (sugars, carboxylic acids, alcohols) in cultivation media without glucose. Potential electron acceptors were added to media in final concentrations of 20 mM (sodium thiosulfate, sodium sulfate, sodium nitrate) or 2 mM (sodium nitrate). Cultures were subcultured at least twice under the same experimental conditions before determination of the growth response. Susceptibility to antibiotics was tested as previously described (Ben Hania et al., 2015). Nongaseous end products of fermentative growth on glucose except formate were determined by gas chromatography according to published protocols (Holdemann et al., 1977; Steer et al., 2001). Formate was quantified by a colorimetric assay (Lang and Lang, 1972). The preparation and extraction of fatty acid methyl esters from biomass and their subsequent separation and identification by gas chromatography was performed as described elsewhere (Miller, 1982; Kaksonen et al., 2006). Extraction and analyses of respiratory lipoquinones
and polar lipids were carried out according to published protocols (Tindall, 1990; Tindall et al., 2007).

Acknowledgements

The owners of the salines Rui Simeao Tavira Sai in Tavira, Portugal and Salinas de Es Trenc in Mallorca, Spain are acknowledged for the permission to take samples. We thank Simone Severitti and Nicole Heyer for excellent technical assistance regarding SMRTbell™ template preparation and sequencing. We are grateful to the Genome Analytics group (HZI Braunschweig) for providing Illumina sequence data. The Identification Service team of the DSMZ (Anja Frühling, Gabriele Pütter, Anika Wasner, Brian Tindall, Peter Schumann and Susanne Verbarg) is acknowledged for cellular fatty acid, respiratory quinone, polar lipid and peptidoglycan analyses. We thank Ramon Rossello-Mora for providing helpful comments about speciation and Aharon Oren for the proposal of a species epithet. The work on this study was funded by the Leibniz Institute DSMZ and the German Research Foundation, project Kl 1000/2-2 of the Research Unit 571 ‘Geobiology species epithet. The work on this study was funded by the Leibniz Institute DSMZ and the German Research Foundation, project Kl 1000/2-2 of the Research Unit 571 ‘Geobiology'.

References

Amin, U.S., Lash, T.D., and Wilkinson, B.J. (1995) Proline betaine is a highly effective osmoprotectant for Staphylococcus aureus. Arch Microbiol 163: 138–142.

Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., et al. (2016) Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat Commun 7: 13219.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. (2016) PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44: W16–W21.

Baker, B.J., Lazar, C.S., Teske, A.P., and Dick, G.J. (2015) Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. Microbiome 3: 14.

Bentkowski, P., van Oosterhout, C., Ashby, B., and Mock, T. (2017) The effect of extrinsic mortality on genome size evolution in prokaryotes. ISME J 11: 1011–1018.

Bertelli, C., Laird, M.R., Williams, K.P., Simon Fraser University Research Computing Group, Lau, B.Y., Hoad, G., et al. (2017) IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res 17: 132.

Boedecker, C., Schüler, M., Reintjes, G., Jeske, O., van Teeseling, M.C.F., Jogler, M., et al. (2017) Determining the bacterial cell biology of Planctomycetes. Nat Commun 8: 14853.

Bondoso, J., Harder, J., and Lage, O.M. (2013) rpoB gene as a novel molecular marker to infer phylogeny in Planctomycetales. Antonie Van Leeuwenhoek 104: 477–488.

Bryant, M. (1972) Commentary on the Hungate technique for culture of anaerobic bacteria. Am J Clin Nutr 25: 1324–1328.

Carver, T., Thomson, N., Bleasby, A., Berriman, M., and Parkhill, J. (2009) DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 25: 119–120.

Chen, I.-M.A., Markowitz, V.M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., et al. (2017) IMG/M: integrated genome and metagenome comparative data analysis system. Nucleic Acids Res 45: D507–D516.

Elbourne, L.D.H., Tetu, S.G., Hassan, K.A., and Paulsen, I.T. (2017) TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. Nucleic Acids Res 45: D320–D324.

Elshahed, M.S., Youssef, N.H., Luo, Q., Najar, F.Z., Roe, B.A., Sisk, T.M., et al. (2007) Phylogenetic and metabolic diversity of Planctomycetes from anaerobic, sulfide- and sulfur-rich Zodletone Spring, Oklahoma. Appl Environ Microbiol 73: 4707–4716.

Fuerst, J.A., and Sagulenko, E. (2011) Beyond the bacterium: planctomycetes challenge our concepts of microbial structure and function. Nat Rev Microbiol 9: 403–413.

Fukunaga, Y., Kurahashi, M., Sakiyama, Y., Ohuchi, M., Yokota, A., and Harayama, S. (2009) Phycisphaerae mikurumensis gen. nov., sp. nov., isolated from a marine alga, and proposal of Phycisphaeraeae fam. nov., Phycisphaerales ord. nov. and Phycisphaerae classis nov. in the phylum Planctomycetes. J Gen Appl Microbiol 55: 267–275.

Gao, F., and Zhang, C.-T. (2008) Ori-Finder: a web-based system for finding oriC in unannotated bacterial genomes. BMC Bioinformatics 9: 79.

Giovannoni, S.J., Cameron Thrash, J., and Temperton, B. (2014) Implications of streamlining theory for microbial ecology. ISME J 8: 1553–1565.

Ben Hania, W., Joseph, M., Bunk, B., Sprotter, C., Klenk, H.-P., Fardeau, M.-L., et al. (2017) Characterization of the first cultured representative of a Bacteroidetes clade specialized on the scavenging of cyanobacteria. Environ Microbiol 19: 1134–1148.

Ben Hania, W., Joseph, M., Schumann, P., Bunk, B., Feibig, A., Sprotter, C., et al. (2015) Complete genome sequence and description of Salinispira pacifica gen. nov., sp. nov., a novel spirochaete isolated form a hypersaline microbial mat. Stand Genomic Sci 10: 7.

Harrison, B.K., Zhang, H., Berelson, W., and Orphan, V.J. (2009) Variations in archaeal and bacterial diversity associated with the sulfate-methane transition zone in continental margin sediments (Santa Barbara Basin, California). Appl Environ Microbiol 75: 1487–1499.

Hayat, S., Hayat, Q., Alyemeni, M.N., Wani, A.S., Pichtel, J., and Ahmad, A. (2012) Role of proline under changing environments: a review. Plant Signal Behav 7: 1456–1466.

Holdemann, L.V., Cato, E.P., and Moore, W.E.C. (1977) Anaerobe Laboratory Manual 4th ed. Blacksburg, VA: Anaerobe Laboratory, Virginia Polytechnic Institute and State University.

Hungate, R. (1950) The anaerobic mesophilic cellulolytic bacteria. Bacteriol Rev 14: 1–49.

Jetten, M.S.M., Wagner, M., Fuerst, J., Van Loosdrecht, M., Kuenen, G., and Strous, M. (2001) Microbiology and application of the anaerobic ammonium oxidation ('anammox') process. Curr Opin Biotechnol 12: 283–288.

Kaksonen, A.H., Spring, S., Schumann, P., Kroppenstedt, R.M., and Puhakka, J.A. (2006) Desulfofracomaculum
thermosubterraneum sp. nov., a thermophilic sulfate-reducer isolated from an underground mine located in a geothermally active area. *Int J Syst Evol Microbiol* **56**: 2603–2608.

Kim, M., Oh, H.-S., Park, S.-C., and Chun, J. (2014) Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* **64**: 346–351.

Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., et al. (2009) VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**: 2283–2285.

Konstantinidis, K.T., Rosselló-Mora, R., and Amann, R. (2017) Uncultivated microbes in need of their own taxonomy. *ISME J* **11**: 2399–2406.

Kovaleva, O.L., Merkel, A.Y., Novikov, A.A., Basierov, R.V., Toshchakov, S.V., and Bonch-Osmolovskaya, E.A. (2015) *Tepidisphaera mucosa* gen. nov., sp. nov., a moderately thermophilic member of the class *Phycisphaerae* in the phylum *Planctomycetes*, and proposal of a new family, *Tepidisphaeraceae* fam. nov., and a new order, *Tepidisphaerales* ord. nov. *Int J Syst Evol Microbiol* **65**: 549–555.

Lage, O.M., and Bondoso, J. (2014) *Planctomycetes* and coralloidae, a striking association. *Front Microbiol* **5**: 267.

Lang, E., and Lang, H. (1972) Spezifische Farbreaktion zum direkten Nachweis der Ameisensäure. *Fresenius' Zeitschrift Für Anal Chemie* **260**: 8–10.

Lee, I., Kim, Y.O., Park, S.C., and Chun, J. (2016) OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* **66**: 1100–1103.

Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li, S.-J., Hua, Z.-S., Huang, L.-N., Li, J., Shi, S.-H., Chen, L.-X., et al. (2014) Microbial communities evolve faster in extreme environments. *Sci Rep* **4**: 6205.

Liesack, W., and Stackebrandt, E. (1989) Evidence for unlinked rrn operons in the planctomycete *Pirellula marina*. *J Bacteriol* **171**: 5025–5030.

Lombard, V., Golaonca Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**: D490–D495.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.

Mackiewicz, P., Zakrzewska-Czerwinska, J., Zawilak, A., Dudek, M.R., and Cebrat, S. (2004) Where does bacterial replication start? Rules for predicting the oriC region. *Nucleic Acids Res* **32**: 3781–3791.

Madukey, N.Z., Tehranchi, A.K., Wang, J.D., and Kreuzer, K.N. (2014) Replication of the *Escherichia coli* chromosome in RNase HI-deficient cells: multiple initiation regions and fork dynamics. * Mol Microbiol* **91**: 39–56.

Mahmoudi, N., Robeson, M.S., Castro, H.F., Fortney, J.L., Techtmann, S.M., Joyner, D.C., et al. (2015) Microbial community composition and diversity in Caspiian Sea sediments. *FEMS Microbiol Ecol* **91**: 1–11.

Markowitz, V.M., Mavromatis, K., Ivanova, N.N., Chen, I.-M., A., Chu, K., and Kyrpides, N.C. (2009) IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* **25**: 2271–2278.

Marraffini, L.A., and Sontheimer, E.J. (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* **11**: 181–190.

Masai, H., and Arai, K. (1996) DnaA- and PriA-dependent promoters: two distinct replication complexess for replication of *Escherichia coli* chromosome. *Front Biosci* **1**: d48–d58.

Miller, L.T. (1982) Single derivatization method for routine analysis of bacterial whole-cell fatty acid methyl esters, including hydroxy acids. *J Clin Microbiol* **16**: 584–586.

Mollet, C., Drancourt, M., and Raoult, D. (1997) *ropB* sequence analysis as a novel basis for bacterial identification. * Mol Microbiol* **26**: 1005–1011.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., et al. (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* **42**: D206–D214.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–1055.

Qin, Q.-L., Xie, B.-B., Zhang, X.-Y., Chen, X.-L., Zhou, B.-C., Zhou, J., et al. (2014) A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol* **196**: 2210–2215.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.

Rawlings, N.D., Barrett, A.J., and Finn, R. (2016) Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **44**: D343–D350.

Robbins, S.J., Evans, P.N., Parks, D.H., Golding, S.D., and 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.

Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2015) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* **43**: D298–D299.

Rocha, E.P.C. (2004) The replication-related organization of bacterial genomes. *Microbiology* **150**: 1609–1627.

Rodriguez-R, L.M., and Konstantinidis, K.T. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *Peer J Prepr* **4**: e1900v1.

Seemann, T. (2014) Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–2069.

Skovgaard, O., Bak, M., Løbner-Olesen, A., and Tommerup, N. (2011) Genome-wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing. *Genome Res* **21**: 1388–1393.

Spring, S., Brinkmann, N., Murjaa, M., Sproer, C., Reitner, J., and Klein, H.-P. (2015) High diversity of culturable prokaryotes in a lithifying hypersaline microbial mat. *Geomicrobiol J* **32**: 332–346.

Spring, S., Bunk, B., Sproer, C., Schumann, P., Rohde, M., Tindall, B.J., et al. (2016) Characterization of the first cultured representative of *Verrucomicrobia* subdivision 5.
indicates the proposal of a novel phylum. *ISME J* 10: 2801–2816.

Steer, T., Collins, M.D., Gibson, G.R., Hippe, H., and Lawson, P.A. (2001) *Clostridium hathewayi* sp. nov., from human faeces. *Syst Appl Microbiol* 24: 353–357.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirow, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14: 178–192.

Tindall, B.J. (1990) Lipid composition of *Halobacterium lacusprofundi*. *FEMS Microbiol Lett* 66: 199–202.

Vasu, K., and Nagaraja, V. (2013) Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev* 77: 353–357.

Wegner, C.E., Richter-Heitmann, T., Klindworth, A., Klockow, C., Richter, M., Achstetter, T., et al. (2013) Expression of sulfatases in *Rhodopirellula baltica* and the diversity of sulfatases in the genus *Rhodopirellula*. *Mar Genomes* 9: 51–61.

Wittmann, J., Bremer, E. (2015) Uptake of amino acids and their metabolic conversion into the compatible solute proline confers osmoprotection to *Bacillus subtilis*. *Appl Environ Microbiol* 81: 250–259.

Zerbino, D.R., and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18: 821–829.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Phylogenetic tree of the PVC superphylum deduced from RpoB protein sequences showing the placement of the newly isolated anaerobic strains within the proposed order *Sedimentisphaerales*.

Fig. S2. Genome structures and origins of replication in two reference strains isolated from a hypersaline cyanobacterial mat.

Fig. S3. Variation in the abundance of genes encoding proteins with known functions in genomes of cultured strains affiliated with the proposed order *Sedimentisphaerales*.

Fig. S4. Venn diagrams displaying the shared gene content among isolate genomes.

Fig. S5. Synteny plots of isolate genomes illustrating various levels of evolutionary relationships.

Fig. S6. Taxonomic comparison of cultured strains affiliated with the proposed order *Sedimentisphaerales*.

Fig. S7. Polar lipid patterns of strains affiliated with the proposed genus *Sedimentisphaera* after two-dimensional thin layer chromatography.

Table S1. Phylogenetic diversity within major lineages of the class *Phycisphaerae* based on 16S rRNA gene sequences.

Table S2. Distribution of key genes of central metabolic pathways in complete genomes of cultured strains and almost complete MAGs representing uncultured members of the proposed order *Sedimentisphaerales*.

Table S3. Cellular fatty acid patterns of species affiliated with the proposed genus *Sedimentisphaera* and type species of other genera of the *Phycisphaerae* class.

Table S4. Phenotypic traits of type strains affiliated with the proposed genus *Sedimentisphaera* compared to representatives of the related genera *Phycisphaera*, *Algisphaera* and *Tepidisphaera*.

© 2018 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*, 20, 2438–2455.