Characterization of the first cultured representative of a Bacteroidetes clade specialized on the scavenging of cyanobacteria

Wajdi Ben Hania, Manon Bartoli, Boyke Bunk, Cathrin Sproer, Hans-Peter Klenk, Marie-Laure Fardeau, Stefan Spring

environmental microbiology

Environmental Microbiology (2017) **19**(3), 1134–1148 doi:10.1111/1462-2920.13639

Characterization of the first cultured representative of a *Bacteroidetes* clade specialized on the scavenging of cyanobacteria

Wajdi Ben Hania,[1] Manon Joseph,[1] Boyke Bunk,[2] Cathrin Spröer,[3] Hans-Peter Klenk,[4†] Marie-Laure Fardeau[1] and Stefan Spring[4*]

[1]*Laboratoire de Microbiologie IRD, MIO, Aix Marseille Université, Marseille, France.*

[2]*Department Microbial Ecology and Diversity Research, 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]*Department Microorganisms, Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany.*

Summary

The anaerobic, mesophilic and moderately halophilic strain L21-Spi-D4^T was recently isolated from the suboxic zone of a hypersaline cyanobacterial mat using protein-rich extracts of *Arthrospira* (formerly *Spirulina*) *platensis* as substrate. Phylogenetic analyses based on 16S rRNA genes indicated an affiliation of the novel strain with the *Bacteroidetes* clade MgMjR-022, which is widely distributed and abundant in hypersaline microbial mats and heretofore comprised only sequences of uncultured bacteria. Analyses of the complete genome sequence of strain L21-Spi-D4^T revealed a possible specialization on the degradation of cyanobacterial biomass. Besides genes for enzymes degrading specific cyanobacterial proteins a conspicuous transport complex for the polypeptide cyanophycin could be identified that is homologous to typical polysaccharide utilization loci of *Bacteroidetes*. A distinct and reproducible co-occurrence pattern of environmental 16S rRNA gene sequences of the MgMjR-022 clade and cyanobacteria in the suboxic zone of hypersaline mats points to a specific dependence of members of this clade on decaying cyanobacteria. Based on a comparative analysis of phenotypic, genomic and ecological characteristics we propose to establish the novel taxa *Salinivirga cyanobacteriivorans* gen. nov., sp. nov., represented by the type strain L21-Spi-D4^T, and *Salinivirgaceae* fam. nov., comprising sequences of the MgMjR-022 clade.

Received 9 September, 2016; revised 26 November, 2016; accepted 1 December, 2016. *For correspondence. E-mail ssp@dsmz.de; Tel. (+0049) 531 2616 233; Fax (+0049) 531 2616 418. †Present address: School of Biology, Newcastle University, Newcastle Upon Tyne, UK

Introduction

Photosynthetically active laminated microbial mats represent important model systems for the study of early life on earth and other planets due to a resemblance to distinct fossilized organo-sedimentary structures known as stromatolites, which represent the first record of life on earth and can be traced back to several billion years old rock formations (Grotzinger and Knoll, 1999; Dupraz and Visscher, 2005). In several hypersaline lakes on the Kiritimati Atoll, Central Pacific, thick gelatinous cyanobacterial mats can be found which are associated with large reticulate microbialites situated in the anoxic zone below the mat. It is assumed that the degradation of extracellular polymeric substances in the deep anoxic zone of these mats plays an important role in the observed lithification process (Arp *et al.*, 2012). Members of the *Bacteroidetes* phylum are well-known for their ability to degrade polymeric compounds (Fernández-Gómez *et al.*, 2013) and have been detected in high abundance and diversity in various hypersaline mats by using cultivation-independent methods (Farías *et al.*, 2014; Wong *et al.*, 2016). Therefore, it is likely that *Bacteroidetes* play a key role in the degradation and cycling of mat compounds.

An analysis of the phylogenetic stratigraphy of the Kiritimati hypersaline mat revealed that different clades of *Bacteroidetes* prevail in distinct layers of the mat (Schneider *et al.*, 2013) reflecting a profound niche separation within this group of bacteria. In the upper photic-oxic zone bacteria affiliated with the genera *Salinibacter* and *Salisaeta* were clearly dominating. Representatives of

these genera are extremely halophilic and probably can use light as additional energy source for mixotrophic growth, which explains their high abundance in association with endoevaporitic deposits on the surface of microbial mats (e.g., Sahl *et al*., 2008). Bacteria representing this clade have traditionally been classified within the family *Rhodothermaceae* of the *Bacteroidetes*, but according to a recent proposal both genera should be assigned to the family *Salinibacteraceae* and transferred from the *Bacteroidetes* to the novel sister phylum *Rhodothermaeota* (Munoz *et al*., 2016). On the other hand, in the deeper layers of the Kiritimati hypersaline mat phylotypes related to members of the *Marinilabiaceae* became dominating and reached proportions of up to 25% of the total number of retrieved bacterial 16S rRNA gene sequences. The family *Marinilabiaceae* and related phylogenetic groups were recently transferred to the order *Marinilabiliales* (Wu *et al*., 2016), which represents mainly facultatively aerobic or anaerobic bacteria with a fermentative metabolism that have been frequently isolated from marine and hypersaline anoxic sediments (e.g., Shalley *et al*., 2013; Iino *et al*., 2015; Wu *et al*., 2016).

In a previous study extracts of the cyanobacterium *Arthrospira* (formerly *Spirulina*) *platensis* containing around 60% single-cell protein were used as substrate to isolate strain L21-Spi-D4^T representing phylotypes affiliated with a novel family-level clade of the class *Bacteroidia* that have been frequently detected in the suboxic zone of hypersaline mats (Spring *et al*., 2015). In this study, we present a detailed analysis of the spatial distribution and characteristics of the novel strain, which revealed specific adaptations to the degradation of cyanobacterial biomass.

Results and discussion

Phylogenetic placement within the Bacteroidia

Phylogenetic trees based on almost complete 16S rRNA gene sequences revealed an affiliation of strain L21-Spi-D4^T with the MgMjR-022 sequence cluster as defined by the SILVA 123 SSU Ref NR99 data set (Fig. 1). Strain L21-Spi-D4^T represents the first cultured isolate of this clade, which so far is comprised almost exclusively of sequences retrieved from hypersaline cyanobacterial mats.

The diversity encountered within the MgMjR-022 clade was determined based on a pairwise sequence analysis of 181 non-redundant high-quality 16S rRNA sequences available from the SILVA 123 SSU Ref NR99 data set. The average sequence identity value within this group was 91.81% and the minimum sequence identity 79.23%. In a recent study sequence identity values within established taxa of different ranks were analyzed (Yarza *et al*., 2014). Based on the analyzed data set a median sequence identity value of 92.25% and a minimum sequence identity value of 87.65% were deduced for a family. Consequently, the observed phylogenetic divergence within the MgMjR-022 clade matches most closely the taxonomic rank of a family. The closest related cultured strains were found within the *Marinilabiliaceae*. Based on a comparison with 16S rRNA sequences of type strains the most closely related type species were *Marinilabilia salmonicolor* NBRC 15948^T (86.2%), *Anaerophaga thermohalophila* Fru22^T (85.8%), *Alkaliflexus imshenetskii* Z-7010^T (85.8%), *Alkalitalea saponilacus* SC/BZ-SP2^T (85.6%) and *Natronoflexus pectinivorans* AP1^T (85.6). These 16S rRNA sequence identity values were below the threshold of 86.5% proposed for assigning sequences to a family (Yarza *et al*., 2014), therefore confirming the placement of strain L21-Spi-D4^T in a distinct novel family within the order *Marinilabiliales*.

Phenotypic characteristics

Morphology. Cells of strain L21-Spi-D4^T were Gram-negative straight to slightly curved slender rods, often with a hook at one end. They had an average size of 0.4–0.5 μm × 5–8 μm, were non motile and occurred single or in pairs. An active bending or flexing of cells as described for most representatives of the *Marinilabiliaceae* was not detected. Formation of cell aggregates was occasionally observed in liquid media, but no visible colonies were formed on solid agar media, so that gliding motility on surfaces could not be determined. In older cultures bulged cells and spherical bodies were found (Fig. 2A and B). Thin cuts viewed by transmission electron microscopy revealed a cell wall structure typical of Gram-negative bacteria comprising an outer membrane, periplasmic space and cytoplasmic membrane. Occasionally, protrusions of the outer membrane could be detected (Fig. 2C and D). Pigments were not formed and no flagella, endospores or other intracellular inclusion bodies became apparent by microscopy.

Chemotaxonomy. The cellular fatty acid composition of strain L21-Spi-D4^T was characterized by large amounts of iso-branched fatty acids, especially *iso*-$C_{15:0}$, which made up more than half of the total amount (Table S1). A clearly distinguishing trait to the patterns of type strains of the most closely related type species was a very high amount of the hydroxy fatty acid *iso*-$C_{17:0}$ 3OH (18.1%). The polar lipid composition of this strain was complex comprising phosphatidylethanolamine in addition to several unidentified polar lipids, including aminolipids and glycolipids (Supporting Information Fig. S1). Cytochromes of *c*-type were not detectable in redox difference spectra of solubilized membrane suspensions and tests for oxidase and catalase were negative. These phenotypic traits point to a strictly fermentative metabolism lacking a respiratory electron transport chain. However, the respiratory lipoquinone menaquinone 7 (MK7) could be detected in this strain,

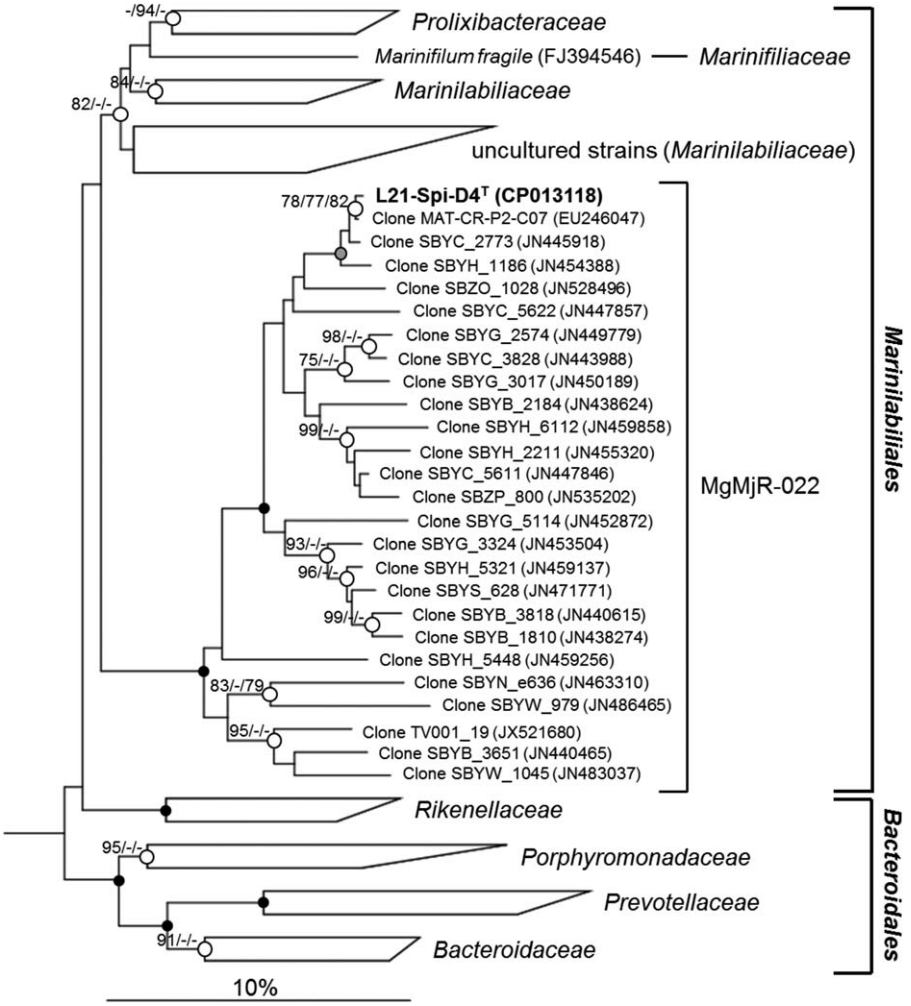

Fig. 1. Phylogenetic placement of strain L21-Spi-D4^T within the class *Bacteroidia* based on almost complete 16S rRNA gene sequences. The tree topology was reconstructed using the neighbor-joining algorithm with the correction of Jukes-Cantor and rooted using the 16S rRNA gene sequence of *Cytophaga hutchinsonii* (CP000383, not shown). 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 gray 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 neighbor-joining, maximum-likelihood and maximum parsimony method. Polygons represent clades of several sequences. Scale bar, 10% estimated sequence divergence.

which is an indication for the presence of membrane-bound electron transfer complexes.

Physiology of growth. Strain L21-Spi-D4^T was moderately halophilic and required at least 50 g l^{-1} NaCl for growth. Optimal conditions for growth were at 9% (w/v) NaCl, a pH of around 7.0 and a temperature of 35°C. Proliferation of this strain depended strictly on oligopeptides supplied as Trypticase, Tryptone peptone or yeast extract, which allowed growth in the absence of additional substrates. In contrast, the synthetic dipeptide alanyl-glutamine (Sigma G8541) or single amino acids supplied as Casamino acids were not utilized, whereas complex proteins (gelatine and serum albumin) were only degraded in the presence of an additional substrate like glucose or yeast extract. Vitamins did not prevent the requirement for peptides and were not necessary for growth. A dependence of growth on oligopeptides was previously also observed in *Porphyromonas gingivalis*, an anaerobic and fastidious representative of the *Bacteroidales* colonizing human gingival pockets (Takahashi *et al.*, 2000). A requirement for peptides could therefore represent a wide spread characteristic of members of this phylogenetic group, independent of the environment from which they were isolated. In the presence of yeast extract or peptone strain L21-Spi-D4^T could use several carbohydrates and pyruvate as substrates for growth. The minimum amount of yeast extract required for growth on glucose was 0.5 g l^{-1}. Fermentation products from glucose were acetate, CO_2 and H_2 (Supporting Information Fig. S2). Upon growth with yeast extract or peptone as sole substrate acetate, succinate, propionate, isovalerate and isobutyrate were produced in an estimated ratio of around 2:1:1:1:0.5 (Supporting Information Fig. S3). Isovalerate and isobutyrate result only from the degradation of branched-chain amino acids (Barker, 1981), which were therefore preferentially used during fermentation.

A summary of all determined phenotypic traits of the novel strain is reported in the formal species description at the end of this paper. A list of characteristics useful for the differentiation of strain L21-Spi-D4^T from type strains of related type species belonging to the family *Marinilabiliaceae* is given in Table S2.

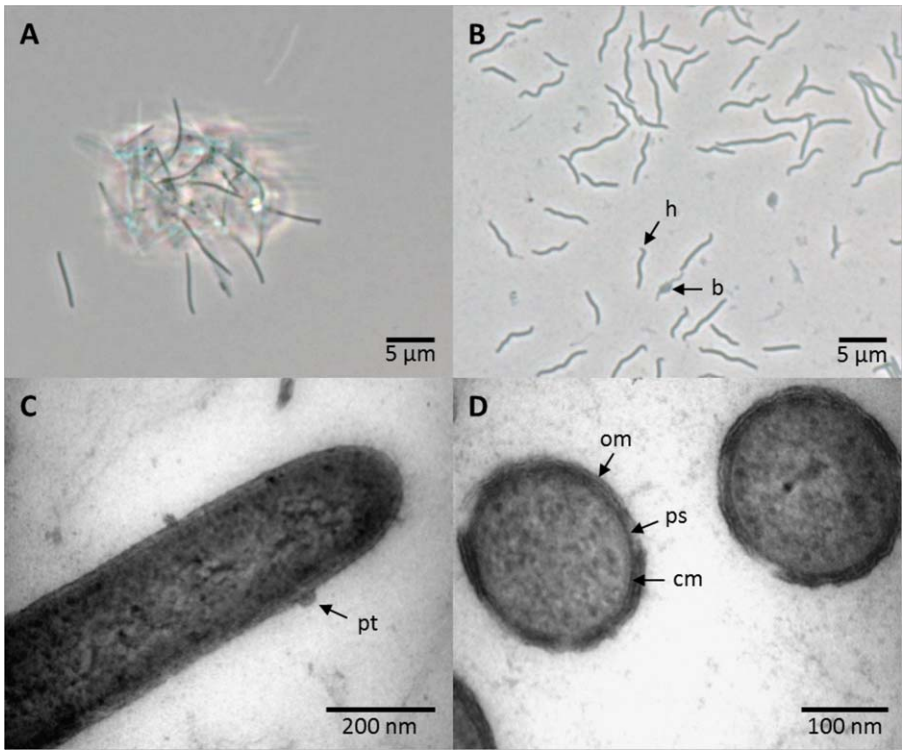

Fig. 2. Shape and ultrastructure of cells of strain L21-Spi-D4^T.
A. Phase contrast micrograph of an aggregate of cells formed upon cultivation in liquid medium.
B. Curved morphology of cells in stationary phase viewed by phase contrast microscopy after squeezing onto an agar-coated slide. Arrows mark hooked ends (h) and bulging of cells (b).
C. Transmission electron micrograph of a negatively stained longitudinal thin section. A typical protrusion (pt) of the outer membrane is marked by an arrow.
D. Electron micrograph of a negatively stained cross-section revealing the characteristics of a Gram-negative cell envelope, which are an outer membrane (om), a periplasmic space (ps) and a cytoplasmic membrane (cm). [Colour figure can be viewed at wileyonlinelibrary.com]

Insights from the genome sequence

General features of the genome. The genome of strain L21-Spi-D4^T is represented by a circular chromosome with a size of 4.81 Mb and a G + C content of 39.0 mol%. In total 3943 genes were predicted, including 3884 protein-coding genes of which 71.5% were assigned a putative function by the IMG genome annotation pipeline (Huntemann *et al.*, 2015). In addition, 59 RNA genes were detected comprising three complete operons for ribosomal RNA (16S-tRNA-23S-5S). A high proportion of protein-coding genes were assigned to the COG functional category amino acid transport and metabolism (172), whereas genes involved in the transport and metabolism of carbohydrates (85) and lipids (82) were less abundant, which points to a specialization on protein-rich substrates. In addition, a large number of genes were found to be involved in signal transduction mechanisms (178) thereby indicating a high capacity of this strain to respond to environmental stimuli. According to the IMG/ER database around 6.9% of all genes of the L21-Spi-D4^T genome could be connected to a transporter classification. Transporters of the ATP-binding cassette (ABC) superfamily (TC:3.A.1) that usually catalyses the uptake of low molecular weight substrates were the most abundant type and comprised 64 genes.

The stability of the genome structure in this strain may be influenced by a high prevalence of transposase genes, which account for around 3.0% of all protein-coding genes, while in related representatives of the *Marinilabiliaceae* this proportion is only 0.7% (average of four genomes analyzed). The high number of transposase genes could indicate a relaxed genome structure resulting from a recent genetic isolation that has been caused by the adaptation to the anoxic zone of hypersaline mats. The presence of 51 pseudogenes having no function may also reflect the lack of a stringent selective pressure to retain genome integrity. Defence mechanisms against a modification of the genome structure by foreign DNA include several types of restriction-modification systems (Roberts *et al.*, 2015) and four genomic regions encoding clustered, regularly interspaced short palindromic repeats (CRISPR) along with several associated *cas* genes.

A graphic summary of the genome structure of strain L21-Spi-D4^T is presented in Supporting Information Fig. S4.

Decomposition of proteins and polysaccharides. Most members of the *Bacteroidetes* are specialized on the degradation of complex substrates. It was found that marine members of this phylum have a distinct preference for the degradation of peptides and proteins, while most non-marine representatives are more specialized on the utilization of carbohydrates (Fernández-Gómez *et al.*, 2013). Strain L21-Spi-D4^T is able to utilize proteins and polysaccharides for growth and its genome encodes a wide array of peptidases and glycoside hydrolases for the depolymerization of complex substrates. A BLAST search of the L21-Spi-D4^T genome using the MEROPS database returned 180 sequences that could be assigned to peptidase

families (around 37 genes per Mb), which illustrates a remarkable capacity of this strain to degrade proteins and peptides. Serine (S) peptidases were the most abundant type, followed by metallo (M) and cysteine (C) peptidases (Table S3). Many genes were assigned to families known to contain extracellular proteases like papain (C01A), thermitase (S08A) or tricorn core protease (S41B). The highest number of sequences was assigned to the S33 peptidase family, but the identified genes seem to encode mainly enzymes related to hydrolases with a function in the degradation of aromatic compounds (e.g., a putative 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase, L21SP5_03418) or the hydrolytic cleavage of amino acid esters (e.g., a putative valacyclovir hydrolase, L21SP5_03297). Such enzymes could play a role in the removal of chemical modifications from proteins to make them more accessible for the degradation by proteases, which is probably also the function of a putative phycocyanobilin lyase (CpcT1, L21SP5_02599). This enzyme was originally described in cyanobacteria and catalyses the site-selective attachment of phycocyanobilin to the apoproteins phycocyanin or phycoerythrocyanin (Zhao *et al*., 2007), while in heterotrophic bacteria it more likely plays a role in the removal of the chromophore from the phycobiliprotein to enable proteolytic digestion. In addition to the biliprotein lyase, several further genes were detected that encode peptidases, which play a role in the specific degradation of cyanobacterial proteins, including two distinct cyanophycinases (L21SP5_00867, L21SP5_00872) and a microcystinase (L21SP5_03670).

The genome of strain L21-Spi-D4^T encodes also a multitude of carbohydrate active enzymes that participate in the degradation of polysaccharides, including 33 distinct glycoside hydrolases (Table S4). Numerous genes were assigned to the glycoside hydrolase (GH) families GH13 and GH16, which comprise mainly enzymes involved in the degradation of polysaccharides composed of α (1→4) and β (1→3) linked glucose monomers respectively. Typical substrates of these enzymes could be bacterial glycogen (α-1,4- and α-1,6-glucan) or chrysolaminarin (β-1,3-glucan), which represent common reserve polymers in cyanobacteria and algae. Interestingly, four genes were allocated to the GH23 family, which is represented by lytic enzymes involved in the degradation of peptidoglycan cell walls. In addition to glycoside hydrolases, five genes with a putative carbohydrate esterase function and two genes encoding sulfatases were detected (L21SP5_01883, L21SP5_02800). Both sulfatases are membrane-bound and probably involved in the digestion of cell walls. The low number of sulfatases in the genome of strain L21-Spi-D4^T indicates that extracellular polymeric substances (EPS) most likely do not represent targeted substrates of this strain. Marine and hypersaline bacteria specialized on the degradation of exopolymers usually require a large number of sulfatases for the specific removal of sulfate residues from polysaccharides (e.g., Spring *et al*., 2016), which are usually sulfated at various sites to make them more recalcitrant against enzymatic degradation.

Substrate uptake. Bacteroidetes have developed specific transport systems for the concomitant degradation and uptake of polysaccharides, presumably to prevent the release of monomeric sugars into the environment where they eventually could be consumed by competing bacteria. Genomic regions that contain genes of these membrane-anchored uptake complexes are designated Polysaccharide Utilization Loci (PUL) (Martens *et al*., 2009) and were found to be widely distributed among representatives of the *Bacteroidetes* (Terrapon *et al*., 2015). The characteristic feature of PULs are the genes *susC* and *susD*, which are always arranged in tandem and encode a TonB dependent transporter and a substrate binding protein attached to the outer membrane respectively. In strain L21-Spi-D4^T two true PULs and a PUL-like gene cluster could be identified.

The most comprehensive locus is located at L21 SP5_02764–02776 and contains thirteen genes. This cluster is obviously involved in the uptake and depolymerization of α-glucans, because it contains five GH13 genes, along with a GH65 and GH31 gene. On the opposite DNA strand, adjacent to the *susCD* genes, a transcriptional regulator of the LacI family (L21SP5_02775) and a transporter of the major facilitator superfamily (MFS) with a putative specificity for maltose (L21SP5_02776) were located. According to the Polysaccharide-Utilization Loci DataBase (Terrapon *et al*., 2015) the observed combination of glycoside hydrolase genes is so far unique among genome-sequenced members of the *Bacteroidetes*. However, a similar PUL comprising three GH13 genes, one GH65 gene and several glycolysis genes has been recently identified in "*Gramella forsetii*" KT0803, where it could be involved in the degradation of dextrins (Kabisch *et al*., 2014). The second PUL is located at L21SP5_02816–02823 and characterized by two GH16 and one GH3 gene, which points to a participation in the degradation and uptake of β-glucans. A PUL with the same pattern of glycoside hydrolase genes was found for instance in "*Gramella forsetii*" KT0803 (Fig. 3A) and was shown to be involved in the utilization of laminarin (Kabisch *et al*., 2014). In contrast to the laminarin utilization locus of "*Gramella forsetii*" KT0803 the corresponding PUL of strain L21-Spi-D4^T contains additionally a gene for a MFS glucose transporter (L21SP5_02818) and two genes for outer membrane proteins containing the carbohydrate-binding modules CBM6 (L21SP5_02820) and CBM9 (L21SP5_02821) instead of the commonly found proteins characterized by a PKD domain. In the model proposed in Fig. 3A it is suggested that the CBM proteins and the GH16 endoglucanases build a conveyor belt-like structure that degrades β-glucans and

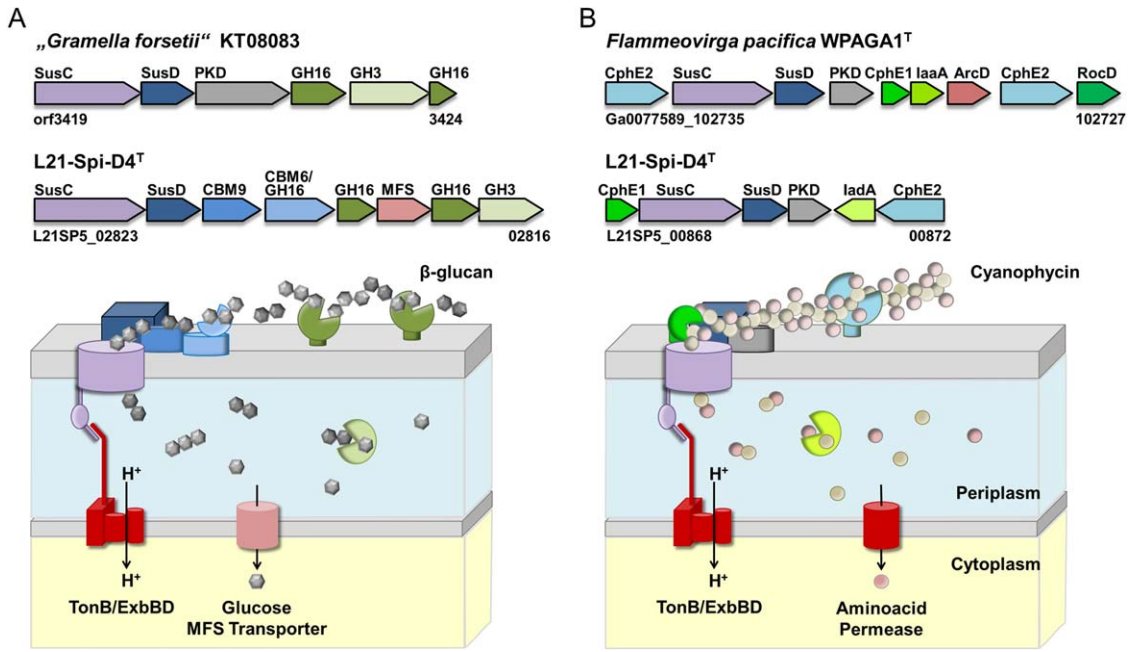

Fig. 3. PUL-like genomic regions for the utilization of β-glucans and cyanophycin. Colors of genes and proteins correspond with their proposed functions: lilac, SusC-like transporters; gray, PKD-domain containing proteins; blue, proteins involved in substrate binding; green, hydrolytic enzymes; red, proteins involved in transportation of substrates to the cytoplasm.
A. Comparison of the β-glucan utilization locus of strain L21-Spi-D4^T with the laminarin-utilization locus of "*G. forsetii*" (*Flavobacteriia*) and proposed model of the uptake complex in strain L21-Spi-D4^T.
B. Gene arrangement of the cyanophycin utilization locus in strain L21-Spi-D4^T compared with a similar region detected in *F. pacifica* (*Cytophagia*) and proposed model of the uptake complex in strain L21-Spi-D4^T.

concomitantly supplies the resulting fragments to the TonB-dependent transporter.

A further locus characterized by the *susCD* tandem is of significant interest, because it is not linked to glycoside hydrolase genes and therefore not involved in the decomposition and uptake of polysaccharides. This genomic region comprises a total of six genes (L21SP5_00868–00872) and features two genes encoding extracellular cyanophycinases (CphE1 and CphE2) and an *iadA* gene for the enzyme isoaspartyl dipeptidase (Fig. 3B). Cyanophycinase is a serine-type exopeptidase hydrolysing cyanophycin to β-aspartyl-arginine dipeptides, which can be hydrolysed to arginine and aspartate by isoaspartyl dipeptidase. Consequently, it is likely that this gene cluster represents a Cyanophycin Utilization Locus (CUL). Cyanophycin is a common and abundant reserve polymer in cyanobacteria, where it is mainly used for the storage and transport of organically-bound nitrogen that has been assimilated by the fixation of dinitrogen (Burnat *et al.*, 2014). Thus, an efficient uptake and utilization of this valuable substrate could represent an important ecological advantage for bacteria feeding on cyanobacteria.

To exclude the possibility of a coincidental colocation of cyanophycinase and *susCD* genes in the L21-Spi-D4^T genome, we tried to detect similar gene clusters in genomes of other bacteria by performing a BLAST search using the SusC protein of L21-Spi-D4^T (L21SP5_00869) as query and the IMG collection of genomes as database. In total twelve distinct gene clusters could be detected, which were present in phylogenetically diverse species of *Bacteroidetes* (Table S5), so that an incidental linkage of *cphE* and *susCD* genes can be excluded. All retrieved TonB-dependent transporters (SusC/RagA-like proteins) that colocated with cyanophycinase genes were closely related and had sequence identities above 45% to the query protein of strain L21-Spi-D4^T. Accordingly, in phylogenetic trees of SusC/RagA proteins of diverse *Bacteroidetes* species TonB-dependent transporters of CULs formed a monophyletic lineage, which was only distantly related to groups of SusC proteins involved in the utilization of β-glucans or α-glucans (Supporting Information Fig. S5). Interestingly, a TonB-dependent transporter involved presumably in the uptake of large protein fragments was previously identified in *P. gingivalis* (Nagano *et al.*, 2007), thereby indicating a wide-spread occurrence of peptide uptake complexes characterized by the SusCD tandem. Except one case all putative CULs contained two cyanophycinase genes. One gene designated as *cpheE1* was always present and encodes a canonical single domain cyanophycinase. The other gene (*cphE2*), which was

missing in the genome of *Paludibacter propionicigenes* DSM 17365^T (Gronow *et al.*, 2011), contains an additional C-terminal region that seems to have a function in the transport and anchoring of the protein to the outer membrane (in most cases a Por secretion system domain). Interestingly, in several cases the cyanophycinase domain of the *cphE2* gene was highly divergent to typical cyanophycinases, which could indicate a relaxed selection pressure on the enzymatic function of the CphE2 protein. In the hypothetical model presented in Fig. 3B, we therefore suggest that the main function of CphE2 is the binding of cyanophycin, which is then supplied to the functional cyanophycinase and the TonB-dependent transporter. Based on this hypothesis it is possible that in *P. propionicigenes* the CphE2 protein was replaced by a lipoprotein containing an N-terminal dipeptidylpeptidase IV domain (Palpr_0472), which is encoded adjacently to the *susD* gene.

The most comprehensive putative CUL comprising nine genes was detected in the genome of *Flammeovirga pacifica* WPAGA1^T (Chan *et al.*, 2015). In addition to three cyanophycinases genes the gene cluster shown in Fig. 3B comprises an isoaspartyl-peptidase (*iaaA*), an arginine/ornithine antiporter (*arcD*) and an ornithine aminotransferase (*rocD*). Based on the genome annotation it can be assumed that in this species L-arginine is degraded via the deiminase pathway, which essentially results in L-ornithine, ATP and NH_3. Therefore, the arginine/ornithine antiporter has an important function in preventing the inhibition of the initial enzyme of the pathway (arginine deiminase) by an accumulation of ornithine.

For the uptake of peptides other than cyanophycin, several transporter systems were identified. Genes of two putative ABC-type peptide transporters were detected, which were not arranged in operons, but dispersed across the genome. Genes for the peptide-binding proteins were located at L21SP5_01631 and L21SP5_02194, while the corresponding permease genes were found at L21SP5_00031 and L21SP5_01704. Interestingly, the ATP binding proteins of both putative oligopeptide ABC transporters were not encoded separately but fused to a single gene (L21SP5_02394). In addition, a proton-dependent transporter of the major facilitator superfamily (L21SP5_02533) and the membrane protein CstA (L21SP5_00639) could be involved in the uptake of peptides. Several proteins were identified that could catalyse the transport of amino acids. A putative sodium:alanine symporter is encoded at L21SP5_03896, while the arginine/ornithine antiporter ArcD is located at L21SP5_03773. Two other genes (L21SP5_02519 and L21SP5_03922), which contain amino acid permease domains could be involved in the unspecific uptake of amino acids.

Fermentative pathways and energy metabolism. Based on the combined analyses of fermentation products and genome sequence the major catabolic pathways in strain L21-Spi-D4^T could be reconstructed. It is proposed that glucose is degraded to pyruvate via the Embden–Meyerhof–Parnas (EMP) pathway (Supporting Information Fig. S6). Remarkably, in the presence of intracellular polyphosphates (synthesized by polyphosphate kinase) and pyrophosphate (formed as byproduct of macromolecule synthesis or by adenylate kinase) ATP independent glycolysis reactions involving the enzymes polyphosphate glucokinase, pyrophosphate–fructose 6-phosphate 1-phosphotransferase and pyruvate, phosphate dikinase could enable the production of 6 mol ATP from 1 mol glucose compared with 2 mol ATP in the canonical EMP pathway. Due to the requirement for complex substrates it could not be determined which particular amino acids were actually used for growth, but the production of acetate, succinate, propionate, isovalerate and isobutyrate may indicate that alanine, asparagine, arginine, histidine, threonine, proline, glutamate, aspartate, leucine and valine could be preferably degraded using pathways involving an oxidative decarboxylation step at 2-oxoacid:ferredoxin oxidoreductases (Supporting Information Fig. S7).

Oxygen had no stimulatory effect on the growth of strain L21-Spi-D4^T and no significant expression of terminal cytochrome *c* oxidases could be detected in laboratory experiments. Therefore, it was unexpected that the genome encodes a complete set of enzymes for the oxidative citrate cycle, a respiratory complex I and several genes encoding subunits of a terminal cytochrome *d* quinol oxidase (L21SP5_01809–01810) and a putative caa_3-type cytochrome *c* oxidase (L21SP5_03574–03578). However, it turned out that genes for a respiratory complex III (cytochrome bc_1 or alternative complex III) were missing or fragmentary, thereby indicating that this complex maybe non-functional. This could indicate a recent degeneration of an aerobic respiratory electron transport chain, perhaps as a result of the dependence on decaying cyanobacteria, which are mainly found in the permanently anoxic niches of microbial mats (Arp *et al.*, 2012; Schneider *et al.*, 2013).

The strictly fermentative metabolism of strain L21-Spi-D4^T allows the generation of energy in the form of ATP by substrate-level phosphorylation involving several carboxylate kinases. Alternatively, a chemiosmotic potential could be utilized for the synthesis of ATP. The methylmalonyl-CoA decarboxylase (L21SP5_03928–03931) and a RNF-type NAD:ferredoxin oxidoreductase complex (L21SP5_03909–03914) may generate a sodium gradient, while the quinol:fumarate oxidoreductase (L21SP5_01812–01814) is assumed to translocate protons (Cecchini *et al.*, 2003). In addition to these complexes, which are typically found in fermentative Gram-negative bacteria, genes of a putative energy-conserving group 4 [NiFe]-hydrogenase could be identified, which were embedded in a large tentative operon (L21SP5_01248–01263) adjacently to a Mrp antiporter and

a NADPH input module. This membrane-bound complex closely resembled modular hydrogenases of the Mrp-Mbh-type, which are mainly present in hyperthermophilic fermentative Archaea and rarely encountered in bacteria (Spring *et al*., 2010; Schut *et al*., 2013). The gene cluster detected in strain L21-Spi-D4^T was almost identical to the locus encoding the Mrp-Mbh-NAD(P)H complex of the thermophilic bacterium *Kosmotoga olearia* (Kole_0561–0574), which was proposed to use an electrochemical gradient to drive the endergonic hydrogen production from NAD(P)H (Schut *et al*., 2013). In combination with a RNF-type electron transport complex the reducing equivalents generated by carbohydrate oxidation could be disposed of as hydrogen without the requirement of a bifurcating cytosolic hydrogenase, which is lacking in strain L21-Spi-D4^T and *Kosmotoga* species. The presence of a sodium translocating NADH:quinone oxidoreductase (L21SP5_03813–03818) representing a tentative respiratory complex I (Verkhovsky and Bogachev, 2010) could indicate that sodium ions are preferred against protons for the generation of an electrochemical potential in strain L21-Spi-D4^T. We propose that the sodium motive force generated during fermentation is utilized by a canonical F_oF_1 ATP synthase (L21SP5_02006–02007/02011–02016), while an alternative V-type ATPase (L21SP5_02781–02787) and a membrane-bound pyrophosphatase (L21SP5_02602) could provide pH homeostasis during fermentation by outward pumping of protons. In turn, a separate Mrp cation-proton antiporter (L21SP5_02298–02304) could use the proton gradient to maintain sodium resistance. In addition, a Mrp-Mbx-NAD(P)H complex (L21Sp5_02700–02711) with unclear function was identified, which is very similar to the corresponding complex in *K. olearia* (Kole_2051–2063). Based on the determined modular structure this complex could represent a reversible membrane-bound transhydrogenase that is coupled to an ion gradient.

A summary of the potential roles of the various identified membrane-bound enzyme complexes in the generation of metabolically useful energy during fermentative growth of strain L21-Spi-D4^T is shown in the hypothetical model presented in Fig. 4.

Distribution profiles and co-occurrence patterns in hypersaline mats

The environmental distribution of members of the MgMjR-022 clade was deduced from the sequence associated information provided by the SILVA 123 SSU Ref data set. It turned out that from a total of 1230 16S rRNA gene sequences assigned to this clade only thirteen sequences were not retrieved from hypersaline mats, which indicates a pronounced adaptation of members of this clade to photosynthetically active microbial mats typically found in hypersaline environments.

The vertical distribution patterns of strain L21-Spi-D4^T and related bacteria of the MgMjR-022 clade in the Kiritimati hypersaline mat could be revealed using data of a previous study, which used large-scale sequencing of 16S rRNA genes to determine the phylogenetic stratigraphy of this laminated microbial mat (Schneider *et al*., 2013). Although, in general it is not feasible to correlate rRNA gene frequency directly with cellular abundance, the counts of 16S rRNA sequences in distinct mat layers allowed some conclusions about the relative abundance and vertical distribution of distinct phylotypes in this ecosystem. In Fig. 5A, the distribution patterns of prevalent *Bacteroidetes* clades in the Kiritimati Lake 21 mat were compared with the vertical profile of the most abundant clade of cyanobacteria, which represent the principle primary producers in hypersaline mats. As expected the highest abundance of cyanobacteria affiliated with the *Halothece* cluster (Garcia-Pichel *et al*., 1998) was found in the photic-oxic zone, which corresponds to the upper three layers of the mat (Schneider *et al*., 2013; Ionescu *et al*., 2015). Below the photic-oxic zone in around 2.5 cm depth the MgMjR-022 clade reaches a first maximum in layer 4, followed by a slight decrease in layer 5 and a second maximum in layer 6. Sequences belonging to the L21-Spi-D4^T species-level clade showed the same distribution pattern as the MgMjR-022 clade and dominated this family-level group by representing around 75% of all assigned phylotypes. It is assumed that layer 5 of the studied Lake 21 microbial mat represents the surface layer of a former cyanobacterial mat, which has been overgrown by the current photosynthetically active mat (Schneider *et al*., 2013; Spring *et al*., 2015). Although 16S rRNA genes of cyanobacteria were not retrieved in significant numbers from layers below the photic-oxic zone, the remains of cyanobacterial cells could be identified by microscopy in the deep layers of this mat (Schneider *et al*., 2013). This discrepancy can be explained by an ongoing degradation of DNA, which seems to be more sensitive to decomposition than other structural compounds of dead cyanobacterial cells. Therefore, it is likely that the distribution pattern of the MgMjR-022 clade reflects the presence of decaying cyanobacterial biomass in the mat layers 4–6. The co-occurrence with inactive or decaying cyanobacterial cells appears to be a hallmark trait of the MgMjR-022 group, because the distribution patterns of other prevalent clades of the *Bacteroidetes* were clearly different and hence likely caused by other environmental factors. For instance, members of the *Salinibacteraceae* are known to express rhodopsins and are probably able to use light energy for aerobic mixotrophic growth (Boichenko *et al*., 2006), which could explain their restricted distribution in the photic-oxic zone of the mat. On the other hand, representatives of the *Saprospiraceae* are frequent commensals of algae. They either actively attack and lyse proliferating cells or use

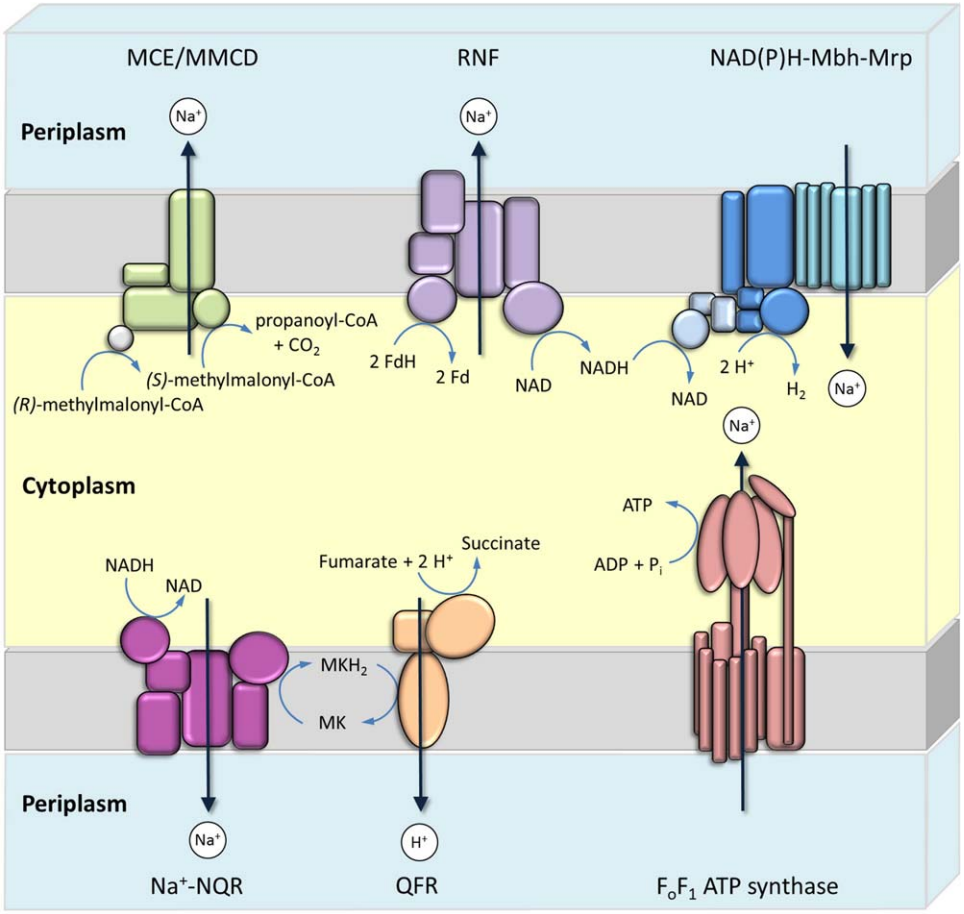

Fig. 4. Proposed participation of membrane-bound enzyme complexes of strain L21-Spi-D4^T in the regeneration of reducing equivalents and the utilization of a sodium motive force for the synthesis of ATP during fermentative growth. Abbreviations: MCE/MMCD, methylmalonyl-CoA epimerase/decarboxylase; RNF, NAD:ferredoxin oxidoreductase complex; NAD(P)H-Mbh-Mrp, membrane-bound energy-conserving hydrogenase with NAD(P)H input module; Na^+-NQR, sodium-translocating NADH:quinone oxidoreductase; QFR, quinol:fumarate oxidoreductase; Fd, oxidized ferredoxin; FdH, reduced ferredoxin; MK, oxidized menaquinone; MKH_2, reduced menaquinone.

algae-derived metabolites as substrates for aerobic chemoheterotrophic growth (Shi *et al.*, 2006; McIlroy and Nielsen, 2014), which is illustrated by their co-occurrence with photosynthetically active cyanobacteria in the upper layers of the mat. In contrast, *Marinilabiliaceae* were found mainly in the deep anoxic zones of the mat, where probably most of the cellular material of cyanobacteria has already been degraded and recalcitrant compounds of the mat matrix represent the main substrates for growth.

The postulated dependence of members of the MgMjr-022 clade on decaying biomass of cyanobacteria was evaluated by the analysis of sequence data obtained previously from the mat of a solar saltern evaporation pond located in Guerrero Negro, Mexico (Harris *et al.*, 2013). This non-lithifying hypersaline mat is characterized by numerous dense layers resulting in a compact mat structure (Fig. 5B), which is clearly different to the gelatinous, partly fluffy, structure typical of the mineralizing mats in Lake 21. The divergent textures of both mat systems are probably caused by geochemically differences between both locations and also reflected in distinguishable microbial community compositions (Schneider *et al.*, 2013). For instance, the dominating clade of cyanobacteria in the Guerrero Negro mat is represented by *Coleofasciculus* (formerly *Microculeus*) species (Siegesmund *et al.*, 2008), which are less halotolerant than members of the *Halothece* cluster found in Kiritimati Lake 21. Despite the observed differences, the distribution pattern of the MgMjR-022 clade in the Guerrero Negro mat displays a similar dependence on inactive cyanobacterial biomass as in the Lake 21 mat. As shown in Fig. 5B sequences representing the MgMjR-022 clade are most frequently found below the layer of active cyanobacterial growth, which seems to be restricted to a narrow oxic zone at the mat surface (0–2 mm depth) (Ley *et al.*, 2006). Above all, the correlation of the vertical distribution of MgMjR-022 clade sequences with the relative abundance of *Coleofasciculus* spp. was not only restricted to the oxic-anoxic transition zone, but also detected in deeper anoxic layers of the mat. However, the low portion of sequences affiliated with the L21-Spi-D4^T phylotype (around 22%) on the total number of sequences assigned to the MgMjR-022 clade represented an interesting deviation to the observed community structure of the Lake 21 mat. One possible reason could be that diversification in the MgMjR-022 clade is mainly driven by an adaptation to the dominating mat-forming cyanobacterial species. Thus,

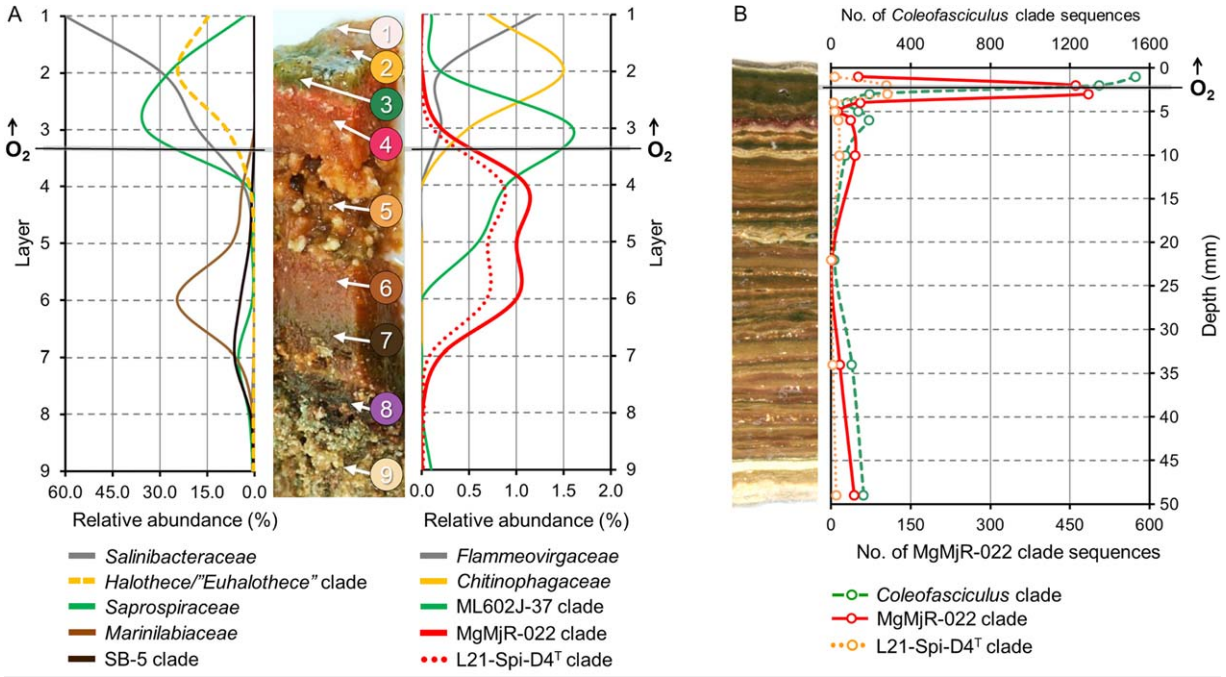

Fig. 5. Spatial distribution of prevalent *Bacteroidetes* clades, *Salinibacteraceae* and the principal mat-forming cyanobacteria in hypersaline microbial mats. Family-level clades are represented by continuous lines, genus-level clades are represented by dashed lines and dotted lines represent a species-level clade comprising strain L21-Spi-D4^T. The transition to anoxic conditions within the mat samples is indicated by a gray line and located during daylight at a depth of around 2.5 cm (within layer 4) in the Kiritimati Lake 21 mat and at 2 mm depth in the Guerrero Negro mat.

A. Distribution patterns based on the proportion of partial 16S rRNA gene sequences in distinct layers of the Kiritimati Lake 21 lithifying microbial mat. On the left, highly abundant *Bacteroidetes* clades (representing at least 5% of the total number of sequences in a distinct layer) and the dominating cyanobacterial clade comprising *Halothece* ("*Euhalothece*") spp. are shown. On the right, the distribution patterns of moderately abundant *Bacteroidetes* clades (representing at least 1% of the total number of sequences in a distinct layer) are shown. Between both graphs a representative section of the Lake 21 microbial mat is depicted. Numbers indicate distinct mat layers used for the generation of the corresponding 16S rRNA gene sequence libraries by high-throughput sequencing. Modified from Schneider *et al.* (2013).

B. Distribution patterns in a microbial mat of a saltern evaporation pond (Exportadora de Sal, Guerrero Negro, Mexico). Frequencies of cloned 16S rRNA gene sequences affiliated with the MgMjR-022 cluster and the L21-Spi-D4^T clade are compared with the dominating cyanobacterial clade comprising *Coleofasciculus* spp. On the left a representative section of the Guerrero Negro microbial mat is shown. The photograph has been taken from Harris *et al.* (2013), with permission.

in the Guerrero Negro mat strains related to L21-Spi-D4^T may have been superseded by a phylotype that is better adapted to the degradation of *Coleofasciculus* cells.

Conclusions and classification

We propose a specialization of strain L21-Spi-D4^T and related bacteria on the utilization of decaying cyanobacterial biomass based on an environmental distribution that is mainly restricted to the anoxic zone of photosynthetically active microbial mats, distinct co-occurrence patterns of members of the MgMjR-022 clade and cyanobacteria as well as several characteristics of the genome sequence of strain L21-Spi-D4^T. The observed strictly anaerobic fermentative metabolism of this strain may have evolved from a facultative aerobic ancestor to avoid competition with highly abundant aerobic commensals of active cyanobacteria (e.g., members of the *Saprospiraceae*) in the upper layers of microbial mats. A differentiation of the novel strain from cultured members of the closely related family *Marinilabiliaceae* is possible due to chemotaxonomic traits and the absence of cellular flexibility and pigmentation. Furthermore, carbohydrate fermentation in strain L21-Spi-D4^T resulted only in the formation of acetate, H_2 and CO_2, whereas members of the *Marinilabiliaceae* perform a mixed-acid fermentation. Consequently, we propose to establish a novel family within the order *Marinilabiliales* comprising mainly cloned 16S rRNA gene sequences of the MgMjR-022 clade and *Salinivirga cyanobacteriivorans* gen. nov., sp. nov., as type species represented by the type strain L21-Spi-D4^T. Despite a low bootstrap support in reconstructed phylogenetic trees placement of the novel family *Salinivirgaceae* within the order *Marinilabiliales* seems to be justified based on the close relationship to members of the *Marinilabiliaceae*, presence of MK-7 as major respiratory quinone and a prevalence in marine to

hypersaline sediments or mats, which distinguishes this family clearly from representatives of the order *Bacteroidales* as defined by Wu *et al*. (2016). Formal descriptions of the suggested novel taxa follow below.

Description of *Salinivirga* gen. nov. (Sa.li.ni.viŕga. L. neut. n. *salinum* salt-cellar; L. fem. n. *virga* rod; N.L. fem. n. *salinivirga* a saline rod).

Unpigmented, Gram-staining negative, straight to slightly undulated rod-shaped cells, occurring single or in pairs. An active bending or flexing of cells was not observed. No flagella or spores are formed. Major cellular fatty acids are *iso*-$C_{15:0}$ and *iso*-$C_{17:0}$ 3OH. The polar lipid composition is complex comprising phosphatidylethanolamine and several distinct unidentified polar lipids, including aminolipids and glycolipids. The main respiratory lipoquinone is menaquinone 7 (MK7). Tests for oxidase and catalase are negative. Obligately anaerobic. Strictly fermentative metabolism. Moderately halophilic, mesophilic and neutrophilic. Complex nutrient requirements. The type species is *S. cyanobacteriivorans*.

Description of *S. cyanobacteriivorans* sp. nov. (cy.a.no.bac.te.ri.i.voŕans. N.L. neut. n. pl. *cyanobacteria* Cyanobacteria; L. pres. part. *vorans* eating, devouring; N.L. part. adj. *cyanobacteriivorans* devouring cyanobacteria).

Shows the following characteristics in addition to those given for the genus. Most cells have a width of 0.4–0.5 μm and a length of 5–8 μm. No colonies are formed on agar plates. Optimal conditions for growth are 35°C, pH 6.9–7.0 and a salinity of 9% (w/v) NaCl; temperatures from 20 to 45°C, pH values from 5.8 to 8.2 and salinities from 50 to 180 g l^{-1} NaCl are tolerated. Growth depends on the presence of oligopeptides, but not vitamins or amino acids. The degradation and utilization of complete proteins requires additional substrates like yeast extract or glucose. The complex compounds Trypticase peptone, Tryptone and yeast extract are suitable substrates for growth. In the presence of yeast extract the following carbon sources are utilized: albumin, galactose, gelatine, glucose, maltose, pyruvate, sucrose and starch. The following compounds are not utilized: acetate, agar, Casamino acids, cellobiose, cellulose, chitin, ethanol, fructose, lactate, lactose, raffinose, trehalose and xylose. The end-products resulting from glucose fermentation are acetate, CO_2 and H_2. Fermentation products formed upon growth with yeast extract or peptone are acetate, succinate, propionate, isobutyrate and isovalerate. Sodium sulfate, sodium thiosulfate, sodium sulfite, elemental sulfur, sodium nitrate and sodium nitrite are not used as terminal electron acceptors. Susceptible to chloramphenicol A, rifampicin and tetracycline (each at 10 mg l^{-1}). Ampicillin, carbenicillin, D-cycloserine and penicillin G are tolerated at least in concentrations up to 100 mg l^{-1}, while gentamicin and kanamycin A is tolerated up to 1000 mg l^{-1}. In addition to the major fatty acids listed in the description of the genus significant amounts of *iso*-$C_{15:0}$ ALDE, *anteiso*-$C_{15:0}$, *anteiso*-$C_{17:0}$ 3OH, *iso*-$C_{16:0}$ 3OH, *iso*-$C_{13:0}$ and *iso*-$C_{15:0}$ 3OH are present. The DNA G + C content of the type strain is 39.0 mol%.

The type strain is L21-Spi-D4^T (=DSM 27204^T = KCTC 15528^T= JCM 31231^T) and was isolated from the suboxic zone of a hypersaline microbial mat at the littoral zone of Lake 21, Kiritimati, Republic of Kiribati.

Description of *Salinivirgaceae* fam. nov. (Sa.li.ni.vir.ga.céae. N.L. fem. n. *Salinivirga* type genus of the family; L. suff. *-aceae* ending to denote a family; N.L. fem. n. pl. *Salinivirgaceae* family of the genus *Salinivirga*).

Rod-shaped, Gram-negative bacteria. Mesophilic, moderately halophilic and anaerobic bacteria mainly found in hypersaline sediments or microbial mats that are characterized by the presence of cyanobacteria. The affiliation of novel species to this family depends on the phylogenetic position, which should be determined based on comparative sequence analyses of 16S rRNA genes. In addition, the 16S rRNA gene sequence identity values of newly described strains affiliated with this family should be above 86.5% to the type strain of the type species *S. cyanobacteriivorans*, which represents the threshold recommended for the definition of families (Yarza *et al*., 2014).

The type genus of the family is *Salinivirga*.

Experimental procedures

Culturing

Strain L21-Spi-D4^T was isolated from an anaerobic enrichment culture inoculated with slurries of a cyanobacterial mat sample retrieved from the hypersaline Lake 21 on the Kiritimati Atoll (Northern Line Islands, Republic of Kiribati). The location of the sampling site and details of the isolation method were described elsewhere (Spring *et al*., 2015). For the preparation of media and incubation under anoxic conditions, the anaerobe cultivation technique of Hungate (Hungate, 1950) with the modifications introduced by Bryant (Bryant, 1972) was used. For the characterization of strain L21-Spi-D4^T a complex basal medium was used, which has been described previously (Ben Hania *et al*., 2015). For routine cultivation 20 mM D-glucose was added as substrate to the complex basal medium. For the determination of nutrient requirements an alternative defined medium was used that contained per litre: 60.0 g NaCl, 6.0 g $MgCl_2 \times 6\ H_2O$, 1.5 g KCl, 1.0 g Na_2SO_4, 1.0 g NH_4Cl, 0.4 g $CaCl_2 \times 2\ H_2O$, 0.4 g K_2HPO_4, 10.0 ml trace elements solution of DSMZ medium 141 (http://www.dsmz.de/microorganisms/medium/pdf/DSMZ_Medium141.pdf), 0.5 mg resazurin, 10.0 ml vitamins solution of DSMZ medium 141, 2.5 g Na_2CO_3, 0.3 g $Na_2S \times 9\ H_2O$ and 0.3 g L-cysteine-HCl $\times\ H_2O$. The medium was prepared under 80% N_2 and 20% CO_2 gas mixture without the vitamins, carbonate, sulfide and cysteine, which were added to the medium after autoclaving from sterile anoxic stock solutions.

All chemicals were obtained from Sigma-Aldrich Chemie (Munich, Germany) and complex nutrients from BD Biosciences (Heidelberg, Germany).

Analyses of phylogeny and environmental distribution

The 16S rRNA gene sequence of strain L21-Spi-D4^T was determined previously (Spring *et al.*, 2015) and deposited in the GenBank/EMBL/DDBJ databases under the accession number KC665951. The 16S rRNA sequence of L21-Spi-D4^T was added to the alignment of the SILVA database (Quast *et al.*, 2013; SSU Ref NR 99 release 123) 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 major clades of the *Bacteroidetes* phylum were selected resulting in a comprehensive data set for phylogenetic analyses. Thereafter, remaining alignment errors revealed by visual inspection were corrected manually. SusC/RagA-like protein sequences were obtained from GenBank or Uniprot and aligned using the ClustalW algorithm implemented in the ARB package. Phylogenetic trees based on aligned data sets of 16S rRNA gene or SusC/RagA 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 PAM for proteins. Maximum likelihood trees were reconstructed using RAxML (version 7.7.2) with the GTRGAMMA model for DNA and PROTGAMMALG for proteins under the rapid bootstrap analysis algorithm. The maximum parsimony program of ARB was used with default settings for nucleotide or amino acid sequences. The robustness of the tree topologies was evaluated by performing 1000 rounds of bootstrap replicates. The phylogenetic diversity within the MgMjR-022 clade was deduced with the similarity option of the ARB distance matrix program. Sequences of low quality (labelled in red in the SILVA 123 SSU Ref NR99 guide tree) were excluded from the calculation.

The relative abundances of various clades in distinct layers of the Lake 21 microbial mat were determined based on a previously published 16S rRNA gene data set deposited under the NCBI Sequence Read Archive project accession number SRA058120 (Schneider *et al.*, 2013). The proportion of sequences affiliated with family-level clades were taken from a taxonomic analysis provided by Dominik Schneider (*pers. communication*), which was basically obtained by similarity searches using BLAST against the SILVA 111 SSU NR99 reference database. Note that sequences that had been retrieved from the Kiritimati hypersaline mat and were affiliated with the *Rhodothermaceae* should now be assigned to the family *Salinibacteraceae* within the phylum *Rhodothermaeota* according to a recent taxonomic proposal (Munoz *et al.*, 2016). The proportion of sequences affiliated with a species-level clade represented by strain L21-Spi-D4^T was determined by using a minimum sequence identity of 97% with the deposited 16S rRNA gene sequence of this isolate (KC665951). The analysis of the distribution of 16S rRNA gene sequences in the Guerrero Negro mat was based on previously published data (Harris *et al.*, 2013) included in the SILVA 123 SSU Ref data set.

Genome sequencing, assembly and comparative genomics

Genomic DNA was isolated form a stationary culture of L21-Spi-D4^T using the Jetflex Genomic DNA Purification Kit (GENOMED Cat. no. 600100; Löhne, Germany) according to the protocol provided by the manufacturer with the modifications described previously (Ben Hania *et al.*, 2015). The complete genome sequence was 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) and the other for the Illumina HiSeq platform (Illumina, San Diego, CA, USA). 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" applying C2 chemistry. In total seven SMRT cells were run, taking 120 minutes movies except one that was 180 minutes. Illumina sequencing was performed on a HiSeq 2500 platform with 2x100 cycles. The paired-end library contained inserts of an average size of 500 bp and delivered 4 million reads.

A draft long read genome assembly named L21-FB-Spi-V4-SP5_HGAP_7SC_std2 was created using the "RS_HGAP_Assembly.2" protocol included in SMRTPortal version 2.2.0 including all seven SMRT cells applying parameters described previously (Ben Hania *et al.*, 2015), but not allowing partial alignments. Finally, one chromosomal contig could be obtained, which was trimmed, circularized and adjusted to *dnaA* as first gene. A total coverage of 281x has been calculated within the long read assembly process. DNA base modifications analysis was performed by "RS_Modification_ and_Motif_Analysis.1" protocol with default settings. Quality check of the final consensus sequences regarding overall coverage as well as SNPs was performed using IGV (Thorvaldsdóttir *et al.*, 2013) after mapping of Illumina short read data onto the draft genome using BWA (Li and Durbin, 2009).

Genome annotation was primarily done using PROKKA version 1.8 (Seemann, 2014). The annotated genome was 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 using BLASTP to search 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-ER) platform (Markowitz *et al.*, 2009).

Determination of phenotypic traits

The Gram reaction was determined with heat fixed liquid cultures stained with BD Difco kit reagents. For electron microscopy, whole cells and thin-sections were prepared as described previously (Ben Hania *et al.*, 2015). The presence of spores was analyzed by phase contrast microscopic observations of young and old cultures and pasteurization tests, performed at 80, 90 and 100°C for 10 and 20 min.

The pH and temperature ranges for growth were determined using basal medium supplemented with 20 mM glucose. Different pH values (5–9) of the medium were adjusted by injecting aliquots of anoxic stock solutions of 0.1 M HCl (acidic pH), 10% $NaHCO_3$ or Na_2CO_3 (basic pH) in Hungate-type tubes. Water baths were used for incubating bacterial

cultures from 15 to 55°C. 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 20 mM or 2.0 g l^{-1} (complex compounds or polymers) in glucose-free medium. To test for electron acceptors, sodium thiosulfate (20 mM), sodium sulfate (20 mM), sodium sulfite (2 mM), elemental sulfur (10.0 g l^{-1}), sodium nitrate (20 mM) or sodium nitrite (2 mM) were added to the medium. Cultures were subcultured at least twice under the same experimental conditions before determination of growth rates. Sulfide production was determined photometrically as colloidal CuS according to the method of Cord-Ruwisch (1985). End-products of fermentation were either measured by high pressure liquid chromatography (HPLC) using the method of Fardeau *et al.* (1997) or by gas chromatography according to published protocols (Holdemann *et al.*, 1977; Steer *et al.*, 2001). Susceptibility to antibiotics was tested as previously described (Ben Hania *et al.*, 2015).

The cellular fatty acid pattern of strain L21-Spi-D4^T was determined from cells grown for three days at 35°C in basal complex medium containing 1.0 g l^{-1} D-glucose as carbon source. The preparation and extraction of fatty acid methyl esters from biomass and their subsequent separation and identification by gas chromatography was done as reported elsewhere (Kaksonen *et al.*, 2006). Extraction and analyses of respiratory lipoquinones and polar lipids were carried out according to previously published protocols (Tindall, 1990; Tindall *et al.*, 2007).

Nucleotide sequence accession number

The annotated complete genome sequence of strain L21-Spi-D4^T was deposited in GenBank with the accession number CP013118. The version described in this paper is the first version, CP013118.1.

Acknowledgements

We thank Simone Severitt 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 ID service of the DSMZ (Anja Frühling, Gabriele Pötter, Dr Brian Tindall and Dr Susanne Verbarg) is acknowledged for cellular fatty acid, respiratory quinone and polar lipid analyses. The help of Aharon Oren and Bernhard Schink in reporting the correct etymology of the proposed taxonomic names is appreciated.

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 of Organo- and Biofilms."

References

Arp, G., Helms, G., Karlinska, K., Schumann, G., Reimer, A., Reitner, J., and Trichet, J. (2012) Photosynthesis versus exopolymer degradation in the formation of microbialites on the atoll of Kiritimati, Republic of Kiribati, Central Pacific. *Geomicrobiol J* **29:** 29–65.

Barker, H.A. (1981) Amino acid degradation by anaerobic bacteria. *Annu Rev Biochem* **50:** 23–40.

Ben Hania, W., Joseph, M., Schumann, P., Bunk, B., Fiebig, A., Spröer, 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.

Boichenko, V.A., Wang, J.M., Antón, J., Lanyi, J.K., and Balashov, S.P. (2006) Functions of carotenoids in xanthorhodopsin and archaerhodopsin, from action spectra of photoinhibition of cell respiration. *Biochim Biophys Acta* **1757:** 1649–1656.

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

Burnat, M., Herrero, A., and Flores, E. (2014) Compartmentalized cyanophycin metabolism in the diazotrophic filaments of a heterocyst-forming cyanobacterium. *Proc Natl Acad Sci U S A* **111:** 3823–3828.

Cecchini, G., Maklashina, E., Yankovskaya, V., Iverson, T.M., and Iwata, S. (2003) Variation in proton donor/acceptor pathways in succinate:quinone oxidoreductases. *FEBS Lett* **545:** 31–38.

Chan, Z., Wang, R., Liu, S., Zhao, C., Yang, S., and Zeng, R. (2015) Draft genome sequence of an agar-degrading marine bacterium *Flammeovirga pacifica* WPAGA1. *Mar Genomics* **20:** 23–24.

Cord-Ruwisch, R. (1985) A quick method for the determination of dissolved and precipitated sulfides in cultures of sulfate-reducing bacteria. *J Microbiol Methods* **4:** 33–36.

Dupraz, C., and Visscher, P.T. (2005) Microbial lithification in marine stromatolites and hypersaline mats. *Trends Microbiol* **13:** 429–438.

Fardeau, M.L., Ollivier, B., Patel, B.K., Magot, M., Thomas, P., Rimbault, A., *et al.* (1997) *Thermotoga hypogea* sp. nov., a xylanolytic, thermophilic bacterium from an oil-producing well. *Int J Syst Bacteriol* **47:** 1013–1019.

Farías, M.E., Contreras, M., Rasuk, M.C., Kurth, D., Flores, M.R., Poiré, D.G., *et al.* (2014) Characterization of bacterial diversity associated with microbial mats, gypsum evaporites and carbonate microbialites in thalassic wetlands: Tebenquiche and La Brava, Salar de Atacama, Chile. *Extremophiles* **18:** 311–329.

Fernández-Gómez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S.G., González, J.M., and Pedrós-Alió, C. (2013) Ecology of marine *Bacteroidetes*: a comparative genomics approach. *ISME J* **7:** 1026–1037.

Garcia-Pichel, F., Nübel, U., and Muyzer, G. (1998) The phylogeny of unicellular, extremely halotolerant cyanobacteria. *Arch Microbiol* **169:** 469–482.

Gronow, S., Munk, C., Lapidus, A., Nolan, M., Lucas, S., Hammon, N., *et al.* (2011) Complete genome sequence of *Paludibacter propionicigenes* type strain (WB4). *Stand Genomic Sci* **4:** 36–44.

Grotzinger, J.P., and Knoll, A.H. (1999) Stromatolites in precambrian carbonates: evolutionary mileposts or environmental dipsticks? *Annu Rev Earth Planet Sci* **27:** 313–358.

Harris, J.K., Caporaso, J.G., Walker, J.J., Spear, J.R., Gold, N.J., Robertson, C.E., *et al.* (2013) Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat. *ISME J* **7:** 50–60.

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.

Huntemann, M., Ivanova, N.N., Mavromatis, K., Tripp, H.J., Paez-Espino, D., Palaniappan, K., *et al*. (2015) The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Stand Genomic Sci* **10:** 86.

Iino, T., Sakamoto, M., and Ohkuma, M. (2015) *Prolixibacter denitrificans* sp. nov., an iron-corroding, facultatively aerobic, nitrate-reducing bacterium isolated from crude oil, and emended descriptions of the genus *Prolixibacter* and *Prolixibacter bellariivorans*. *Int J Syst Evol Microbiol* **65:** 2865–2869.

Ionescu, D., Spitzer, S., Reimer, A., Schneider, D., Daniel, R., Reitner, J., *et al*. (2015) Calcium dynamics in microbialite-forming exopolymer-rich mats on the atoll of Kiritimati, Republic of Kiribati, Central Pacific. *Geobiology* **13:** 170–180.

Kabisch, A., Otto, A., König, S., Becher, D., Albrecht, D., Schüler, M., *et al*. (2014) Functional characterization of polysaccharide utilization loci in the marine *Bacteroidetes* "*Gramella forsetii*" KT0803. *ISME J* **8:** 1492–1502.

Kaksonen, A.H., Spring, S., Schumann, P., Kroppenstedt, R.M., and Puhakka, J.A. (2006) *Desulfotomaculum 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.

Ley, R.E., Harris, J.K., Wilcox, J., Spear, J.R., Miller, S.R., Bebout, B.M., *et al*. (2006) Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Appl Environ Microbiol* **72:** 3685–3695.

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

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.

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.

Martens, E.C., Koropatkin, N.M., Smith, T.J., and Gordon, J.I. (2009) Complex glycan catabolism by the human gut microbiota: the *Bacteroidetes* Sus-like paradigm. *J Biol Chem* **284:** 24673–24677.

McIlroy, S.J., and Nielsen, P.H. (2014) The family *Saprospiraceae*. In *The Prokaryotes – Other Major Lineages of Bacteria and the Archaea*. Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., and Thompson, F. (eds), Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 863–889.

Munoz, R., Rosselló-Móra, R., and Amann, R. (2016) Revised phylogeny of *Bacteroidetes* and proposal of sixteen new taxa and two new combinations including *Rhodothermaeota* phyl. nov. *Syst Appl Microbiol* **39:** 281–296.

Nagano, K., Murakami, Y., Nishikawa, K., Sakakibara, J., Shimozato, K., and Yoshimura, F. (2007) Characterization of RagA and RagB in *Porphyromonas gingivalis*: study using gene-deletion mutants. *J Med Microbiol* **56:** 1536–1548.

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.

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.

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.

Sahl, J.W., Pace, N.R., and Spear, J.R. (2008) Comparative molecular analysis of endoevaporitic microbial communities. *Appl Environ Microbiol* **74:** 6444–6446.

Schneider, D., Arp, G., Reimer, A., Reitner, J., and Daniel, R. (2013) Phylogenetic analysis of a microbialite-forming microbial mat from a hypersaline lake of the Kiritimati Atoll, Central Pacific. *PLoS One* **8:** e66662.

Schut, G.J., Boyd, E.S., Peters, J.W., and Adams, M.W.W. (2013) The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. *FEMS Microbiol Rev* **37:** 182–203.

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

Shalley, S., Pradip Kumar, S., Srinivas, T.N.R., Suresh, K., and Anil Kumar, P. (2013) *Marinilabilia nitratireducens* sp. nov., a lipolytic bacterium of the family *Marinilabiliaceae* isolated from marine solar saltern. *Antonie Van Leeuwenhoek* **103:** 519–525.

Shi, M., Zou, L., Liu, X., Gao, Y., Zhang, Z., Wu, W., *et al*. (2006) A novel bacterium *Saprospira* sp. strain PdY3 forms bundles and lyses cyanobacteria. *Front Biosci* **11:** 1916–1923.

Siegesmund, M.A., Johansen, J.R., Karsten, U., and Friedl, T. (2008) *Coleofasciculus* gen. nov. (*Cyanobacteria*): morphological and molecular criteria for revision of the genus *Microcoleus* Gomont(1). *J Phycol* **44:** 1572–1585.

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

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

Spring, S., Rachel, R., Lapidus, A., Davenport, K., Tice, H., Copeland, A., *et al*. (2010) Complete genome sequence of *Thermosphaera aggregans* type strain (M11TL). *Stand Genomic Sci* **2:** 245–259.

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.

Takahashi, N., Sato, T., and Yamada, T. (2000) Metabolic pathways for cytotoxic end product formation from glutamate- and aspartate-containing peptides by *Porphyromonas gingivalis*. *J Bacteriol* **182:** 4704–4710.

Terrapon, N., Lombard, V., Gilbert, H.J., and Henrissat, B. (2015) Automatic prediction of polysaccharide utilization loci in *Bacteroidetes* species. *Bioinformatics* **31:** 647–655.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, 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.

Tindall, B.J., Sikorski, J., Smibert, R.A., and Krieg, N.R. (2007) Phenotypic characterization and the principles of comparative systematics. In *Methods for General and Molecular Microbiology*, 3rd ed. Reddy, C., Beveridge, T., Breznak, J., Marzluf, G., Schmidt, T., and Snyder, L. (eds). Washington, DC: American Society of Microbiology, pp. 330–393.

Verkhovsky, M.I., and Bogachev, A.V. (2010) Sodium-translocating NADH:quinone oxidoreductase as a redox-driven ion pump. *Biochim Biophys Acta* **1797:** 738–746.

Wong, H., Ahmed-Cox, A., and Burns, B. (2016) Molecular ecology of hypersaline microbial mats: current insights and new directions. *Microorganisms* **4:** 6.

Wu, W., Zhao, J., Chen, G., and Du, Z. (2016) Description of *Ancylomarina subtilis* gen. nov., sp. nov., isolated from coastal sediment, proposal of *Marinilabiliales* ord. nov. and transfer of *Marinilabiliaceae*, *Prolixibacteraceae* and *Marinifilaceae* to the order *Marinilabiliales*. *Int J Syst Evol Microbiol* **66:** 4243–4249.

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.H., *et al*. (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* **12:** 635–645.

Zhao, K.H., Zhang, J., Tu, J.M., Böhm, S., Plöscher, M., Eichacker, L., *et al*. (2007) Lyase activities of CpcS- and CpcT-like proteins from *Nostoc* PCC7120 and sequential reconstitution of binding sites of phycoerythrocyanin and phycocyanin beta-subunits. *J Biol Chem* **282:** 34093–34103.

Supporting information

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

Fig. S1. Polar lipids pattern of strain L21-Spi-D4^T revealed after two dimensional thin layer chromatography.
Fig. S2. HPLC chromatograms of culture supernatants of strain L21-Spi-D4^T upon growth in medium containing either glucose and yeast extract or only yeast extract as substrates.
Fig. S3. GC analyses of culture supernatants of strain L21-Spi-D4^T upon growth with Trypticase peptone as substrate.
Fig. S4. Circular map of the chromosome of strain L21-Spi-D4^T.
Fig. S5. Phylogenetic tree of SusC/RagA protein sequences showing the positions of TonB-dependent transporters of strain L21-Spi-D4^T among related proteins of various representatives of the *Bacteroidetes*.
Fig. S6. Suggested pathway for the fermentation of glucose to acetate in strain L21-Spi-D4^T.
Fig. S7. Suggested pathways for the fermentation of amino acids in strain L21-Spi-D4^T.
Table S1. Cellular fatty acid patterns of strain L21-Spi-D4^T and type strains of the phylogenetically most closely related type species of the *Marinilabiliaceae*.
Table S2. Differential traits of strain L21-Spi-D4^T and type strains of the phylogenetically most closely related type species of the *Marinilabiliaceae*.
Table S3. Classification and number of peptidase sequences identified in the L21-Spi-D4^T genome by the MEROPS database.
Table S4. Classification, number and predicted activities of glycoside hydrolase genes identified in the L21-Spi-D4^T genome by the CAZy database.
Table S5. Gene composition of putative cyanophycin utilization loci identified in genomes available in the IMG database.