Proteiniphilum saccharofermentans str. M3/6[T] isolated from a laboratory biogas reactor is versatile in polysaccharide and oligopeptide utilization as deduced from genome-based metabolic reconstructions

Geizecler Tomazetto[a,*], Sarah Hahnke[c], Daniel Wibberg[b,*], Alfred Pühler[b], Michael Klocke[c], Andreas Schlüter[b]

[a] Brazilian Bioethanol Science and Technology Laboratory – CTBE/CNPEM, 10000 Giuseppe Maximo Scolfaro St, Zip Code 13083-852 Campinas, SP, Brazil
[b] Center for Biotechnology (CeBiTec), Genome Research of Industrial Microorganisms, Bielefeld University, Universitätsstr. 27, 33615 Bielefeld, Germany
[c] Department Bioengineering, Leibniz Institute for Agricultural Engineering and Bioeconomy (ATB), Max-Eyth-Allee 100, 14469 Potsdam, Germany

A B S T R A C T

Proteiniphilum saccharofermentans str. M3/6[T] is a recently described species within the family *Porphyromonadaceae* (phylum *Bacteroidetes*), which was isolated from a mesophilic laboratory-scale biogas reactor. The genome of the strain was completely sequenced and manually annotated to reconstruct its metabolic potential regarding biomass degradation and fermentation pathways. The *P. saccharofermentans* str. M3/6[T] genome consists of a 4,414,963 bp chromosome featuring an average GC-content of 43.63%. Genome analyses revealed that the strain possesses 3396 protein-coding sequences. Among them are 158 genes assigned to the carbohydrate-active-enzyme families as defined by the CAZy database, including 116 genes encoding glycosyl hydrolases (GHs) involved in pectin, arabinogalactan, hemicellulose (arabinan, xylan, mannan, β-glucans), starch, fructan and chitin degradation. The strain also features several transporter genes, some of which are located in polysaccharide utilization loci (PUL). PUL gene products are involved in glycan binding, transport and utilization at the cell surface. In the genome of strain M3/6[T], 64 PUL are present and most of them in association with genes encoding carbohydrate-active enzymes. Accordingly, the strain was predicted to metabolize several sugars yielding carbon dioxide, hydrogen, acetate, formate, propionate and isovalerate as end-products of the fermentation process. Moreover, *P. saccharofermentans* str. M3/6[T] encodes extracellular and intracellular proteases and transporters predicted to be involved in protein and oligopeptide degradation. Comparative analyses between *P. saccharofermentans* str. M3/6[T] and its closest described relative *P. acetatigenes* str. DSM 18083[T] indicate that both strains share a similar metabolism regarding decomposition of complex carbohydrates and fermentation of sugars.

© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Biogas can be produced by anaerobic digestion (AD) of a wide range of plant materials, organic wastes, and residual organic materials. The biogas-production process is regarded as eco-friendly technology to generate energy from biomass [1,2]. AD is commonly divided into four phases, i.e. hydrolysis, acidogenesis, acetogenesis and methanogenesis, which are conducted by complex consortia consisting of several hundreds of microbial species. Despite the fact that the overall biological process, which finally leads to the production of biogas, is well known, the majority of microbial biogas community members and their metabolic activities in particular are largely unknown [3–7].

In recent years, several bioreactors were taxonomically profiled by high-throughput sequencing of the 16S rRNA marker gene [5, 8–10]. These studies reported that members of the classes *Clostridia* and *Bacteroidia* frequently dominate biogas communities. Members of both classes are responsible for degradation of complex carbohydrates and proteins to monomers and are able to ferment sugar molecules yielding volatile organic acids [11,12]. To deduce functional profiles of biogas communities, shotgun metagenome sequencing has been done [6,13–18].

The first metagenomic studies were based on non-assembled short reads and on small numbers of short contigs [4,14–16,19], providing a gene content overview of microbial communities involved in anaerobic digestion. Recently, deep metagenome

sequencing of DNA from biogas communities enabled assembly of sequence reads and binning of contigs to bin-genomes improving gene prediction and functional interpretation of metagenome sequence data [6,10,20]. However, the reconstruction of complete genome sequences from metagenome sequence data is demanding which is due to highly related sequences originating from different organisms [21,22]. Culture-independent approaches helped to elucidate taxonomic structures and gene contents of many microbiomes [23–26]. Complementary, traditional cultivation-based microbiological analyses including genome sequencing still yield reliable genome sequence information of individual micro-organisms and corresponding phenotypic features, which are also useful for interpretation of bin-genomes. Recently, a number of new bacterial species were isolated from biogas reactors [80]. Subsequent genome sequencing and metabolic reconstructions based on genome sequence information led to the prediction of the role of these microorganisms within the biogas process [27–31]. Among these newly characterized strains, *Proteiniphilum saccharofermentans* str. M3/6[T] was isolated from a mesophilic laboratory-scale biogas reactor [31]. Microbiological characterization revealed that, besides utilization of complex proteinaceous substrates such as yeast extract and peptone, the isolate was able to ferment mono- and disaccharides. Moreover, it produced extracellular enzymes involved in degradation of complex carbohydrates, namely β-glucan, xylan, arabinoxylan, starch, arabinogalactan, phosphoric acid-swollen cellulose and carboxymethyl cellulose (CM-cellulose). Considering these phenotypic features indispensable for effective biomass conversion, it was worthwhile to establish and analyze the complete genome sequence of *P. saccharofermentans* str. M3/6[T] to uncover its genetic potential regarding carbohydrate-active enzymes involved in AD of biomass. The genome was manually annotated and interpreted to reconstruct metabolic pathways dedicated to biomass degradation and fermentation processes. Gene clusters encoding polysaccharide utilization loci (PUL) were analyzed in detail. Obtained findings are of importance regarding the biotechnological process of biomass conversion to biofuels.

2. Material and methods

2.1. Strain cultivation and DNA isolation

P. saccharofermentans str. M3/6[T] was cultivated at 37 °C in anoxic basal medium with yeast extract and proteose peptone (5 g l^{-1} each) as described by Hahnke et al. [31]. The extraction of genomic DNA was performed using the Gentra Puregene Yeast/Bact. Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The obtained DNA was purified using the Nucleo-oSpin® gDNA Clean-up kit (Macherey-Nagel, Düren, Germany).

2.2. Sequencing, assembly and annotation

The total genomic DNA was used for the construction of a standard shotgun library applying the Nextera® Mate-Pair Library Preparation Kit (Illumina), according to the manufacturer's protocol. The genomic library was sequenced on the Illumina MiSeq system. After processing of the raw data, the reads were assembled into contigs using the Newbler *De Novo* Assembler (version 2.8). Genome finishing was performed using the CONSED software package [32] for ordering and joining the contigs. Genome analyses, interpretation and reconstruction of metabolic pathways were performed as recently described [33]. Briefly, the assembled genome sequence was imported into the annotation platform GenDB [34] for automatic prediction of genes. All predicted genes were analyzed and validated manually by means of BLAST against different databases including Pfam, TIGRFAM, InterPro, SwissProt, and the non-redundant NCBI protein sequence database (NR). Putative tRNA and rRNA genes were identified with RNAmmer [35], RBSfinder [36] and tRNAscan-SE [37] and SignalP [38] and TMHMM [39] were used to predict signal peptides and transmembrane proteins. To identify phage-related genes and genomic islands (GI), the *P. saccharofermentans* str. M3/6[T] genome was uploaded in PHAST [40] and IslandViewer [41], respectively. The CRISPRfinder [42] in combinantion with the CRISPRdb database [43] was applied to identified CRISPR arrays (Clustered Regularly Interspaced Short Palindromic Repeats) in the strain M3/6[T] genome. The *cas* gene predicted by GenDB was manually verified by means of BLAST against the database cited above. Finally, comparisons between the genome of strain M3/6[T] and that of the type strain of its most closely related species, *Proteiniphilum acetatigenes* DSM 18083[T], were carried out using the EDGAR tool [44]. Synteny analyses, identification of orthologous genes and classification of genes as core genes or singletons were done within EDGAR.

2.3. Reconstruction of metabolic pathways

To determine the diversity of carbohydrate-active enzyme (CAZyme) families present in the *P. saccharofermentans* str. M3/6[T] genome, all predicted gene products were compared against the HMM profile-based database dbCAN [45] using hmmsearch in the HMMER software package [46]. Predicted CAZymes were also analyzed using Priam profiles [47]. The metabolic pathways of biomass degradation represented in the genome of strain M3/6[T] were reconstructed based on EC numbers of predicted enzymes in combination with Pathway tools [48]. Finally, ABC transporters were predicted and classified comparing the predicted proteins to the TCDB database [49].

2.4. Nucleotide sequence accession number

The genome of *P. saccharofermentans* str. M3/6[T] was deposited in the EMBL-EBI database (European Bioinformatics Institute database) under the accession number LT605205.

3. Results and discussion

3.1. General features of the Proteiniphilum saccharofermentans str. M3/6[T] genome

The strain M3/6[T] was isolated from a two-phase Upflow Anaerobic Solid-State (UASS) reactor fed with 95% maize silage and 5% wheat straw as substrates [31]. The bacterium belongs to the family *Porphyromonadaceae* within the phylum *Bacteroidetes* and it was characterized as an acidogenic microorganism producing acetate, propionate and isovalerate [31]. To uncover the genetic potential of the strain in the context of the biogas production process, its genome sequence was established, manually annotated and analyzed including reconstruction of metabolic pathways and comparative examination.

Sequencing on the Illumina MiSeq system resulted in 3,174,424 sequence reads corresponding to a 165-fold coverage of the 4.4 Mb genome. The Newbler assembler (version 2.8) was used to assemble obtained reads into 60 (>100 bp) contigs. *In silico* finishing applying the platform CONSED led to closure of all gaps between contigs and circularization of the genome. The finished chromosomal sequence consists of 4,414,963 bp and has a GC content of 43.63% (Table 1, Fig. 1). Gene prediction resulted in identification of 3396 protein coding sequences (CDS), 48 tRNA genes, and six ribosomal RNA (*rrn*) operons. Among the CDSs, 58.5% could be classified according to COG categories comprising 20 higher-ranking functional groups (Table 1, Supplementary file 1: Table S1).

Table 1
General features of the *P. saccharofermentans* str. M3/6^T genome.

Feature	Chromosome
Genome size (bp)	4,414,963
GC content (%)	43.63
Total genes	3450
Protein coding genes	3396
Genes with functional prediction	2385
Genes assigned to COGs (%)	58.5
Genes encoding proteins with signal peptides	472
Genes encoding proteins with transmembrane helices	847
rRNA operons	6
tRNA genes	48

The strain M3/6^T genome contains some putative phage genes, but no complete prophage cluster could be identified. Regarding mobile genetic elements (MGE), 89 transposase genes were identified. Moreover, the strain possesses two CRISPR-*cas* systems which commonly play a role in preventing invasion of phages and mobile genetic elements [50].

No virulence, antibiotic resistance or pathogenicity-related genes were identified in the strain M3/6^T genome by IslandViewer [41].

3.2. Genes involved in complex carbohydrate or protein degradation

A number of species of the phylum *Firmicutes*, namely of the families *Clostridiaceae* and *Ruminococcaceae* are able to hydrolyze long-chained carbohydrates such as cellulose [51,52]. Most prominent and well-studied is the thermophilic species *Clostridium thermocellum*, but also other members of the genera *Clostridium*, *Acetivibrio*, and *Ruminococcus* adapted to mesophilic or

thermophilic process regimes are involved in cellulose degradation [53]. In some biogas reactors operated at mesophilic temperatures, also a higher abundance of members of the phylum *Bacteroidetes* was observed indicating their participation in hydrolysis of organic compounds [54,55]. In contrast, in thermophilic biogas reactors operated at 50–60 °C, *Bacteroidetes* were not or only detectable in minor amounts [7]. Interestingly, upon further increase of the process temperature to 65 °C and above, *Bacteroidetes* members proliferate again leading to a reduced abundance of *Firmicutes* [56]. However, most of the *Bacteroidetes* species detected in biogas reactors are not sufficiently characterized yet to interpret their role in AD.

To overcome these limitations, the genome of the newly isolated *P. saccharofermentans* str. M3/6^T was screened for the presence of carbohydrate-active enzymes listed in the CAZy database (www.cazy.org) [57].

According to the CAZy classification scheme, the genome of the strain M3/6^T encodes 116 glycoside hydrolyses (GH), 25 glycosyltransferases (GT), seven carbohydrate esterases (CE), two polysaccharide lyases (PL), and seven gene products featuring carbohydrate binding motifs (CBM). A summarizing overview is given in Fig. 2 and Table 2 and in Supplementary file 1 (Tables S2 and S3). Among them, the strain M3/6^T encodes a variety of carbohydrate-active enzymes required for degradation of different complex carbohydrates, including pectin, arabinogalactan, hemicellulose (arabinan, xylan, mannan, β-glucans), starch, fructan, and chitin. A recent study reported that the microorganism showed only weak extracellular enzyme activities against β-glucan, starch, arabinogalactan, xylan, arabinoxylan, CM-cellulose, and phosphoric acid-swollen cellulose [31]. However, it is important to note that different environmental conditions may cause different physiological responses and that the
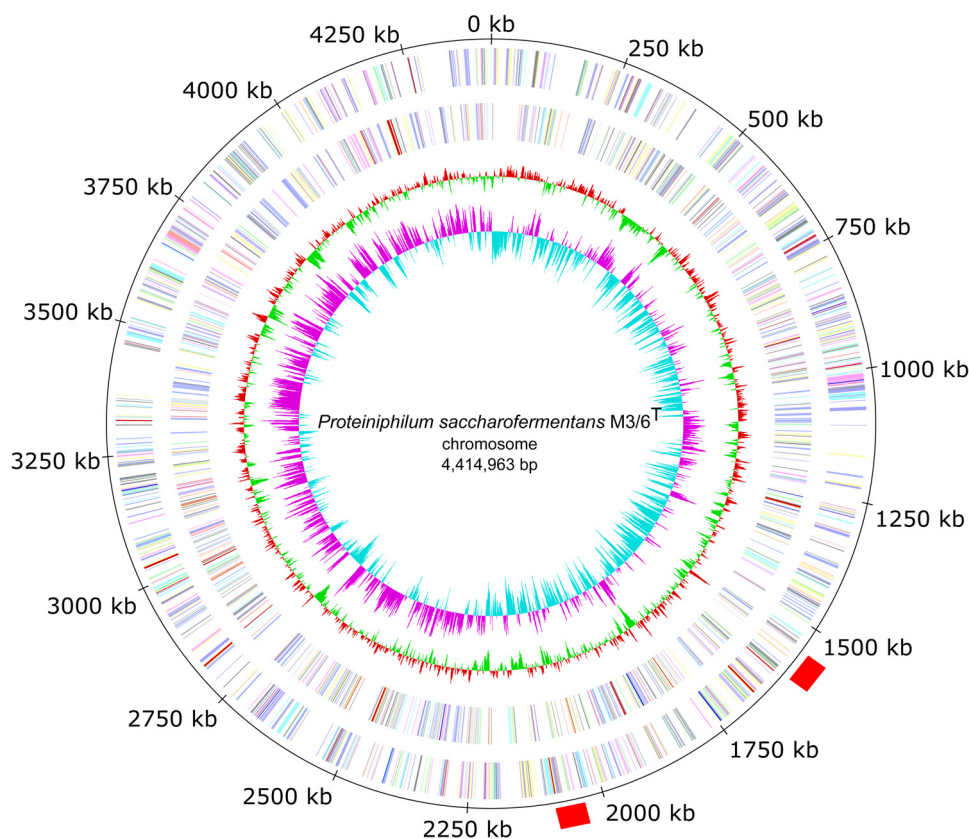

Fig. 1. Genome plot of *P. saccharofermentans* str. M3/6^T. The first outer circle represents the genome scale in kb. The origin of replication was determined based on GC skew analyses. Second and third circle: predicted protein-coding sequences (CDS) on the forward and the reverse strands colored according to the assigned COG categories. Fourth and fifth inner circles represent the GC-content and GC-skew, respectively. Red squares on the outer circle indicate CRISPR-*cas* systems.

Locus	PUL	Genes predicted
PSM36_0005 to PSM36_0009	1	LacI nPULg HP TonB SusD HP HP HP GH116
PSM36_0023 to PSM36_0026	2	nPULg nPULg SusC SusD
PSM36_0099 to PSM36_0102	3	SusD TonB Anti-σ ECF-σ
PSM36_0108 to PSM36_0111	4	SusD TonB HP ECF-σ
PSM36_0120 to PSM36_0132	5	ECF-σ Anti-σ HP HP HP HP TonB SusD HP GH16 GH3 HP ECF-σ
PSM36_0171 to PSM36_0175	6	TonB SusD SusE GH53 GH2
PSM36_0202 to PSM36_0206	7	GH16 SusD TonB Anti-σ ECF-σ
PSM36_0293 to PSM36_0298	8	GH32 GH32 HP HP SusD SusC HTCS
PSM36_0311 to PSM36_0312	9	SusD SusC
PSM36_0321 to PSM36_0322	10	SusD SusC
PSM36_0340 to PSM36_0341	11	SusD SusC
PSM36_0404 to PSM36_0409	12	GH78 GH20 HP HP SusD SusC
PSM36_0430 to PSM36_0431	13	SusD SusC
PSM36_0456 to PSM36_0460	14	CE GH92 nPULg SusD SusC
PSM36_0495 to PSM36_0501	15	GH13 GH13 GH53 CBM CBM SusD SusC
PSM36_0633 to PSM36_0639	16	CBM HP CBM SusD SusD Anti-σ ECF-σ
PSM36_0732 to PSM36_0733	17	SusC SusD
PSM36_0891 to PSM36_0953	18	GH51 GH35 GH43 GH43 GH53 nPULg nPULg GH43 nPULg nPULg GH51 nPULg GH97 nPULg HP HP HP GH43 nPULg GH35 SusD SusC HP GH28 GH28 PL11 GH105 HP HP HP GH28 GH2 GH28 GH88 CE HP GH105 HP HP GH43 TonB nPULg GH43 GH2 GH30 SusD SusC HP HP GH43 HP SusD SusC SusD SusC HP GH2 HP GH88 SusD SusC
PSM36_1051 to PSM36_1052	19	ECF-σ nPULg SusC SusD
PSM36_1078 to PSM36_1083	20	GH32 GH2 HP SusD SusC ECF-σ
PSM36_1086 to PSM36_1091	21	ECF-σ HP GH38 GH92 SusC SusD
PSM36_1102 to PSM36_1105	22	ECF-σ HP SusC SusD
PSM36_1106 to PSM36_1111	23	ECF-σ HP SusC SusD GH2 GH88
PSM36_1152 to PSM36_1159	24	GH43 SusC SusD GH43 HP GH2 nPULg GH43
PSM36_1177 to PSM36_1182	25	SusC SusD HP HP GH43 GH127
PSM36_1186 to PSM36_1189	26	GH127 nPULg SusC SusD
PSM36_1217 to PSM36_1220	27	ECF-σ Anti-σ SusC SusD
PSM36_1225 to PSM36_1228	28	ECF-σ Anti-σ SusC SusD
PSM36_1236 to PSM36_1237	29	SusC SusD
PSM36_1425 to PSM36_1428	30	GH99 HP SusD SusC
PSM36_1558 to PSM36_1565	31	AraC SusD SusC HP HP GH109 nPULg ECF-σ
PSM36_1589 to PSM36_1592	32	SusD SusC Anti-σ ECF-σ
PSM36_2010 to PSM36_2020	33	ECF-σ HP SusC SusD GH42 nPULg nPULg GH9 CBM GH2
PSM36_2062 to PSM36_2080	34	GH2 HP Sulf_1 GH105 Sulf_1 GH2 HP GH2 GH30 Sulf_1 HP SusD SusC SusD SusC GH2 HP nPULg AraC
PSM36_2169 to PSM36_2170	35	SusC SusD
PSM36_2196 to PSM36_2202	36	GH78 GH78 GH127 ECF-σ Anti-σ SusC SusD
PSM36_2203 to PSM36_2207	37	GH99 GH99 HP SusC SusD
PSM36_2227 to PSM36_2232	38	GH127 GH78 nPULg nPULg SusD SusC
PSM36_2279 to PSM36_2286	39	Petp Petp SusD SusD Petp HP SusD SusC
PSM36_2300 to PSM36_2303	40	SusD SusC nPULg GH43
PSM36_2402 to PSM36_2403	41	SusC SusD
PSM36_2461 to PSM36_2468	42	EFC HP Anti-σ SusC SusD GH20 GH42 GH42
PSM36_2477 to PSM36_2482	43	GH92 nPULg SusD SusC Anti-σ ECF-σ
PSM36_2483 to PSM36_2484	44	SusC SusD
PSM36_2502 to PSM36_2505	45	SusD SusC Anti-σ ECF-σ
PSM36_2561 to PSM36_2562	46	SusD SusC
PSM36_2570 to PSM36_2577	47	SusC SusD nPULg HP GH3 Sulf_1 GH43
PSM36_2600 to PSM36_2607	48	GH2 SusC SusD HP GH65 GH29 HP GH88
PSM36_2614 to PSM36_2616	49	SusC SusD GH76
PSM36_2645 to PSM36_2646	50	SusC SusD
PSM36_2690 to PSM36_2670	51	GH31 HP GH78 nPULg HP GH127 SusD SusC SusD SusC
PSM36_2705 to PSM36_2718	52	GH33 GH5 GH105 GH95 GH28 GH78 HP GH43 HP HP HP HP SusD SusC
PSM36_2896 to PSM36_2903	53	GH51 GH51 nPULg nPULg SusD SusC Anti-σ ECF-σ
PSM36_2973 to PSM36_2974	54	SusD SusC
PSM36_2984 to PSM36_2992	55	HTCS GH43 GH105 nPULg SusD TonB GH105 nPULg ECF-σ
PSM36_3005 to PSM36_3014	56	SusC SusD SusC SusC HP SusD SusD SusC SusC
PSM36_3045 to PSM36_3046	57	SusC SusD
PSM36_3090 to PSM36_3091	58	SusC SusD
PSM36_3112 to PSM36_3117	59	GH33 GH2 HP nPULg MFS GH130 GH26 GH26 HP HP SusD SusC
PSM36_3149 to PSM36_3157	60	PL8 nPULg GH88 nPULg HP HP SusC SusD
PSM36_3208 to PSM36_3212	61	ECF-σ Anti-σ SusC SusD GH26
PSM36_3304 to PSM36_3310	62	GH43 GH5 HP SusD SusC GH5 GH31
PSM36_3362 to PSM36_3367	63	GH95 GH29 GH3 GH29 SusD SusC
PSM36_3423 to PSM36_3431	64	SusC SusD HP HP CE GH43 GH28 SusD SusC

Fig. 2. Schematic overview of PUL (Polysaccharide Utilization Loci) predicted in the *P. saccharofermentans* str. M3/6^T genome. To facilitate visualization of gene arrangements, these are colored according to the function of the encoded proteins SusC (blue), SusD (purple), SusE (light blue), HP (hypothetical protein, gray), GHs (Glycoside Hydrolase, pink), CBM (Carbohydrate-Binding Module, green), PL (Polysaccharide Lyases, light orange), CE (Carbohydrate Esterase, dark orange), TonB (TonB-linked outer membrane protein, light pink), Pept (Peptidase, light green), MFS (Major Facilitator Superfamily, light pink), regulators (AraC, LacI, HTCS, ECF, Gntr, Anti-sigma, yellow). Genes that do not encode PUL components were marked with' non-PUL genes' (nPULg, dark brown).

pure-culture conditions applied to determine extracellular enzyme activities did not correspond to the environmental conditions in the biogas reactor from which the strain M3/6^T was isolated. Thus, regarding its repertoire of carbohydrate-active enzymes, the potential of the strain to degrade carbohydrates might be more versatile as previously reported. The repertoire of carbohydrate active enzymes encoded in the genome of strain M3/6^T is described below.

3.2.1. Pectin

Pectin is present in the primary cell wall of higher plants and responsible for its structural integrity. It is composed of homogalacturonan (HG), rhamnogalacturonan I (RGI) and rhamnogalacturonan II (RGII), all containing D-galacturonic acid in their side chains. The *P. saccharofermentans* str. M3/6^T genome encodes the enzymes pectinesterase (EC 3.1.1.11), polygalacturonase (EC 3.2.1.15), exo-poly-α-D-galacturonosidase (EC 3.2.1.82) and pectate lyases. Pectin degradation starts with the action of pectinesterases, followed by polygalacturonases and pectate lyases catalyzing deesterification of pectin. Pectic acid is released which than is broken down by endo- or exo-polygalacturonases producing mono- or di-galacturonate. Genes involved in degradation of RGI, encoding rhamnogalacturan acetylesterase (EC 3.1.1.86), rhamnogalacturonyl hydrolase (EC 3.2.1.172) and α-L-rhamnosidase (EC 3.2.1.40) were also identified in the genome of strain M3/6^T. These enzymes act on RGI and produce rhamnose residues. Furthermore, two genes for polysaccharide lyases (PL), which catalyze random cleavage of pectin, were predicted. One gene encodes a family PL8 enzyme whereas the second one represents family PL11.

3.2.2. Arabinogalactan

Arabinogalactan (AG) is a polysaccharide consisting of galactose and arabinose residues. AGs are associated with pectin in plant cell walls and occur in two structurally different forms denominated as type I and type II. Two genes encoding endo-β-1,4-galactanase (EC 3.2.1.89) responsible for degradation of AG type I were predicted in the *P. saccharofermentans* str. M3/6^T genome. Endo-β-1,4-galactanases act on D-galactosidic linkages in AGs releasing galactotetraose. Moreover, a putative galactan endo-1,6-β-galactosidase (EC 3.2.1.164) was predicted.

3.2.3. Hemicellulose

Hemicellulose is a heteropolymer of the plant cell wall closely associated with cellulose and lignin. It is composed of a variety of polysaccharides derived from sugars such as D-xylose, D-galactose, D-mannose, D-glucose, and L-arabinose. The *P. saccharofermentans* str. M3/6^T genome possesses 47 genes encoding enzymes predicted to be responsible for arabinan, xylan, mannan and β-glucan degradation.

Arabinan is a polysaccharide composed of arabinose residues, which are linked by α-1,5 glycosidic bonds in its backbone and α-1,2 and α-1,3 connections in its side chains. The strain M3/6^T genome encodes arabinan endo-1,5-α-L-arabinanase (EC 3.2.1.99) that catalyzes the endo-hydrolysis of arabinan into arabinan oligosaccharides, which are further decomposed to L-arabinose residues by α-L-arabinofuranosidase (EC 3.2.1.55). Genes encoding β-L-arabinofuranosidase (EC 3.2.1.185) are also present in the genome. The enzyme hydrolyzes side chain α-(1,2) glycosidic bonds.

Xylan is composed of D-xylose residues. Complete hydrolysis of xylan requires the synergistic action of GH and CE. In strain M3/6^T, three of seven enzymes commonly described as enzymes needed

Table 2
Carbohydrate active enzymes (CAZy) encoded in the *P. saccharofermentans* str. M3/6^T genome.

Substrate target	Enzyme name	EC[a] number	Family[b]
Pectin	Pectinesterase	3.1.1.11	CE8
	Exo-poly-α-D-galacturonosidase	3.2.1.82	GH28
	α-L-rhamnosidase	3.2.1.40	GH33
	α-L-rhamnosidase	3.2.1.40	GH78
	Rhamnogalacturonyl hydrolase	3.2.1.172	GH105
	Pectate lyase superfamily protein		GH28
Arabinogalactan	Arabinogalactan endo-β-1,4-galactanase	3.2.1.89	GH53
	Putative Galactan endo-1,6-β-galactosidase	3.2.1.164	GH30
Arabinan (Hemicellulose)	Putative α-L-arabinofuranosidase	3.2.1.55	GH43
	α-L-arabinofuranosidase	3.2.1.55	GH51
	β-L-arabinofuranosidase	3.2.1.185	GH127
	Arabinan endo-1,5-α-L-arabinosidase	3.2.1.99	GH43
Xylan (Hemicellulose)	Xylan 1,4-β-xylosidase	3.2.1.37	GH43
	Xylan 1,3-β-xylosidase	3.2.1.72	
	Acetyl xylan esterase	3.1.1.72	
Mannan (Hemicellulose)	β-mannanase	3.2.1.78	GH26
	Putative mannan endo-1,6-α-mannosidase.	3.2.1.101	GH76
Lichenan	Licheninase	3.2.1.73	GH16
Cellulose	Cellulase	3.2.1.4	GH5
	Cellulose 1,4-β-cellobiosidase	3.2.1.91	GH9
	β-glucosidase	3.2.1.21	GH3
Chitin	Chitinase	3.2.1.14	GH18
	β-N-acetylhexosaminidase	3.2.1.52	GH20
	N-acetylglucosamine kinase	2.7.1.59	
	Putative β-hexosaminidase	3.2.1.52	GH3
Starch	α-amylase	3.2.1.1	GH13
	Putative glucan 1,4-α-glucosidase	3.2.1.3	GH15
	Putative α-amylase	3.2.1.1	GH57
	4-α-glucanotransferase	2.4.1.25	
	Glycogen Phosphorylase	2.4.1.1	GT35
Pullulan	Neopullulanase	3.2.1.135	GH13
Fructan	Fructan β-fructosidase	3.2.1.80	GH32
	β-fructofuranosidase	3.2.1.26	GH32

Enzymes encoded in the *P. saccharofermentans* str. M3/6^T genome were compared against the HMM profile-based database dbCAN using hmmsearch in the HMMER software package. Results were manually curated considering best BLAST hits, alignments, and e-values. In addition, PRIAM profiles were used for classifications.
   [a] Enzyme Commission number.
   [b] CAZy families.

for complete xylan hydrolysis were predicted. Acetyl xylan esterase (EC 3.1.1.72) catalyzes release of acetyl groups from polymeric xylan and two other enzymes successively cleave xylose residues from the non-reducing termini of xylan molecules: xylan 1,4-β-xylosidase (EC 3.2.1.37) and xylan 1,3-β-xylosidase (EC 3.2.1.72).

Mannans are polysaccharides composed of mannose, glucose and galactose residues. The strain M3/6^T genome harbors genes encoding β-mannanases (EC 3.2.1.78) that hydrolyze linear mannan by randomly cleaving its backbone. (1,4)-β-manno-oligomers are released. The strain may also catalyze cleavage of α-(1-6) mannosidic bonds in unbranched mannan oligossacharides producing mannose residues by a putative endo-1,6-α-mannosidase (EC 3.2.1.101).

β-Glucans are polysaccharides consisting of repeated glucose residues linked by β-1,3 and β-1,4 glycosidic bonds. Hydrolysis of β-glucans is mainly carried out by four types of β-glucanases: β-1,3(4)-glucanase (EC 3.2.1.6), laminarinase (E.C. 3.2.1.39), cellulase (E.C. 3.2.1.4), and lichenase (E.C. 3.2.1.73). The strain M3/6^T genome possesses genetic determinants encoding a cellulase and a lichenase. Lichenases catalyze cleavage of β-1-4-D-glucosidic bonds of β-glucans yielding shorter glucans. Subsequently, the

resulting glucans can be hydrolyzed by a cellulase releasing glucose units.

3.2.4. Cellulose

Cellulose is the major polysaccharide of plant cell walls consisting of repeated D-glucose residues linked by β-1,4 glycosidic bonds. The *P. saccharofermentans* str. M3/6^T genome encodes an endoglucanase (EC 3.2.1.4) that catalyzes the cleavage of β-(1-4) glucosidic bonds releasing cellodextrin, which can be further hydrolyzed by 1,4-β-cellobiosidase (EC 3.2.1.91) producing cellobiose. Subsequently, the latter disaccharide is cleaved into glucose monomers by a β-glucosidase (EC 3.2.1.21).

3.2.5. Chitin

Chitin is frequently considered to represent the second most abundant polysaccharide in nature after cellulose. It is present in several organisms such as shrimps, crabs and insects as well as in the cell walls of fungi, yeast and green algae. The polysaccharide is composed of repeating units of β-(1,4)-linked N-acetyl-β-D-glucosamine. Even if chitin is not present in biogas reactors in high amounts, a total of nine genes encoding putative chitinolytic enzymes were identified in *P. saccharofermentans* str. M3/6^T,

including a chitinase (EC 3.2.1.14), β-*N*-acetyl-hexosaminidase (EC 3.2.1.52) and *N*-acetyl-glucosamine kinase (EC 2.7.1.59). Chitinases act on *N*-acetyl-β-ᴅ-glucosamine in chitin yielding chitodextrins, which can be further hydrolyzed by chitinases releasing chitotriose. Subsequently, chitotriose is hydrolyzed to N,N'-diacetylchitobiose by β-*N*-acetyl-hexosaminidase, producing *N*-acetyl-ᴅ-glucosamine.

3.2.6. Starch

Starch is the main storage polysaccharide of higher plants. It is a mixture of two polysaccharides, namely α-amylose and amylopectin. α-Amylose consists of several thousand glucose residues commonly connected *via* α-(1,4) glycosidic bonds. On the other hand, amylopectin is made up of hundreds of shorter linear chains (α-1,4-glucan) with branches of α-(1,6) glycosidic bonds occurring every 24–30 glucose residues. CAZy analyses of the *P. saccharofermentans* str. M3/6[T] genome revealed seven genes encoding enzymes involved in starch degradation, including α-amylase (EC 3.2.1.1), a putative glucan 1,4-α-glucosidase (EC 3.2.1.3), 4-α-glucanotransferase (EC 2.4.1.25), and glycogen phosphorylase (EC 2.4.1.1). The α-amylase acts on the inner part of the starch chain cleaving α-(1,4) glycosidic linkages to release maltose or maltodextrin. The α-(1,4) and α-(1,6) glycosidic bonds of the external glucose residues of the polysaccharide are cleaved by glucan 1,4-α-glucosidase yielding glucose. 4-α-Glucanotransferase cleaves and transfers an 1,4 glycosidic bond present in maltose to a new position in an acceptor molecule releasing a glucose residue. Finally, α-1,4-glycosidic bonds of maltodextrin are cleaved by glycogen phosphorylase to release glucose residues.

3.2.7. Pullulan

Pullulan is a linear polysaccharide frequently described as repeating unit of maltotriose linked by α-(1,6) glycosidic bonds. Nevertheless, also α-(1,4) linkages can occur. Pullulan is one of the exo-polysaccharides produced by the yeast-like fungus *Aureobasidium pullulans* [58,59]. Pullulanases are classified according to their substrate specificities and reaction products. There are five groups: pullulanases type I (PULI), pullulanases type II (PULII), amylopullulanase, isopullulanase and neopullulanase [60]. A gene encoding neopullulanase (EC 3.2.1.135) was identified in *P. saccharofermentans* str. M3/6[T]. The enzyme catalyzes the breakdown of α-(1,4) glycosidic linkages in pullulan to produce maltose, glucose and panose as main end-products [61].

3.2.8. Fructan

Fructan is a polysaccharide composed of fructose residues linked by β-(2,1)- or β-(2,6)-glycosidic bonds. Fructan is classified in three types: inulin (linear structure linked by β-(2,1)-glycosidic bonds), levan (linear structure linked by β-(2,6)-glycosidic bonds) and graminan (branched structure linked by β-(2,1)- and β-(2,6)-glycosidic bonds). Analyses indicated that the *P. saccharofermentans* str. M3/6[T] genome encodes two enzymes involved in fructan degradation, namely fructan β-fructosidase (EC 3.2.1.80) and β-fructofuranosidase (EC 3.2.1.26). Fructan β-fructosidase is able to cleave β-(2,1)- and β-(2,6)-glycosidic bonds in inulin and levan at the non-reducing end releasing sucrose (or fructose residues). The disaccharide sucrose can be hydrolyzed into fructose and glucose by fructofuranosidase. Moreover, thirty-two other genes of strain M3/6[T] were predicted to encode glycoside hydrolases. These enzymes could not be further specified.

With respect to the increasing demand on microbial capabilities regarding the degradation of organic substrates rich in protein components, additionally the potential of *P. saccharofermentans* str. M3/6[T] to decompose proteins was also examined based on its functional genome annotation. The proteolytic system of strain M3/6[T] consists of intracellular and extracellular proteases and transporters (Supplementary file 1: Table S4). The intracellular proteolytic system is composed of Lon- and Clp serine proteases and prolyl oligo-peptidases, which play an essential role in degrading abnormal proteins or proteins participating in regulatory processes. Among the extracellular proteases, strain M3/6[T] encodes aminopeptidases which are exopeptidases acting on N-terminal amino acid residues from small oligopeptides or proteins. These findings are in accordance with the earlier observation that peptone supports anaerobic growth of]*P. saccharofermentans* str. M3/6[T] [31].

3.3. Polysaccharide utilization loci (PUL)

Polysaccharide utilization loci (PUL) were identified in the genomes of *Bacteroidetes* members recovered from rumen or human gut samples [62–65]. Commonly, the genes *sus*D and *sus*C are central elements of PUL clusters. They encode an outer membrane glycan binding protein (SusD) and an outer membrane protein (SusC) involved in transfer of the maltooligosaccharide to the periplasmic space for complete degradation [62,63,65,66]. Genes coding for carbohydrate-active enzymes and transcriptional regulators (e.g. HTCS, ECF-σ/Anti-σ regulators, SusR, AraC, GntR and LacI) were frequently located in close proximity to PUL clusters. Corresponding gene products contribute to the transcriptional regulation and functionality of the multi-component complex [67].

In *P. saccharofermentans* str. M3/6[T], 64 distinct PUL were identified (Fig. 2). Among them, 38 PUL comprise 115 out of 116 GH genes predicted in the genome of the strain. GHs encoded in PUL represent 33 different CAZyme families. The most frequent ones are GH43 and GH2, indicating that corresponding PUL are mainly dedicated to xylan degradation and metabolism of galactose, respectively. At the example of *G*ramella flava JLKT2011, a recent study confirmed involvement of PUL and associated GH genes in xylan and homogalacturonan utilization by exploiting multi-omics data [68]. PUL of *P. saccharofermentans* str. M3/6[T] are also associated with PL, CE and peptidase genes. Likewise, PUL of *Prevotella* species are linked to peptidase genes, suggesting that the encoded complex is involved in protein degradation and peptide transfer across the membrane [64]. Moreover, 25 PUL of strain M3/6[T] are linked to transcriptional regulator genes indicating that the encoded regulators are involved in control of PUL expression. Overall, the strain M3/6[T] seems to be capable to utilize a wide spectrum of complex polysaccharides and proteins.

3.4. Putative fermentation metabolic pathways

P. saccharofermentans str. M3/6[T] encodes a versatile set of carbohydrate-active enzymes indicating that the strain is able to generate energy by fermentation of sugar molecules. To obtain deeper insights into fermentation pathways, the central metabolism of strain M3/6[T], was reconstructed from genome sequence data. In this context, also the corresponding transport mechanisms were considered.

In total, 225 genes encoding transporters for amino acids, peptides, monosaccharides, inorganic and metal ions are present in the genome of strain M3/6[T]. A summary on monosaccharide, amino acid and peptide transporters is given in Supplementary file 2 (Table S1). Five different transporter families encoded in the genome may be involved in uptake of xylose, fucose, arabinose, and maltose. Four other transporter families were predicted to facilitate peptide and/or amino acid transport into the cell.

P. saccharofermentans strain M3/6[T] is able to ferment a variety of mono- and disaccharides producing carbon dioxide, hydrogen, acetate, formate, propionate and isovalerate as end-products of the fermentation process [31]. Glucose, arabinose, cellobiose, fructose,

galactose, lactose, maltose, rhamnose, sucrose, and trehalose were among the substrates that supported growth of strain M3/6[T]. Except for rhamnose and fucose degradation, all genes encoding relevant enzymes involved in utilization of these simple sugar molecules were predicted (Fig. 3 and Supplementary file 1: Table S5). Moreover, metabolic reconstruction also indicated that strain M3/6[T] is able to degrade xylose, lyxose, mannose, and melibiose as well as some amino acids (e.g. proline, alanine, asparagine). The majority of metabolites of corresponding degradation pathways are converted to pyruvate, and subsequently yield fermentation end-products (see above and Supplementary file 1: Table S5). Hence, the strain M3/6[T] is able to produce acids from sugar fermentation suggesting that its function in the biogas production process is associated with the acidogenic phase.

3.5. P. saccharofermentans M3/6[T] harbors two CRISPR-cas systems

Bacteriophages and viruses of micro-eukaryotes are considered to represent the most abundant and genetically diverse group on earth [69–71]. Viral infections are responsible for lysis of microbial cells thereby affecting microbial community structures [70,71]. A recent study indicated that phages can affect biogas microbial communities thus causing profound effects on the performance of the anaerobic digestion process [72]. Obtained results suggested that more than 40% of variation within biogas community compositions might be caused by phages.

Clustered regularly interspaced short palindromic repeats (CRISPR-cas) systems are considered to represent prokaryotic immune systems against foreign DNA molecules and in some cases, also against RNA molecules [73–76]. Briefly, they are composed of a cas operon followed by a leader sequence and arrays of highly conserved short repeat sequences, which are separated by variable sequences (called 'spacers') originating from phage or plasmid DNA. Moreover, cas operons are classified in three major types and subtypes according to their architecture [76,77].

P. saccharofermentas strain M3/6[T] harbors two CRISPR-cas systems (Fig. 3 and Supplementary file S2). The first CRISPR-cas system was classified as type II since it comprises the cas9 gene near cas1 and cas2. A predicted leader region (promoter region) is located downstream of the cas2 gene and features on average an AT-content of 63.3%. The CRISPR array is composed of 39 47-bp direct repeats and 38 spacers of 29–30 bp.

The second CRISPR-cas system belonging to the subtype I-B is composed of an operon of eight genes, a leader region (211 bp featuring an AT-content of 66.98%), and a large CRISPR array of 114 29-bp direct repeats and 113 spacers of 34–39 bp (Fig. 4).

The host retains genome signatures from phages as spacer sequences in CRISPR arrays after phage infection. These can be used to identify corresponding phages based on database searches. All spacer sequences located in P. saccharofermentas str. M3/6[T] CRISPR arrays were compared against all phage genomes in the NCBI nucleotide sequence database applying the BLASTn-short function (BLAST+package) with an e-value threshold of 1.0×10^{-10}. Spacer sequences could not be assigned to any known bacteriophage genome stored in the database indicating that phage diversity is not sufficiently covered in the NCBI database.
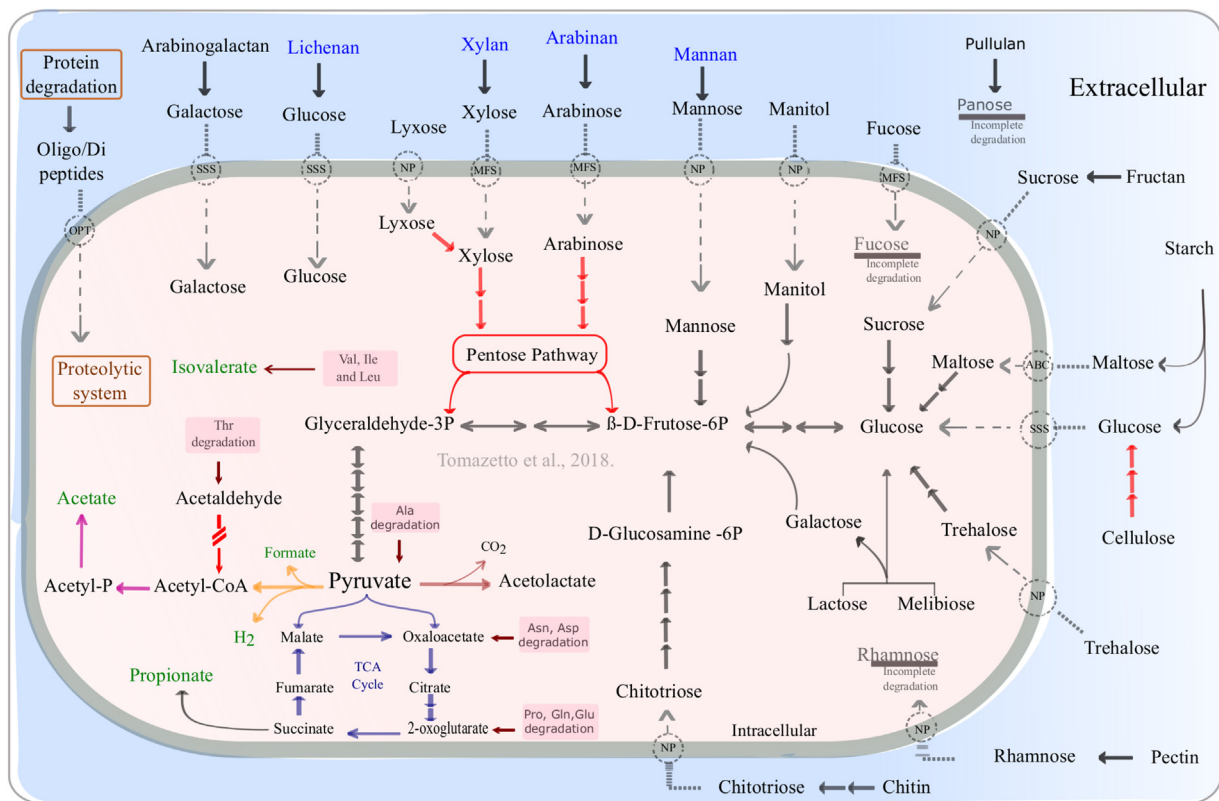

Fig. 3. Schematic overview on carbohydrate and protein degradation pathways based on enzymes predicted from the *P. saccharofermentans* str. M3/6[T] genome sequence. Intracellular and extracellular reactions are separated by the cell envelope. Names in blue denote the hemicellulose polysaccharides. The following intracellular metabolic pathways are shown: glycolysis, pentose phosphate pathway, central pyruvate metabolism and the tricarboxylic acid cycle (TCA). Arrows symbolize enzymatic reactions. Crossed red arrows mark enzymatic reactions for which corresponding enzymes were not predicted in the *P. saccharofermentans* str. M3/6[T] genome. Green metabolites represent fermentation pathway end-products (acetate, propionate, isovalerate, formate, and molecular hydrogen). Abbreviations: Asp, aspartic acid; Gln, glutamine; His, histidine; Ile, isoleucine; Leu, leucine; Thr, threonine; Val, valine; OPT, Oligopeptide Transporter Family proteins; ABC, ABC transporter; SSS, Solute-Sodium Symporter Family proteins; MFS, Major Facilitator Superfamily proteins; NP, no prediction.
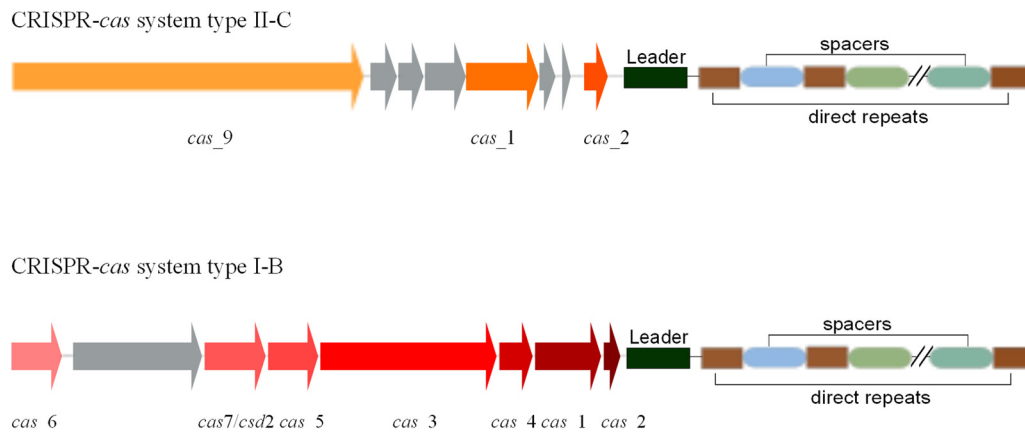
CRISPR-*cas* system type II-C

cas_9 *cas_1* *cas_2*

CRISPR-*cas* system type I-B

cas 6 *cas7/csd2 cas 5* *cas 3* *cas 4 cas 1 cas 2*

Fig. 4. CRISPR-*cas* systems of the *P. saccharofermentans* str. M3/6^T genome. CRISPR-*cas* system type II-C: *cas* operon, eight genes (7460 bp); Leader, 109 bp; CRISPR array (2967 bp), 39 47-bp-direct repeats and 38 spacers of 29–30 bp. CRISPR-*cas* system type I-B: *cas* operon, eight genes (9131 bp); Leader, 211 bp; CRISPR array (7307 bp), 114 29-bp-direct repeats and 113 spacers of 34–39 bp; Coding sequences (CDS) colored in grey encode hypothetical proteins. The coordinates and sequences of the CRISPR-*cas* systems are provided as Supplemental material (File S2).

3.6. Comparative genomic analyses

The *P. saccharofermentans* str. M3/6^T genome was compared to its closest described relative *P. acetatigenes* str. DSM 18083^T that is available in the NCBI nucleotide sequence database.

According to the Integrated Microbial Genomes pipeline (IMG/EM) [78], the strain DSM 18083^T is able to degrade the same types of complex carbohydrates and mono- and disaccharides as predicted for strain M3/6^T. Genes for utilization of pectin, xylan and pullulan are not present in strain DSM 18083^T. Automatic IMG/EM annotation indicated that strain DSM 18083^T possesses 90 PUL gene clusters compared to 64 that are present in strain M3/6^T. However, both strains are able to degrade β-glycan compounds.

The comparative genomics tool EDGAR [44] was applied to calculate the shared core-genome between *P. saccharofermentans* str. M3/6^T and *P. acetatigenes* str. DSM 18083^T. The latter strain was isolated from a reactor treating brewery wastewater (Leibniz-Institute DSMZ GmbH, Braunschweig, Germany), while *P. saccharofermentans* str. M3/6^T originates from a biogas reactor. Therefore, both strains were isolated from biotechnological environments. Nothing is known on the occurrence and dissemination of these strains in natural habitats. Moreover, *P. acetatigenes* str. DSM 18083^T and *P. saccharofermentans* str. M3/6^T represent the only types strains of the genus *Proteiniphilum*. Therefore, comparative genomic analyses are limited. It appeared that both strains share 2482 orthologous genes, representing 73.5% and 68.2% of all genes predicted for strains M3/6T and DSM 18083^T, respectively. Moreover, these strains feature an Average Nucleotide Identity (ANI) value of 91.4% confirming that these bacteria belong to different species [79].

To complement genomic analyses regarding genome architectures, a synteny plot was calculated. As illustrated in Fig. S1 (Supplementary file S1), there are considerable rearrangements, deletions and insertions within the strain M3/6^T genome compared to that of strain DSM 18083^T, suggesting that both strains have undergone diverging evolution for a longer period of time probably promoting adaptation to their specific ecological niches. However, these adaptations cannot be completely evaluated due to missing metadata on natural habitats of these *Proteiniphilum* species.

4. Concluding remarks

Proteiniphilum saccharofermentans str. M3/6^T features a versatile metabolism contributing to substrate hydrolysis and acidogenesis within the biogas process as revealed by systematic genome analyses. The strain possesses a set of genes encoding carbohydrate-active enzymes needed for decomposition of pectin, arabinogalactan, hemicellulose (arabinan, xylan, mannan, β-glucans), cellulose, starch, fructan, chitin and pullulan. Most of these genes as well as peptidase genes are associated with PUL, suggesting their involvement in degradation of a broad spectrum of substrates, especially polysaccharides. Metabolites of carbohydrate decomposition feed fermentation pathways finally yielding volatile organic acids. Therefore, this strain is a promising candidate for the development of inoculant cultures regarding enhancement of biomass decomposition, bioaugmentation processes as well as production of volatile organic acids in anaerobic digestion and may also contribute to increased biogas yields.

Finally, two CRISPR-*cas* systems were detected in the genome of strain M3/6^T, indicating that the bacterium possesses protection mechanisms against invasive foreign DNA elements (e.g., phages or plasmids). CRISPR-spacer sequences of M3/6^T do not match sequences within databases, indicating that *Proteiniphilum* phages and/or plasmids are not very well represented in corresponding databases. However, presence of a CRISPR-*cas* 'immune system' may render the strain insensitive towards infections by specific bacteriophages which also has implications on the performance of the strain in anaerobic digestion processes.

Competing interests

The authors declare that they do not have any competing interests.

Acknowledgements

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.btre.2018.e00254.

References

[1] J.B. Holm-Nielsen, T. Al Seadi, P. Oleskowicz-Popiel, The future of anaerobic digestion and biogas utilization, Bioresour. Technol. 100 (2009) 5478–5484, doi:http://dx.doi.org/10.1016/j.biortech.2008.12.046.

[2] P. Weiland, Biogas production: current state and perspectives, Appl. Microbiol. Biotechnol. 85 (2010) 849–860, doi:http://dx.doi.org/10.1007/s00253-009-2246-7.

[3] B. St-Pierre, A.-D.G. Wright, Comparative metagenomic analysis of bacterial populations in three full-scale mesophilic anaerobic manure digesters, Appl. Microbiol. Biotechnol. 98 (2013) 2709–2717, doi:http://dx.doi.org/10.1007/s00253-013-5220-3.

[4] A. Li, Y. Chu, X. Wang, L. Ren, J. Yu, X. Liu, X. Yan, J. Yan, L. Zhang, S. Wu, S. Li, A pyrosequencing-based metagenomic study of methane-producing microbial comunity in solid-stat biogas reactor, Biotechnol. Fuels 6 (2013) 3.

[5] Y. Stolze, M. Zakrzewski, I. Maus, F. Eikmeyer, S. Jaenicke, N. Rottmann, C. Siebner, A. Pühler, A. Schlüter, Comparative metagenomics of biogas-producing microbial communities from production-scale biogas plants operating under wet or dry fermentation conditions, Biotechnol. Biofuels 8 (2015), doi:http://dx.doi.org/10.1186/s13068-014-0193-8.

[6] S. Campanaro, L. Treu, P.G. Kougias, D. De Francisci, G. Valle, I. Angelidaki, Metagenomic analysis and functional characterization of the biogas microbiome using high throughput shotgun sequencing and a novel binning strategy, Biotechnol. Biofuels 9 (2016) 26, doi:http://dx.doi.org/10.1186/s13068-016-0441-1.

[7] I. Maus, D.E. Koeck, K.G. Cibis, et al., Unraveling the microbiome of a thermophilic biogas plant by metagenome and metatranscriptome analysis complemented by characterization of bacterial and archaeal isolates, Biotechnol. Biofuels (2016) 868–874, doi:http://dx.doi.org/10.1007/s10745-006-9094-1.

[8] S. Theuerl, F. Kohrs, D. Benndorf, I. Maus, D. Wibberg, A. Schlüter, R. Kausmann, M. Heiermann, E. Rapp, U. Reichl, A. Pühler, M. Klocke, Community shifts in a well-operating agricultural biogas plant: how process variations are handled by the microbiome, Appl. Microbiol. Biotechnol. 99 (18) (2015), doi:http://dx.doi.org/10.1007/s00253-015-6627-9.

[9] A. Fontana, V. Patrone, E. Puglisi, L. Morelli, D. Bassi, M. Garuti, L. Rossi, F. Cappa, Effects of geographic area, feedstock, temperature, and operating time on microbial communities of six full-scale biogas plants, Bioresour. Technol. 218 (2016) 980–990, doi:http://dx.doi.org/10.1016/j.biortech.2016.07.058.

[10] Y. Stolze, A. Bremges, M. Rumming, C. Henke, I. Maus, A. Pühler, A. Sczyrba, A. Schlüter, Identification and genome reconstruction of abundant distinct taxa in microbiomes from one thermophilic and three mesophilic production-scale biogas plants, Biotechnol. Biofuels 9 (2016) 156, doi:http://dx.doi.org/10.1186/s13068-016-0565-3.

[11] F.A. Rainey, Class II. Clostridia class. nov., Bergey's Manual of Systematic Bacteriology, vol. 3, Springer, New York, 2009, pp. 736–1298.

[12] N.R. Krieg, et al., Phylum XIV. Bacteroidetes phyl. nov, Bergey's Manual of Systematic Bacteriology, vol. 4, Springer, 2010, pp. 25.

[13] L. Krause, N.N. Diaz, R.A. Edwards, K.-H. Gartemann, H. Krömeke, H. Neuweger, A. Pühler, K.J. Runte, A. Schlüter, J. Stoye, R. Szczepanowski, A. Tauch, A. Goesmann, Taxonomic composition and gene content of a methane-producing microbial community isolated from a biogas reactor, J. Biotechnol. 136 (2008) 91–101, doi:http://dx.doi.org/10.1016/j.jbiotec.2008.06.003.

[14] A. Schlüter, T. Bekel, N.N. Diaz, M. Dondrup, R. Eichenlaub, K.-H. Gartemann, I. Krahn, L. Krause, H. Krömeke, O. Kruse, J.H. Mussgnug, H. Neuweger, K. Niehaus, A. Pühler, K.J. Runte, R. Szczepanowski, A. Tauch, A. Tilker, P. Viehöver, A. Goesmann, The metagenome of a biogas-producing microbial community of a production-scale biogas plant fermenter analysed by the 454-pyrosequencing technology, J. Biotechnol. 136 (2008) 77–90, doi:http://dx.doi.org/10.1016/j.jbiotec.2008.05.008.

[15] S. Jaenicke, C. Ander, T. Bekel, R. Bisdorf, M. Dröge, K.-H. Gartemann, S. Jünemann, O. Kaiser, L. Krause, F. Tille, M. Zakrzewski, A. Pühler, A. Schlüter, A. Goesmann, Comparative and joint analysis of two metagenomic datasets from a biogas fermenter obtained by 454-pyrosequencing, PLoS One 6 (2011) e14519, doi:http://dx.doi.org/10.1371/journal.pone.0014519.

[16] A. Rademacher, M. Zakrzewski, A. Schlüter, M. Schönberg, R. Szczepanowski, A. Goesmann, A. Pühler, M. Klocke, Characterization of microbial biofilms in a thermophilic biogas system by high-throughput metagenome sequencing, FEMS Microbiol. Ecol. 79 (2012) 785–799, doi:http://dx.doi.org/10.1111/j.1574-6941.2011.01265.x.

[17] A. Hanreich, U. Schimpf, M. Zakrzewski, A. Schlüter, D. Benndorf, R. Heyer, E. Rapp, A. Pühler, U. Reichl, M. Klocke, Metagenome and metaproteome analyses of microbial communities in mesophilic biogas-producing anaerobic batch fermentations indicate concerted plant carbohydrate degradation, Syst. Appl. Microbiol. 36 (2013) 330–338, doi:http://dx.doi.org/10.1016/j.syapm.2013.03.006.

[18] C. Sundberg, W.A. Al-Soud, M. Larsson, E. Alm, S.S. Yekta, B.H. Svensson, S.J. Sørensen, A. Karlsson, 454 pyrosequencing analyses of bacterial and archaeal richness in 21 full-scale biogas digesters, FEMS Microbiol. Ecol. 85 (2013) 612–626, doi:http://dx.doi.org/10.1111/1574-6941.12148.

[19] R. Wirth, E. Kovács, G. Maróti, Z. Bagi, G. Rákhely, K.L. Kovács, Characterization of a biogas-producing microbial community by short-read next generation DNA sequencing, Biotechnol. Biofuels 5 (2012) 41, doi:http://dx.doi.org/10.1186/1754-6834-5-41.

[20] L. Treu, P.G. Kougias, S. Campanaro, I. Bassani, I. Angelidaki, Deeper insight into the structure of the anaerobic digestion microbial community: the biogas

[21] microbiome database is expanded with 157 new genomes, Bioresour. Technol. 216 (2016) 260–266, doi:http://dx.doi.org/10.1016/j.biortech.2016.05.081.

[21] M. Pop, Genome assembly reborn: recent computational challenges, Brief. Bioinform. 10 (2009) 354–366, doi:http://dx.doi.org/10.1093/bib/bbp026.

[22] H.B. Nielsen, M. Almeida, A.S. Juncker, S. Rasmussen, J. Li, S. Sunagawa, Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes, Nat. Biotechnol. 32 (2014) 822–828, doi:http://dx.doi.org/10.1038/nbt.2939.

[23] J.C. Clemente, E.C. Pehrsson, M.J. Blaser, K. Sandhu, Z. Gao, B. Wang, M. Magris, G. Hidalgo, M. Contreras, Ó. Noya-alarcón, O. Lander, J. Mcdonald, M. Cox, J. Walter, P.L. Oh, J.F. Ruiz, S. Rodriguez, N. Shen, S.J. Song, J. Metcalf, The microbiome of uncontacted amerindians, Sci Adv. 1 (April (3)) (2015) e1500183.

[24] R. Mondav, B.J. Woodcroft, E.-H. Kim, C.K. McCalley, S.B. Hodgkins, P.M. Crill, J. Chanton, G.B. Hurst, N.C. VerBerkmoes, S.R. Saleska, P. Hugenholtz, V.I. Rich, G. W. Tyson, Discovery of a novel methanogen prevalent in thawing permafrost, Nat. Commun. 5 (2014) 1–7, doi:http://dx.doi.org/10.1038/ncomms4212.

[25] R.A. White III, E.M. Bottos, T.R. Chowdhury, J.D. Zucker, C.J. Brislawn, C.D. Nicora, S.J. Fansler, K.R. Glaesemann, K. Glass, J.K. Jansson, Moleculo long-read sequencing facilitates assembly and genomic binning from complex soil metagenomes, mSystems 1 (2016) 1–15, doi:http://dx.doi.org/10.1128/mSystems.00045-16.

[26] B.M. Satinsky, C.S. Fortunato, M. Doherty, C.B. Smith, S. Sharma, N.D. Ward, A.V. Krusche, P.L. Yager, J.E. Richey, M.A. Moran, B.C. Crump, Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011, Microbiome 3 (2015) 39, doi:http://dx.doi.org/10.1186/s40168-015-0099-0.

[27] D.E. Koeck, D. Wibberg, T. Koellmeier, J. Blom, S. Jaenicke, A. Winkler, A. Albersmeier, V.V. Zverlov, A. Pühler, W.H. Schwarz, A. Schlüter, Draft genome sequence of the cellulolytic *Clostridium thermocellum* wild-type strain BC1 playing a role in cellulosic biomass degradation, J. Biotechnol. 168 (2013) 62–63, doi:http://dx.doi.org/10.1016/j.jbiotec.2013.08.011.

[28] G. Tomazetto, S. Hahnke, I. Maus, D. Wibberg, A. Pühler, A. Schlüter, M. Klocke, Complete genome sequence of *Peptoniphilus* sp. strain ING2-D1G isolated from a mesophilic lab-scale completely stirred tank reactor utilizing maize silage in co-digestion with pig and cattle manure for biomethanation, J. Biotechnol. 192 (2014) 59–61, doi:http://dx.doi.org/10.1016/j.jbiotec.2014.09.011.

[29] S. Hahnke, I. Maus, D. Wibberg, G. Tomazetto, A. Pühler, M. Klocke, A. Schlüter, Complete genome sequence of the novel *Porphyromonadaceae* bacterium strain ING2-E5B isolated from a mesophilic lab-scale biogas reactor, J. Biotechnol. 193 (2015) 34–36, doi:http://dx.doi.org/10.1016/j.jbiotec.2014.11.010.

[30] D.E. Koeck, I. Maus, D. Wibberg, A. Winkler, V.V. Zverlov, W. Liebl, A. Pühler, W. H. Schwarz, A. Schlüter, Draft genome sequence of *Herbinix hemicellulosilytica* T3/55T, a new thermophilic cellulose degrading bacterium isolated from a thermophilic biogas reactor, J. Biotechnol. 214 (2015), doi:http://dx.doi.org/10.1016/j.jbiotec.2015.07.022.

[31] S. Hahnke, T. Langer, D.E. Koeck, M. Klocke, Description of *Proteiniphilum saccharofermentans* sp. nov., *Petrimonas mucosa* sp. nov. and *Fermentimonas caenicola* gen. nov., sp. nov. isolated from mesophilic lab-scale biogas reactors and emended description of the genus Proteiniphilum, Int. J. Syst. Evol. Microbiol. (2016) 1–10, doi:http://dx.doi.org/10.1099/ijsem.0.000902.

[32] D. Gordon, C. Abajian, P. Green, Consed: a graphical tool for sequence finishing, Genome Res. 8 (1998) 195–202, doi:http://dx.doi.org/10.1101/gr.8.3.195.

[33] G. Tomazetto, S. Hahnke, D.E. Koeck, D. Wibberg, I. Maus, A. Pühler, M. Klocke, A. Schlüter, Complete genome analysis of *Clostridium bornimense* strain M2/40^T: a new acidogenic Clostridium species isolated from a mesophilic two-phase laboratory-scale biogas reactor, J. Biotechnol. 232 (2016), doi:http://dx.doi.org/10.1016/j.jbiotec.2015.08.001.

[34] F. Meyer, A. Goesmann, A.C. McHardy, D. Bartels, T. Bekel, J. Clausen, J. Kalinowski, B. Linke, O. Rupp, R. Giegerich, A. Pühler, GenDB—an open source genome annotation system for prokaryote genomes, Nucleic Acids Res. 31 (2003) 2187–2195, doi:http://dx.doi.org/10.1093/nar/gkg312.

[35] K. Lagesen, P. Hallin, E.A. Rødland, H.H. Stærfeldt, T. Rognes, D.W. Ussery, RNAmmer: consistent and rapid annotation of ribosomal RNA genes, Nucleic Acids Res. 35 (2007) 3100–3108, doi:http://dx.doi.org/10.1093/nar/gkm160.

[36] B.E. Suzek, M.D. Ermolaeva, M. Schreiber, S.L. Salzberg, A probabilistic method for identifying start codons in bacterial genomes, Bioinformatics 17 (2001) 1123–1130, doi:http://dx.doi.org/10.1093/bioinformatics/17.12.1123.

[37] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for inproved detection of transfer RNA genes in genomic sequence, Nucleic Acids Res. 25 (1997) 955–964, doi:http://dx.doi.org/10.1093/nar/25.5.0955.

[38] T.N. Petersen, S. Brunak, G. Von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, Nat. Methods 8 (2011) 785–786, doi:http://dx.doi.org/10.1038/nmeth.1701 Nat. Publ. Gr..

[39] A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden markov model: application to complete genomes, J. Mol. Biol. 305 (2001) 567–580, doi:http://dx.doi.org/10.1006/jmbi.2000.4315.

[40] Y. Zhou, Y. Liang, K.H. Lynch, J.J. Dennis, D.S. Wishart, PHAST: a fast phage search tool, Nucleic Acids Res. 39 (2011) 1–6, doi:http://dx.doi.org/10.1093/nar/gkr485.

[41] B.K. Dhillon, M.R. Laird, J.A. Shay, G.L. Winsor, R. Lo, F. Nizam, S.K. Pereira, N. Waglechner, A.G. McArthur, M.G.I. Langille, F.S.L. Brinkman, Island viewer 3: more flexible, interactive genomic island discovery, visualization and analysis, Nucleic Acids Res. 43 (2015) 104–108, doi:http://dx.doi.org/10.1093/nar/gkv401.

[42] I. Grissa, G. Vergnaud, C. Pourcel, CRISPRFinder: a web tool to identify clustered regulary interspaced short palindromic repeats, Nucleic Acids Res. 35 (2007) 52–57, doi:http://dx.doi.org/10.1093/nar/gkn228.

[43] I. Grissa, G. Vergnaud, C. Pourcel, The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats, BMC Bioinform. 8 (2007) 172, doi:http://dx.doi.org/10.1186/1471-2105-8-172.

[44] J. Blom, S.P. Albaum, D. Doppmeier, A. Pühler, F.-J. Vorhölter, M. Zakrzewski, A. Goesmann, EDGAR: a software framework for the comparative analysis of prokaryotic genomes, BMC Bioinform. 10 (2009) 154, doi:http://dx.doi.org/10.1186/1471-2105-10-154.

[45] Y. Yin, X. Mao, J. Yang, X. Chen, F. Mao, Y. Xu, dbCAN: a web resource for automated carbohydrate-active enzyme annotation, Nucleic Acids Res. 40 (2012) W445–51, doi:http://dx.doi.org/10.1093/nar/gks479.

[46] R.D. Finn, J. Clements, S.R. Eddy, HMMER web server: interactive sequence similarity searching, Nucleic Acids Res. 39 (2011) W29–37, doi:http://dx.doi.org/10.1093/nar/gkr367.

[47] C. Claudel-Renard, C. Chevalet, T. Faraut, D. Kahn, Enzyme-specific profiles for genome annotation: PRIAM, Nucleic Acids Res. 31 (2003) 6633–6639, doi:http://dx.doi.org/10.1093/nar/gkg847.

[48] P.D. Karp, S.M. Paley, M. Krummenacker, M. Latendresse, J.M. Dale, T.J. Lee, P. Kaipa, F. Gilham, A. Spaulding, L. Popescu, T. Altman, I. Paulsen, I.M. Keseler, R. Caspi, Pathway tools version 13.0: integrated software for pathway/genome informatics and systems biology, Brief. Bioinform. 11 (2009) 40–79, doi:http://dx.doi.org/10.1093/bib/bbp043.

[49] M.H. Saier, V.S. Reddy, B.V. Tsu, M.S. Ahmed, C. Li, G. Moreno-Hagelsieb, The transporter classification database (TCDB): recent advances, Nucleic Acids Res. 44 (2016) D372–D379, doi:http://dx.doi.org/10.1093/nar/gkv1103.

[50] R. Wilkinson, B. Wiedenheft, A CRISPR method for genome engineering, F1000Prime Rep. 6 (2014) 3, doi:http://dx.doi.org/10.12703/P6-3.

[51] R.H. Doi, Cellulases of mesophilic microorganisms: cellulosome and noncellulosome producers, Ann. N. Y. Acad. Sci. 1125 (2008) 267–279, doi:http://dx.doi.org/10.1196/annals.1419.002.

[52] D.E. Koeck, A. Pechtl, V.V. Zverlov, W.H. Schwarz, Genomics of cellulolytic bacteria, Curr. Opin. Biotechnol. 29C (2014) 171–183, doi:http://dx.doi.org/10.1016/j.copbio.2014.07.002.

[53] V.V. Zverlov, W.H. Schwarz, Bacterial cellulose hydrolysis in anaerobic environmental subsystems—Clostridium thermocellum and Clostridium stercorarium, thermophilic plant-fiber degraders, Ann. N. Y. Acad. Sci. 1125 (2008) 298–307, doi:http://dx.doi.org/10.1196/annals.1419.008.

[54] B. Klocke, M. Mundt, K. Sontag, M. Schönberg, M. Linke, Microbial community structures within a two-stage biogas reactor system with bioleaching of rye silage, in: Leibniz-Institut Potsdam-Bornim (Ed.), Wieviel Biogas steckt in Pflanzen? Öffentliches Abschlusssymposium zu den Forschungsergebnissen des "Biogas-Crops-Network" (BCN), Leibniz-Institut für Agrartechnik, Potsdam-Bornim, 2009, pp. 126–139 (ISSN: 0947–7314).

[55] K. Kampmann, S. Ratering, I. Kramer, M. Schmidt, W. Zerr, S. Schnell, Unexpected stability of Bacteroidetes and Firmicutes communities in laboratory biogas reactors fed with different defined substrates, Appl. Environ. Microbiol. 78 (2012) 2106–2119, doi:http://dx.doi.org/10.1128/AEM.06394-11.

[56] A. Rademacher, C. Nolte, M. Schönberg, M. Klocke, Temperature increases from 55 to 75 °C in a two-phase biogas reactor result in fundamental alterations within the bacterial and archaeal community structure, Appl. Microbiol. Biotechnol. 96 (2012) 565–576, doi:http://dx.doi.org/10.1007/s00253-012-4348-x.

[57] B.I. Cantarel, P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics, Nucleic Acids Res. 37 (2009) 233–238, doi:http://dx.doi.org/10.1093/nar/gkn663.

[58] R.S. Singh, G.K. Saini, Pullulan-hyperproducing color variant strain of Aureobasidium pullulans FB-1 newly isolated from phylloplane of Ficus sp, Bioresour. Technol. 99 (2008) 3896–3899, doi:http://dx.doi.org/10.1016/j.biortech.2007.08.003.

[59] R.S. Singh, H. Singh, G.K. Saini, Response surface optimization of the critical medium components for pullulan production by Aureobasidium pullulans FB-1, Appl. Biochem. Biotechnol. 152 (2009) 42–53, doi:http://dx.doi.org/10.1007/s12010-008-8180-9.

[60] S.L. Hii, J.S. Tan, T.C. Ling, A. Bin Ariff, Pullulanase: role in starch hydrolysis and potential industrial applications, Enzyme Res. 2012 (2012), doi:http://dx.doi.org/10.1155/2012/921362.

[61] T. Imanaka, T. Kuriki, Pattern of action of Bacillus stearothermophilus neopullulanase on pullulan, J. Bacteriol. 171 (1989) 369–374.

[62] A.E. Naas, A.K. Mackenzie, J. Mravec, J. Schückel, W.G.T. Willats, V.G.H. Eijsink, P.B. Pope, Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation? MBio 5 (2014) 4–9, doi:http://dx.doi.org/10.1128/mBio.01401-14.

[63] C.P. Rosewarne, P.B. Pope, J.L. Cheung, M. Morrison, Analysis of the bovine rumen microbiome reveals a diversity of Sus-like polysaccharide utilization loci from the bacterial phylum Bacteroidetes, J. Ind. Microbiol. Biotechnol. 41 (2014) 601–606, doi:http://dx.doi.org/10.1007/s10295-013-1395-y.

[64] T. Accetto, G. Avgustin, Polysaccharide utilization locus and CAZYme genome repertoires reveal diverse ecological adaptation of Prevotella species, Syst. Appl. Microbiol. 38 (2015) 453–461, doi:http://dx.doi.org/10.1016/j.syapm.2015.07.007.

[65] A.K. Mackenzie, A.E. Naas, S.K. Kracun, J. Schückel, J.U. Fangel, J.W. Agger, W.G.T. Willats, V.G.H. Eijsink, P.B. Pope, A polysaccharide utilization locus from an uncultured Bacteroidetes phylotype suggests ecological adaptation and substrate versatility, Appl. Environ. Microbiol. 81 (2015) 187–195, doi:http://dx.doi.org/10.1128/AEM.02858-14.

[66] E.C. Martens, N.M. Koropatkin, T.J. Smith, J.I. Gordon, Complex glycan catabolism by the human gut microbiota: the Bacteroidetes sus-like paradigm, J. Biol. Chem. 284 (2009) 24673–24677, doi:http://dx.doi.org/10.1074/jbc.R109.022848.

[67] E.C. Martens, E.C. Lowe, H. Chiang, N.A. Pudlo, M. Wu, P. Nathan, D.W. Abbott, B. Henrissat, H.J. Gilbert, D.N. Bolam, J.I. Gordon, Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts, PLoS Biol. 9 (2011), doi:http://dx.doi.org/10.1371/journal.pbio.1001221.

[68] K. Tang, Y. Lin, Y. Han, N. Jiao, Characterization of potential polysaccharide utilization systems in the marine Bacteroidetes Gramella flava JLT2011 using a multi-omics approach, Front. Microbiol. 8 (2017) 1–13, doi:http://dx.doi.org/10.3389/fmicb.2017.00220.

[69] J.S. Weitz, T. Poisot, J.R. Meyer, C.O. Flores, S. Valverde, M.B. Sullivan, M.E. Hochberg, Phage-bacteria infection networks, Trends Microbiol. 21 (2013) 82–91, doi:http://dx.doi.org/10.1016/j.tim.2012.11.003.

[70] A.A. Pratama, J.D. van Elsas, The "Neglected" soil virome—potential role and impact, Trends Microbiol. (2018) 1–14, doi:http://dx.doi.org/10.1016/j.tim.2017.12.004.

[71] C.V. Tikhe, C. Husseneder, Metavirome sequencing of the termite gut reveals the presence of an unexplored bacteriophage community, Front. Microbiol. 8 (2018) 1–16, doi:http://dx.doi.org/10.3389/fmicb.2017.02548.

[72] J. Zhang, Q. Gao, Q. Zhang, T. Wang, H. Yue, L. Wu, J. Shi, Z. Qin, J. Zhou, J. Zuo, Y. Yang, Bacteriophage-prokaryote dynamics and interaction within anaerobic digestion processes across time and space, Microbiome 5 (2017) 57, doi:http://dx.doi.org/10.1186/s40168-017-0272-8.

[73] R. Jansen, J.D.A. Van Embden, W. Gaastra, L.M. Schouls, Identification of genes that are associated with DNA repeats in prokaryotes, Mol. Microbiol. 43 (2002) 1565–1575.

[74] A. Bolotin, B. Quinquis, A. Sorokin, S.D. Ehrlich, Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin, Microbiology (2005) 2551–2561, doi:http://dx.doi.org/10.1099/mic.0.28048-0.

[75] G. Gasiunas, T. Sinkunas, V. Siksnys, Molecular mechanisms of CRISPR-mediated microbial immunity, Cell. Mol. Life Sci. 71 (2014) 449–465, doi:http://dx.doi.org/10.1007/s00018-013-1438-6.

[76] K.S. Makarova, Y.I. Wolf, O.S. Alkhnbashi, F. Costa, S.A. Shah, S.J. Saunders, R. Barrangou, S.J.J. Brouns, E. Charpentier, D.H. Haft, P. Horvath, S. Moineau, F.J.M. Mojica, R.M. Terns, M.P. Terns, M.F. White, A.F. Yakunin, R.A. Garrett, J. van der Oost, R. Backofen, E.V. Koonin, An updated evolutionary classification of CRISPR-Cas systems, Nat. Rev. Microbiol. 13 (2015) 722–736, doi:http://dx.doi.org/10.1038/nrmicro3569 Nat. Publ. Gr..

[77] K.S. Makarova, S.J.J. Brouns, P. Horvath, D.F. Sas, Y.I. Wolf, Evolution and classification of the CRISPR-Cas systems, Nat. Rev. Microbiol. 9 (2012) 467–477, doi:http://dx.doi.org/10.1038/nrmicro2577.Evolution.

[78] V.M. Markowitz, I.-M. Chen, K. Chu, E. Szeto, K. Palaniappan, Y. Grechkin, A. Ratner, B. Jacob, A. Pati, M. Huntemann, K. Liolios, I. Anderson, K. Mavromatis, N.N. Ivanova, N.C. Kyrpides, IMG/M: the integrated metagenome data management and comparative analysis system, Nucleic Acids Res. 40 (2011) 123–129, doi:http://dx.doi.org/10.1093/nar/gkr975.

[79] K.T. Konstantinidis, J.M. Tiedje, Genomic insights that advance the species definition for prokaryotes, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 2567–2572, doi:http://dx.doi.org/10.1073/pnas.0409727102.

[80] I. Maus, A. Bremges, Y. Stolze, S. Hahnke, K.G. Cibis, D.E. Koeck, Y.S. Kim, J. Kreubel, J. Hassa, D. Wibberg, A. Weimann, S. Off, R. Stantscheff, V.V. Zverlov, W.H. Schwarz, H. König, W. Liebl, P. Scherer, A.C. McHardy, A. Sczyrba, M. Klocke, A. Pühler, A. Schlüter, Genomics and prevalence of bacterial and archaeal isolates from biogas-producing microbiomes, Biotechnol. Biofuels 10 (1) (2018) 264.