Biotechnology Reports 18 (2018) e00254

Contents lists available at ScienceDirect

Biotechnology Reports

journal homepage: www.elsevier.com/locate/btre

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

ARTICLE INFO

Article history:
Received 4 July 2017
Received in revised form 26 April 2018
Accepted 26 April 2018
Available online xxx

Keywords:
Carbohydrate-active enzymes
Polysaccharide utilization loci
Anaerobic digestion
Biomethanation
Metabolic pathway reconstruction
Bioconversion

ABSTRACT

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.

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

* Corresponding authors.
E-mail address: dwibberg@cebitec.uni-bielefeld.de (D. Wibberg).

https://doi.org/10.1016/j.btre.2018.e00254

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 NucleoSpin® 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).