Journal of Biotechnology 231 (2016) 268–279

Contents lists available at ScienceDirect

Journal of Biotechnology

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

An integrated metagenome and -proteome analysis of the microbial community residing in a biogas production plant

Vera Ortseifen[a,b], Yvonne Stolze[a], Irena Maus[a], Alexander Sczyrba[c], Andreas Bremges[c], Stefan P. Albaum[d], Sebastian Jaenicke[e], Jochen Fracowiak[f], Alfred Pühler[a], Andreas Schlüter[a,*]

[a] Genome Research of Industrial Microorganisms, Center for Biotechnology (CeBiTec), Bielefeld University, Universitätsstraße 27, 33615 Bielefeld, Germany
[b] Microbial Genomics and Biotechnology, Center for Biotechnology (CeBiTec), Bielefeld University, Universitätsstraße 27, 33615 Bielefeld, Germany
[c] Computational Metagenomics, Center for Biotechnology (CeBiTec), Bielefeld University, Universitätsstraße 27, 33615 Bielefeld, Germany
[d] Bioinformatics Resource Facility, Center for Biotechnology (CeBiTec), Bielefeld University, Universitätsstraße 27, 33615 Bielefeld, Germany
[e] Bioinformatics and Systems Biology, Justus Liebig University, 35392 Giessen, Germany
[f] Fraunhofer Institute for Molecular Biology and Applied Ecology, Department of Bioresources, Winchester Strasse, D-35394 Giessen, Germany

ARTICLE INFO

Article history:
Received 18 March 2016
Received in revised form 8 June 2016
Accepted 12 June 2016
Available online 14 June 2016

Keywords:
Biogas microbial community
Integrated metagenome/-proteome study
Database impact on protein identification
Contig context information
Taxonomic profile

ABSTRACT

To study the metaproteome of a biogas-producing microbial community, fermentation samples were taken from an agricultural biogas plant for microbial cell and protein extraction and corresponding metagenome analyses. Based on metagenome sequence data, taxonomic community profiling was performed to elucidate the composition of bacterial and archaeal sub-communities. The community's cytosolic metaproteome was represented in a 2D-PAGE approach. Metaproteome databases for protein identification were compiled based on the assembled metagenome sequence dataset for the biogas plant analyzed and non-corresponding biogas metagenomes. Protein identification results revealed that the corresponding biogas protein database facilitated the highest identification rate followed by other biogas-specific databases, whereas common public databases yielded insufficient identification rates. Proteins of the biogas microbiome identified as highly abundant were assigned to the pathways involved in methanogenesis, transport and carbon metabolism. Moreover, the integrated metagenome/-proteome approach enabled the examination of genetic-context information for genes encoding identified proteins by studying neighboring genes on the corresponding contig. Exemplarily, this approach led to the identification of a *Methanoculleus* sp. contig encoding 16 methanogenesis-related gene products, three of which were also detected as abundant proteins within the community's metaproteome. Thus, metagenome contigs provide additional information on the genetic environment of identified abundant proteins.

1. Introduction

Significant progress has been made regarding the analysis of complex microbial communities. This has been achieved by recent developments in metagenome research such as new high-throughput sequencing techniques and advanced bioinformatics tools for data interpretation (Tanca et al., 2013). Metagenome sequence data provide information about the genetic repertoire of all microorganisms in an environment in a culture-independent approach (Kolmeder and de Vos, 2014). However, the question concerning gene expression leading to distinct metabolic activities cannot be resolved by metagenome analyses (Heyer et al., 2013). Therefore, the goal of metatranscriptome and -proteome analyses is to link genetic information to function. Metatranscriptomics based on next generation sequencing arose as a powerful method that could provide detailed insight into gene transcription reflecting microbial activity in a mixed community (Alberti et al., 2014). There are additional processes interfering at the post-transcriptional level that can be assessed by methods addressing the metaproteome of microbial communities (Hettich et al., 2013). Moreover, the characterization of metaproteomic datasets complements the knowledge of metabolic activities and the linkage to microbial communities and their ecological functions (Wang et al., 2014). The metaproteome approach has been applied for diverse environments, e.g. soil and sediments, the human intestinal tract and oral cavity as well

* Corresponding author at: Center for Biotechnology, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany.
E-mail address: aschluet@CeBiTec.Uni-Bielefeld.de (A. Schlüter).

http://dx.doi.org/10.1016/j.jbiotec.2016.06.014
0168-1656/

as bioengineered systems. One major challenge already appeared in one of the first metaproteome studies. Schulze et al. (2005) analyzed the proteins isolated from dissolved organic matter of four different environments. Only few proteins could be identified due to the lack of corresponding metagenome data (Schulze et al., 2005; Heyer et al., 2015). Even analyzing protein expression in natural and bioengineered systems such as a laboratory-scale activated sludge system resulted in poor protein identification rates because of missing genomic information (Wilmes and Bond, 2004). As a result, different studies focused on integrated approaches using corresponding metagenome datasets (Lauro et al., 2011; Wilmes et al., 2008). An integrative metagenomics/-proteomics approach has recently been applied to compare healthy and disease conditions of patients suffering from Crohn's disease regarding the intestinal-tract microbiome (Erickson et al., 2012). Likewise, also non-annotated, unassembled metagenome data could be used as basis for the compilation of a specific database for metaproteome analyses (Rooijers et al., 2011). Moreover, a recently published study followed up on the question whether the application of different protein databases for protein identification impacts metaproteome analyses (Tanca et al., 2013). Hereby, not only appropriate protein databases deduced from corresponding metagenome datasets were generated, but also public databases, namely NCBI, SwissProt and TrEMBL parsed at different taxonomic levels were used and compared. Tanca et al. (2013) could confirm a considerable impact of database selection, also for the trustworthiness of the metaproteome results.

High-throughput omics-technologies help to characterize biogas communities and relationships between community members in detail. Metagenome studies were carried out for biogas-producing microbial communities to uncover their taxonomic profiles and functional potential. First, the community composition was determined by the construction of 16S-rRNA gene clone libraries and subsequent sequencing of corresponding amplicons (Huang et al., 2002; Klocke et al., 2007; McHugh et al., 2003; Mladenovska et al., 2003). Moreover, the development of next generation sequencing technologies offers the possibility of whole metagenome shotgun analyses (Stolze et al., 2015; Jaenicke et al., 2011b; Kröber et al., 2009; Schlüter et al., 2008). Based on the metagenome sequence information, analyses like functional assignments and taxonomic profiling could be performed. Complementary, high-throughput metatranscriptome sequencing was used for profiling of the metabolically active community from a production-scale biogas plant (Zakrzewski et al., 2012). Obtained results among others indicated a high transcriptional activity of archaeal species. Also, metaproteome analyses of microbial communities involved in biogas-producing fermentations were performed (Hanreich et al., 2013, 2012; Heyer et al., 2013).

Here, an integrative metagenome/-proteome study was carried out to investigate a maize-degrading microbial community in batch fermentation. Deeply sequenced metagenome datasets of biogas communities from production-scale biogas plants were exploited for the compilation of biogas-specific protein databases to improve the identification of proteins in a metaproteomics approach for one of the biogas plants analyzed. Protein identification results were compared to those obtained by applying public protein databases or heterologous biogas protein databases. Highly expressed proteins within the microbial biogas community were interpreted in the context of pathways important for the biogas production process. Due to the availability of deeply sequenced metagenome data for the biogas plant analyzed, assembled contigs provided context information for genes encoding identified highly expressed proteins. Integrative metagenome/-proteome analysis is discussed as a preferable approach for functional studies addressing complex microbial communities.

2. Material and methods

2.1. Sampling of a production-scale biogas plant and DNA extraction

At four production-scale biogas plant sites a sample of the fermenter digestate was taken directly from the main fermenter of the plants, respectively. Biogas plant 1, 2 and 3 operate under mesophilic conditions, whereas biogas plant 4 (BGP 4) runs thermophilic at 54 °C. Of those four biogas plants BGP 1, 3 and 4 were already described in the literature (Stolze et al., 2015; Maus et al., 2016). Moreover, fed substrates differ between those four plants. BGP 1 uses 45% maize silage, 22% sugar beet and 33% poultry manure. In BGP 2 50% maize silage, 10% grass and 40% pig/cattle manure was used. BGP 3 uses only two fed substrates, namely 67% maize silage and 33% pig manure. BGP 4 operates with 60% maize silage, 30% grass and 10% pig manure as substrates. Metagenome sequence datasets of the four biogas plants are available in the Short Read Archive (SRA).

BGP1: http://www.ncbi.nlm.nih.gov/sra/?term=SRA357211 and http://www.ncbi.nlm.nih.gov/sra/?term=SRA357213;
BGP2: http://www.ncbi.nlm.nih.gov/sra/?term=SRA357208 and http://www.ncbi.nlm.nih.gov/sra/?term=SRA357209;
BGP3: http://www.ncbi.nlm.nih.gov/sra/?term=SRA357214 and http://www.ncbi.nlm.nih.gov/sra/?term=SRA357221;
BGP4: http://www.ncbi.nlm.nih.gov/sra/?term=SRA357222 and http://www.ncbi.nlm.nih.gov/sra/?term=SRA357223.

Before sampling at the sampling tap of the plants, approximately 25 liters of the fermenter digestate were discarded. One liter was then filled into a gastight vessel, respectively, excess air was let out and the bottle hermetically sealed with a screw cap. Within 2 h, at the most, the bottles were then transferred to the laboratory in a closed styrofoam box to keep the sample material warm. At the laboratory a total amount of approximately 100 g of the fresh sample were used for DNA extraction in quadruplicates following the protocol described in Schlüter et al. (2008), applying minor changes to enable larger sample volumes. Two times 50 ml of the sample were filled into a screw cap tube and frozen at −20 °C for subsequent extraction of cytosolic proteins.

2.2. Preparation and high-throughput sequencing of metagenome libraries for biogas-producing microbial communities

For preparation and high-throughput sequencing two replicates were used. A total of 100 ng metagenomic DNA was sheared to 270 bp fragments using a focused-ultrasonicator (Covaris Inc.). The sheared DNA fragments were purified and size selected using SPRI beads (Beckman Coulter Inc.). Obtained fragments were blunt-end–repaired, phosphorylated and A-tailed. Subsequently, T-tailed adapters, containing sequences used during cluster formation and Illumina compatible adapters (IDT, Inc.), were ligated to the purified DNA fragments applying the KAPA-Illumina library creation kit (KAPA Biosystems Inc.). The prepared sequencing libraries were quantified by applying the KAPA Biosystem's next-generation sequencing library qPCR kit (KAPA Biosystems Inc.) and run on a Roche LightCycler 480 real-time PCR instrument (Roche Inc.). Sequencing of the quantified sample libraries was performed on the Illumina HiSeq 2000 sequencer using the Illumina TruSeq SBS v3-HS kit, following a 2 × 150 indexed high-output run protocol.

2.3. Metagenome assembly and quality control

For each BGP, we first combined all sequencing reads of the two replicates prior to assembly. To further analyze the effect of sequencing depth – and assembly quality –, we randomly subsampled 12.5%, 25%, and 50%, respectively, of BGP3's sequencing