Proteotyping of laboratory-scale biogas plants reveals multiple steady-states in community composition

F. Kohrs[1], R. Heyer[1], T. Bissinger[2], R. Kottler[2], K. Schallert[1], S. Püttker[1], A. Behne[3], E. Rapp[2,3], D. Benndorf[1,], U. Reichl[1,2]*

[1] Otto von Guericke University, Bioprocess Engineering, Universitätsplatz 2, 39106 Magdeburg, Germany

[2] Max Planck Institute for Dynamics of Complex Technical Systems, Bioprocess Engineering, Sandtorstraße 1, 39106 Magdeburg, Germany

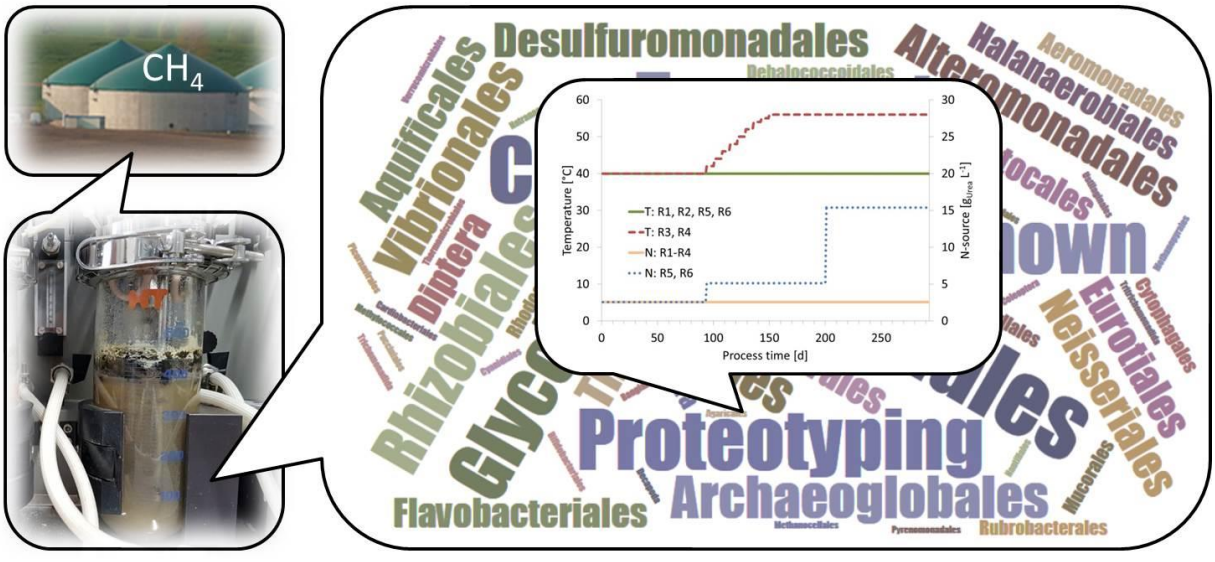
[3] glyXera GmbH, Leipziger Strasse 44 (Zenit Building), 39120 Magdeburg, Germany

* Corresponding author. Tel.: +49 391 67 52160. E-mail address: benndorf@mpi-magdeburg.mpg.de (D. Benndorf)

Highlights

- Proteotyping is a sensitive tool for analysis of microbiomes from biogas plants
- Enrichment of microbial communities resulted in multiple steady-states
- Thermophilic reactors showed increased abundance of cellulose-degrading proteins
- Gap between laboratory and full-scale biogas reactors should be carefully considered

Graphical abstract



Abstract

Complex microbial communities are the functional core of anaerobic digestion processes taking place in biogas plants (BGP). So far, however, a comprehensive characterization of the microbiomes involved in methane formation is technically challenging. As an alternative, enriched communities from laboratory-scale experiments can be investigated that have a reduced number of organisms and are easier to characterize by state of the art mass spectrometric-based (MS) metaproteomic workflows.

Six parallel laboratory digesters were inoculated with sludge from a full-scale BGP to study the development of enriched microbial communities under defined conditions. During the first three month of cultivation, all reactors (R1-R6) were functionally comparable regarding biogas productions (375-625 NL $L_{reactor\ volume}^{-1}$ d^{-1}), methane yields (50-60%), pH values (7.1-7.3), and volatile fatty acids (VFA, <5 mM). Nevertheless, a clear impact of the temperature (R3, R4) and ammonia (R5, R6) shifts were observed for the respective reactors. In both reactors operated under thermophilic regime, acetic and propionic acid (10-20 mM) began to accumulate. While R4 recovered quickly from acidification, the levels of VFA remained to be high in R3 resulting in low pH values of 6.5-6.9. The digesters R5 and R6 operated under the high ammonia regime (>1 gNH_3 L^{-1}) showed an increase to pH 7.5-8.0, accumulation of acetate (>10 mM), and decreasing biogas production (<125 NL $L_{reactor\ volume}^{-1}$ d^{-1}).

Tandem MS (MS/MS)-based proteotyping allowed the identification of taxonomic abundances and biological processes. Although all reactors showed similar performances, proteotyping and terminal restriction fragment length polymorphisms (T-RFLP) fingerprinting revealed significant differences in the composition of individual microbial communities, indicating multiple steady-states. Furthermore, cellulolytic enzymes and cellulosomal proteins of *Clostridium thermocellum* were identified to be specific markers for the thermophilic reactors (R3, R4). Metaproteins found in R3 indicated hydrogenothrophic methanogenesis, whereas metaproteins of acetoclastic methanogenesis were identified in R4. This suggests not only an individual evolution of microbial communities even for the case that BGPs are started at the same initial conditions under well controlled environmental conditions, but also a high compositional variance of microbiomes under extreme conditions.

Keywords

Anaerobic Digestion, Laboratory-Scale Biogas Plant, Community Profiling, Proteotyping, Metaproteomics, MetaProteomeAnalyzer

Abbreviations

6-FAM: 6-carboxyfluorescein, AD: anaerobic digestion; ACN: acetonitrile, RT: room temperature (21°C), bp: base pair, BGP: biogas plant, FA: formic acid, HRT: hydraulic retention time, Ino: inoculum sample, LC: liquid chromatography, M: mesophil(ic) sample, MS: mass spectrometry/spectrometer, MS/MS: tandem MS, N/high-N: process with high load of nitrogen source, R: reactor, SAO: syntrophic acetate oxidation; SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis, SpC: spectral count, T: thermophile(ic) sample, TRF: terminal restriction fragment, VFA: volatile fatty acid(s), xCGE-LIF: multiplexed capillary gel electrophoresis with laser induced fluorescence detection.

Introduction

Conversion of agricultural waste into biogas is a sustainable source of renewable energy. The so called anaerobic digestion (AD) is performed in large parallel or serial digester systems of different sizes and designs, commonly referred to as biogas plants (BGP). Additional classifications are made depending on the process temperature [1], the type (e.g. silage and/or manure and dung) and consistency (e.g. moisture content) of the used substrate [2, 3], and the ammonium or ammonia concentrations [4]. Independent from these conditions, AD process is subdivided into the four steps hydrolysis, fermentation, acetogenesis and methanogenesis [3], which are executed by different groups of microorganisms forming complex microbial communities — the microbiome [5].

Laboratory-scale digester systems (ranging from a few hundred milliliters to several liters working volume) are commonly used as a scale-down model to investigate AD [6-8]. These systems benefit from a better control over the cultivation parameters, and allow well-directed disturbances without risking costly malfunction of full-scale BGP. In a highly controlled process, the substrate is fully defined and continuous stirring enables homogeneous mixing and representative sampling. This is in contrast to full-scale BGP with occasional dead zones or floating layers [3, 9], and varying and non-sterile substrates [10]. However, microbial communities evolving in laboratory bioreactors operating under well-defined process conditions loose part of their complexity [11]. While this facilitates analytics,

the question arises to what extent results can be transferred to the optimization of full-scale BGP.

Many different analytical methods are routinely applied to study microbial communities in BGP. Genomic approaches, like cloning and sequencing of microbial DNA [12] or fingerprinting of 16S rRNA genes (e.g. T-RFLP — terminal restriction fragment length polymorphisms, [13]), allow to explore the diversity of *Archaea* and *Bacteria* of microbial communities. As a complementary approach, metaproteomics turned out to be well suited to capture the physiological state and functions of a microbial population [14]. State of the art methods rely on gel-free approaches [15] pushed by the rapid development of high resolving mass spectrometers (MS) for protein identification, and powerful tools for data analysis [16]. Their application revealed great potential for the characterization of mixed microbial communities, which was referred recently as proteotyping [17]. So far, this term was only used for the identification of single microorganisms by characteristic protein mass spectra derived from Matrix-Assisted Laser Desorption/Ionization-Time-Of-Flight MS analysis (MALDI-TOF-MS) [18]. In a recent review, however, the term proteotyping was extended to cover classification, characterization and identification of microorganisms as well as microbial communities by tandem MS and MS/MS-based shotgun proteomics [19]. The first comprehensive proteotyping of microbial communities in technical biocoenoses aimed at the correlation of biological processes of the microbiome in BGP with respective process parameters [17]. Applications of biostatistics and data mining tools (e.g. principal component analysis or clustering) allowed identifying correlations of taxonomies, functions and metaproteins with process parameters (e.g. temperature, substrate, reactor design or nitrogen content) from extensive lists of identified proteins — without laborious hit-by-hit evaluation.

In this study six parallel digesters were inoculated with sludge from a full-scale BGP to enrich microbial communities under defined conditions. After three month of cultivation, steady-state operation was achieved for all digesters. Subsequently, the temperature and the ammonia concentration were increased for two reactors each. Based on metaproteomics the following questions were addressed: How similar are stable microbial communities operating under exactly the same environmental conditions? Can marker species or functions be determined representing the different process regimes using proteotyping?

Material and methods

Reactor setup

For enrichment, a Sixfors multi bioreactor system (INFORS AG, Bottmingen, Switzerland) with six parallel 500 mL glass vessels was used (R1-R6; 400 mL working volume). Each reactor was equipped with an integrated Pt100 temperature probe, a pH electrode (Type 405-DPAS-SC-K8S, Mettler-Toledo GmbH, Gießen, Germany), and gastight tubing (Santoprene® LEZ-SAN, ID 1.6 mm, thickness 1.6 mm, Medorex, Nörten-Hardenberg, Germany) connected to a Luer/Lock sampling valve (Eppendorf AG, Hamburg, Germany). Agitation was performed by a magnetic propeller stirrer (INFORS AG, Bottmingen, Switzerland); an exhaust gas cooler supplied with 20°C cold water removed any aqueous vapor from the produced biogas. All remaining reactor ports were closed by either plugs or bypassed with gastight tubing.

Sample origin and reactor inoculations

Approximately 5 L of fresh biomass were collected from a local BGP (Magdeburg, Germany, see Supplementary_Note 1 for details about setup, process and substrates) for reactor inoculations and directly transferred to the laboratory. Sludge was centrifuged (3,000 x g) in 50 mL vessels for 3 min at room temperature (RT) to remove crude fibers to obtain more homogenous suspensions in the digesters. For each of the six laboratory-scale reactors 190 mL supernatant were pooled and mixed in the reactor with 210 mL tap water preheated to 40°C. First regular feeding/sampling was performed two days later along with the installation of the "GärOnA" gas analysis system, which will in the following be referred to as the starting point of the cultivation (0 d).

Medium and feeding

The basic composition of the medium is described by Bensmann et al. [20] with the following exception: meat extract, yeast extract, and peptone were replaced by 11.5 g L^{-1} glucose and 34.4 g L^{-1} cellulose as carbon source to exclude any undefined compounds. In addition 2.56, 5.13 or 15.38 g L^{-1} urea were used as a nitrogen source, depending on the process regime (see below). Finally, the vitamin solution was supplemented by 0.05 mg L^{-1} nicotinic acid, 0.05 mg L^{-1} pantothenic acid, 0.02 mg L^{-1} biotin, 0.02 mg L^{-1} folic acid and 0.05 mg L^{-1} 4-aminobenzoic acid. The carbon, nitrogen and phosphor ratio (C:N:P) of the media was 105:7:1 (R1-R4) and 122:41:1 (R5-R6), respectively.

To allow adaption of microorganisms to the defined media, the substrate dosage was increased slowly from 5 mL per feeding during the first ten days to 7.5 mL from day 10-14 and finally 10 mL until the end of the experiment. Sampling was performed directly before feeding to keep the reactor volume constant, and to achieve quasi-continuous process conditions. During the main phase of cultivations (day 14-298), an average of 4.5 feedings per week was performed, which corresponds to a hydraulic retention time (HRT) of 62 d (Table 1).

Process regimes and sampling

Process conditions of the first 93 d (pre-shift state) are described in Table 1. At day 94, changes in the temperature regime (R3, R4) and in the ammonium concentration (R5, R6) were initiated while R1 and R2 were kept at constant mesophilic conditions as a control (Figure 1-3). Temperatures of R3 and R4 were increased stepwise by 2°C per week to 54°C and finally by 1°C per week to 56°C to target a thermophilic regime. For R5 and R6, the urea content in the medium was increased from 2.56 g L^{-1} to 5.13 g L^{-1} at day 94 and furthermore to 15.38 g L^{-1} at day 200, aiming for ammonium levels exceeding 3 g L^{-1}.

Before sampling, tubing connecting exhaust gas coolers and gas analysis system was clamped to avoid air entering reactors. All pH values reported were measured immediately after sampling with an external probe (Basic Meter PB-11 and PY-P20 electrode, Sartorius AG, Göttingen, Germany), and stored at -20°C. This was necessary as the online electrodes showed a significant drift in pH value over process time.

Gas analysis

For qualitative and quantitative biogas measurements, the anaerobic lab fermentation system "GärOnA" with integrated gas analysis (Gesellschaft zur Förderung von Medizin- Bio- und Umwelttechnologie e.V., Halle, Germany) was used. Therefore, the exhaust gas cooler of each reactor was connected to a pressure transducer head. The produced gas volume was calculated from the headspace of the reactor and the resulting pressure change using the ideal gas law. Daily gas production is reported in NL $L_{reactor\ volume}^{-1}$ d^{-1}, referring to gas volumes at standard conditions (0°C and 1,013.25 mbar) per L reactor volume. At an overpressure of 1,050 mbar, the pressure transducer head was vented and the released gas directly transferred to an online coupled gas chromatographic system (GC, ECH

Elektrochemie Halle GmbH, Halle Germany). The GC was equipped with a thermal conductivity detector, a stainless steel GC column (operated at 180°C, ShinCarbon ST, Restek Corporation, Bellefonte PA, USA) and a 200 µL sample loop (operated at 40°C). Argon 5.0 (Westfalen AG, Münster, Germany) was used as mobile phase for determination of CH_4 at a pressure of 2.5 bar.

Metaproteomics

An established metaproteomic workflow was applied as published earlier [17, 21-27] (see Supplementary_Note 3). Following the proteotyping approach by Heyer et al. [17], non-redundant metaproteins were assembled by grouping protein identifications according to their UniRef50 clusters [28]. Respective common ancestor taxonomies where first inherited from peptides to proteins and subsequently from proteins to metaproteins. Metaproteins, UniProtKB keywords for biological processes and taxonomic orders along with spectral counts (SpC) were exported as comma separated values by the MetaProteomeAnalyzer for further processing using the statistic toolbox of MATLAB (The MathWorks GmbH, Ismaning, Germany, version 8.3.0.532 R2014a). MATLAB scripts used are described in the supplement of the proteotyping publication of Heyer et al. [17]. In the following text, to avoid confusions, the UniProKB keywords are set in 'inverted commas'.

Community fingerprinting (T-RFLP)

T-RFLP analyses were carried out following established protocols [29-33]. A detailed description including some minor modifications made is described in Supplementary_Note 3.

xCGE-LIF-based analysis of terminal restriction fragments

Normalized electropherograms were obtained by analyzing the terminal restriction fragments (TRF) from previous amplicon digests via multiplexed capillary gel electrophoresis with laser induced fluorescence detection (xCGE-LIF) utilizing a 4-capillary DNA sequencer (ABI PRISM 3100-Avant Genetic Analyzer, Applied Biosystems, Foster City CA, USA) [34]. Each of the four parallel capillaries had an effective capillary length of 50 cm (Applied Biosystems). Undiluted POP-6™ polymer (Applied Biosystems) was used as separation matrix. Samples were diluted 1:10 in Hi-Di™ Formamide (Applied Biosystems) and injected for 20 s at 1.5 kV. Separation was performed for 6,500 s at 12.2 kV and 50°C. An internal standard (*Gene*

200 *Scan*™ 500 ROX™, Applied Biosystems) ranging from 35-500 base pairs (bp) was added to
201 the samples after diluting with Hi-Di™ Formamide [35].

202 *Processing of xCGE-LIF data*

203 The generated electropherogram data files were converted from their proprietary FSA file
204 format to XML format using the Data File Converter software (Applied Biosystems). Further
205 automated data processing and analysis, including migration time normalization based on
206 the internal bp standard added to each sample, was performed using the glyXtool software
207 (version 5.2.5., glyXera GmbH, Magdeburg, Germany). The following processing parameters
208 were chosen: (1) settings file: Standard_POP6.set; (2) peak number threshold: 100; (3) peak
209 picking intensity threshold (equals the limit of quantification): 10 times signal-to-noise ratio;
210 (4) left peak picking limit: -10 migration time units; (5) right peak picking limit: 700 migration
211 time units and (6) baseline correction using moving averages. Variations in the DNA amount
212 loaded for each PCR were considered by normalizing peak heights to the total peak height of
213 an electropherogram [36]. Peaks common to sets of electropherograms (corresponding to
214 the six measured sample replicates) where identified via migration time matching with a
215 maximum normalized migration time tolerance of ±0.5 units. Peaks within the normalized
216 electropherograms were referred to as TRF. Only TRF detected in more than three of the six
217 replicates [34] were considered for further processing. Based on these, consensus TRF were
218 constructed with the mean migration times and the median of their intensities. Since
219 building of consensus TRF influenced the total TRF sums, the normalizations of TRF to the
220 total TRF intensities of a sample were repeated before comparing samples on basis of
221 respective consensus profiles. For community profiles, only TRF with a normalized intensity
222 of >3% were considered [34].

223 *Determination of ammonium and volatile fatty acids*

224 Two mL of fresh bioreactor sample were centrifuged for 5 min at 16,400 x g, and
225 supernatants filtered (0.2 µm Whatman SPARTANTM 13/0.2 RC Filter Units, GE Healthcare
226 Life Sciences, München, Deutschland) using disposable syringes. Filtrates were diluted (1:10,
227 1:20, 1:40, and 1:50) with ultra-pure water (Millipore Type 1 ultrapure water, Merck KGaA,
228 Darmstadt, Germany) before measuring ammonium (NH_4^+) in a BioProfile® (Nova
229 Biomedical, Waltham MA, USA). Respective concentrations of ammonia (NH_3) were
230 calculated from ammonium measurements taking into account pH value and temperature
231 [37]. Determination of ammonium nitrogen (NH_4^+) was carried out to verify that the applied

urea load in the substrate of R5 and R6 (day 93, Table 1) corresponded to a high-N regime (NH_4^+ >3 g L^{-1}) [38]. Since this was not the case at day 183, a second medium shift was applied that resulted in NH_4^+ concentrations of 5.05 g L^{-1} (0.60 gNH_3 L^{-1}) and 5.07 g L^{-1} at day 240 (0.66 gNH_3 L^{-1}, 1HRT: hydraulic retention time.

Table 2), respectively.

Formate, acetate, lactate, propionate and butyrate were quantified using an anion-exchange HPLC method (Dionex ICS-5000 Reagent-Free HPIC System, Thermos Fisher Scientific Inc.). Culture samples, stored at -20°C, were centrifuged (5 min at 16,400 x g and RT), the supernatant was sterile filtered (0.2 µm), diluted 1:100 with ultra-pure water, and transferred to a 2 mL glass vial (WIC 41150 and WIC 43945/B, WICOM Germany GmbH, Heppenheim, Germany). Five µL were injected (Dionex AS-AP Autosampler) to the HPLC system with an eluent generator (Dionex ICS-5000+ EG Eluent Generator with KOH cartridge) and electrolytic suppression (Dionex AERS 500 2 mm). VFA were separated using two anion-exchange columns (Dionex IonPac AS11 Analytical Column 2 x 250 mm, all Thermo Fisher Scientific Inc.) using 0.35 ml min^{-1} eluent flow of a potassium hydroxide gradient: 0-2 min, linear 0.2-0.5 mM; 2-8 min, isocratic 0.5 mM; 8-10 min, linear 0.5-3.0 mM; regeneration: 3 min 50 mM, 5 min 0.2 mM). VFA were monitored with a conductivity detector and quantified via an external standard containing the respective acid mix in different concentrations (1-310 µM). Method validation confirmed a limit of quantification for the undiluted samples under 0.5 mM (formate: 0.40 mM, acetate: 0.23 mM, lactate: 0.25 mM, propionate: 0.28 mM, butyrate: 0.18 mM).

Results

Abiotic process data

After a short delay during the first two weeks of enrichment (start-up phase) the daily biogas production increased to 0.481 ±0.09 NL $L_{reactor\ volume}^{-1}$ d^{-1} for R1, 0.500 ±0.09 NL $L_{reactor\ volume}^{-1}$ d^{-1} for R2, 0.538 ±0.105 NL $L_{reactor\ volume}^{-1}$ d^{-1} for R3, 0.548 ±0.096 NL $L_{reactor\ volume}^{-1}$ d^{-1} for R4, 0.502 ±0.102 NL $L_{reactor\ volume}^{-1}$ d^{-1} for R5, and 0.500 ±0.098 NL $L_{reactor\ volume}^{-1}$ d^{-1} for R6. The high-N regimes R5 and R6 showed decreasing gas productions to less than 125 NL $L_{reactor\ volume}^{-1}$ d^{-1} after day 200 (Figure 1-3). The methane content of the produced biogas was also stable for all regimes with small fluctuations (R1: 53.41 ±6.3% CH_4, R2: 53.6 ±6.2% CH_4, R3: 53.2 ±6.3% CH_4, R4: 53.5 ±5.9% CH_4, R5: 54.5 ±7.9% CH_4, R6:

54.3 ±7.4% CH_4). These variations can be explained by the addition of feed as CO_2 release starts earlier than methane production, accounting for an initial higher percentage in the biogas. Surprisingly, no reduction in methane content of R5 and R6 was observed with the decrease in net gas production starting at day 200.

For all reactors, the initial pH of 7.5 (0 d) dropped slowly to a range of 7.1-7.3 within the first 93 d, which is close to the pH value of the medium (Figure 1-3). For the high-N regime (R5 and R6), the pH values increased again to pH 7.5; and due to the second shift in the urea content of the medium (day 200) to about pH 8.0 (Figure 3). The pH of the mesophilic reactors (R1 and R2, sample prefix M) and the thermophilic reactors (R3 and R4, sample prefix T) remained in the range of 7.1-7.3 until 175 d of cultivation. Later on, the pH of R3 and R4 dropped to a range of 6.5-6.9 (Figure 1-2). Temporarily, the pH of R3 was below 6.2 (day 261) which was probably due to an accumulation of VFA.

For all reactors, lactate, formate and butyrate concentrations were stable below 5 mmol (Supplementary_Figure 2-7). For R1 and R2, the concentrations of all VFA were stable (<5 mmol) over the whole course of time (Figure 1). This was also true for the other regimes during the first phase of cultivation (0-93 d). However, after the temperature shift (R3, R4), acetate concentration (10-20 mM) and propionate concentration (10 mM) increased (Figure 2). While high levels of both VFA were found over the rest of the cultivation time in R3, the concentrations dropped again to the base level below 5 mM (day 200) in R4. The high-N regimes (R5, R6) resulted in a slow increase in acetate concentration to 5 mM as a response to the first shift (93-200 d, Figure 3). Elevated acetate concentrations (>10 mM) were only found during the very late phase of R5 and R6 (240-298 d).

Community fingerprinting

SDS-PAGE of protein extracts was performed for the inoculum (sample prefix Ino) as a reference for monitoring of the impact of process conditions on the microbiomes,) and the process samples of day 93 and 261 (Figure 4). Surprisingly, the protein patterns derived from all six biological replicates already showed differences at day 93. Besides minor changes, R1 revealed a unique band of very high intensity at 18 kDa while R2 and R3 had a unique band below 15 kDa. Furthermore, for R2 and R5, some high molecular weight protein structures (in >180 kDa) were missing that are visible in the other lanes. Furthermore, when comparing R1-R6 at day 261 after the applied shifts, the protein profiles showed significant differences

(Figure 4, e.g. R1 and R2: the unique band below 15 kDa is still present; R4: fewer high molecular weight structures >100 kDa compared to R3; R5 and R6: differing patterns between 15-25 kDa).

Additional T-RFLP analyses targeting the 16S rRNA genes of *Archaea* (Figure 5) and *Bacteria* (Figure 6) confirmed these findings, and showed differences in TRF profiles and abundances for both, the *Archaea* and the *Bacteria*, before (day 93) and after the temperature and nitrogen shifts (day 261). Of the eight dominant archaeal TRF were identified in the inoculum after 93 d of cultivation, only 2-4 of these TRF were still present in R1-R6. All reactors revealed new individual TRF that were not found in the inoculum before. After 261 d of cultivation, TRF composition changed again in all reactors compared to their respective sample at day 93. TRF 106 was one of the most abundant archaeal fragments in R1-R6; additionally, samples taken at different time points of cultivation at the same process regime resembled each other more. Only four bacterial TRF were identified in the inoculum (Figure 6), whereas R1-R6 after 93 d of cultivation showed more TRF with a higher heterogeneity. Similar to the archaeal fingerprints, the TRF of bacteria also varied between all reactors. For instance, for the mesophilic R2, the compositional complexity was further reduced to only three TRF after 261 d of cultivation. The thermophilic R3 was dominated by TRF 67 that accounted for almost 80% of all TRF. Finally, for both of the high-N regimes (R5, R6), TRF 216 was identified as the most abundant.

MS/MS analyses and proteotyping

First, the reproducibility of MS/MS measurements was confirmed by hierarchical clustering of UniRef50 metaproteins of all samples using normalized SpC. As shown in Figure 7, the technical replicates clustered closely together (suffixes 1 and 2). Further inspection revealed two main cluster branches separating the sample of R3 (day 261) from all other regimes. Within the latter branch, both replicate protein extractions as well as the replicate measurements of the inoculum formed a subcluster of very high similarity. Furthermore, while all samples of day 93 were clearly separated, they still formed an own branch which was different to all other samples. Finally, all samples taken at the endpoints (day 261) were separated across the cluster tree. While samples from the mesophilic reactors (R1, R2) were clustered in a common branch, samples from the thermophilic reactors (R3, R4) were clearly

separated from all other samples and from themselves. Interestingly, the samples from the high-N regimes (R5, R6) resembled the inoculum.

To identify reasons for the large variations in the microbial communities of the six enrichments (day 93), the identified taxonomic orders were depicted in box plots (Figure 8). Overall, the inoculum samples were dominated by *Bacillales* (17%), *Methanosarcinales* (12.5%), *Enterobacterales* (11%), *Rhizobiales* (8.5%) and *Methanococcales* (4.5%). The dispersion – the range from the lowest to the highest percentage value of a taxonomic order— did not exceed 2% of the total community in any case. However, for day 93, the box plots differed clearly regarding the dominating orders and the scattering of the single values. *Bacillales* (12%), *Enterobacterales* (10.5%), *Methanosarcinales* (10%), *Rhizobiales* (8%) and *Lactobacillales* (6.5%) were most abundant at day 93. Also, the abundances of orders show a high degree of dispersion (up to 6.5% for the most abundant order *Methanosarcinales*). This trend became even more distinct for later time points of cultivations, especially for the thermophilic regime in R3 and R4 (Supplementary_Figure 8).

Despite all the differences in community composition, the volume of the produced biogas and the methane concentrations of all reactors were similar over most of the process time (Figure 1-3) which implies functional similarity. Therefore, different process regimes were analysed on a functional level using UniProtKB keywords [39] for biological processes linked to metaproteins (Figure 9). Four different groups can be distinguished. (i), the inoculum samples were closely related to each other in the lower right region. (ii), all sample taken at day 93 and most samples of day 261 formed a cluster near the centre of the plot, showing high functional similarity. (iii), final stages of R3 are clearly separated from all other samples in the bottom left region; (iv), R1 at day 261 was separated from the centre in the upper region.

Biological processes with a larger impact on the separation of clusters can be derived from the respective loading plot (Figure 9). For example, inocula were separated by higher abundance of the biological processes of ‘Transport’ and ‘Methanogenesis’. The mesophilic R1 (day 261) revealed a high abundance of metaproteins related to ‘Glycolysis’ and ‘Amino-acid biosynthesis’, whereas, the final stage of the thermophilic R3 was dominated by metaproteins related to ‘Carbohydrate metabolism’, ‘Polysaccharide degradation’, and ‘Cellulose degradation’.

Reactor-specific metaproteins

In order to specify the previous findings on the taxonomic and functional levels, and to eventually pinpoint reactor-specific markers, quantitative changes in the relative abundance of metaproteins were analysed for each reactor between day 93 and day 261. More than 100 metaproteins of each reactor showed changes in abundances (Supplementary_Note 2); some of these metaproteins are highlighted in the following (numbers in brackets relate to the sum of SpC of either day 93 or day 261; listed taxonomies refer to the lowest common ancestor).

For the thermophilic reactor R3, the SpC of the archaeal methyl-coenzyme M reductase subunit alpha (UniRef50_P07962) of *Methanosarcina barkeri* (SpC: 28 to 4) and acetyl-CoA decarbonylase/synthase complex subunit alpha 2 (UniRef50_Q46G04) of *Methanosarcina* (SpC: 11 to 0) were clearly reduced after the temperature increase. The cell surface glycoprotein 1 (UniRef50_Q01866, SpC: 0 to 83), cellulose 1,4-beta-cellobiosidase (UniRef50_P0C2S1, SpC: 0 to 199), cellulosomal-scaffolding protein A (UniRef50_Q06851, SpC: 0 to 60), and cellulosomal-scaffolding protein B (UniRef50_Q06851, SpC: 0 to 85) of *Clostridium thermocellum* were identified for the first time with high abundance under thermophilic conditions. Furthermore, 60 kDa chaperonin (UniRef50_Q9C667) of *Bacteria* (SpC: 19 to 81), in particular the 60 kDa chaperonin 1 (UniRef50_Q08854) of *Streptomyces avermitilis* (SpC: 18 to 53), were highly abundant. In addition, the number of several proteins assigned to *Thermotoga maritima* MSB8 increased. In the thermophilic reactor R4, F420-dependent methylenetetrahydromethanopterin dehydrogenase (UniRef50_P55300) of *Methanomicrobiales* (SpC: 18 to 7) and methyl-coenzyme M reductase subunit gamma (UniRef50_P12973) of *Methanothermus fervidus* (SpC: 10 to 1) decreased. In contrast to R3, acetyl-CoA decarbonylase/synthase complex subunits gamma 2 and epsilon 1 (UniRef50_Q9C4Z0, UniRef50_Q46G05) of *Methanosarcina* were highly abundant (SpC: 2 to 32 and 2 to 18). The previously highlighted 60 kDa chaperonins (UniRef50_Q08854) of *Streptomyces avermitilis* (SpC: 19 to 53), bacterial 60 kDa chaperonin (UniRef50_P35635, SpC: 3 to 22), and cellulose-scaffolding protein B (UniRef50_Q01866, SpC: 0 to 24) of *Clostridium thermocellum* as well as the metaproteins from *Thermotoga maritima* MSB8 also increased in R4.

Under high-N concentrations in reactors R5 and R6, only a few metaproteins were identified with high abundance. This included 5,10-methylenetetrahydromethanopterin reductase (UniRef50_Q50744) of *Methanothermobacter thermautotrophicus* (SpC: 2 to 6), acetyl-CoA decarbonylase/synthase complex subunit alpha 2 (UniRef50_Q46G04) of *Methanosarcina* (Spc: 1 to 12) as well as polyribonucleotide nucleotidyltransferase (UniRef50_Q8FPB7) of *Propionibacterium acnes* (SpC: 0 to 6) and 60 kDa chaperonin 1 (UniRef50_Q08854) of *Streptomyces avermitilis* (SpC: 12 to 42) in R5. In R6, acetyl-CoA decarbonylase/synthase complex subunit alpha 2 (UniRef50_Q46G04) of *Methanosarcina* (SpC: 2 to 16) and 60 kDa chaperonin 1 (UniRef50_Q08854) of *Streptomyces avermitilis* (SpC: 21 to 46) were increased.

Discussion

Reactor performance

Methane contents for all cultivation regimes monitored were stable with means of 53.2-54.5% and maximum fluctuations below ±8% (Figure 1-3). Except for the later cultivation phases of R5 and R6, the corresponding biogas productions were also at steady-state (Figure 3). To evaluate the production of biogas, the theoretical methane content was estimated using the empirical formula of Boyle et al. [40] for the media supplied (Table 1) as 48.7% (R1-R4; R5 and R6 until day 93) and 43.3% (R5 and R6 at day 200), respectively. The minor difference to the measured methane contents can be explained by a high solubility of CO2 in the aqueous reactor sludge, resulting in relatively higher methane concentrations in the headspace of the reactors. Subsequently, solved CO2 is removed from the reactors during the regular sampling. This also influenced the volumetric production of biogas. Overall, the measured biogas volumes were in the range of 60%-100% of the calculated biogas volumes (Supplementary_Figure 1) between 50 d and 210 d of cultivation. The initial higher yields can be explained by a take-over of substrates from the inoculum, which were gradually converted. The final decrease in biogas yield below 20% in high N-regimes R5 and R6 after 250 d resulted from the second substrate shift, where ammonium accumulated to >5 g L^{-1}. High ammonium concentrations subsequently resulted in an increase in pH value from 7.5-8 (Figure 3), the chemical equilibrium was shifted towards toxic ammonia concentrations resulting in elevated levels of 0.60-0.66 gNH_3 L^{-1} (1HRT: hydraulic retention time.

Table 2). This is in agreement with earlier studies, were a toxification of the methanogenic community and a negative impact on the biogas production was reported for ammonia concentrations of 0.7 gNH_3 L^{-1} for adapted [41] and 0.1-0.15 gNH_3 L^{-1} for unadapted communities [42], respectively. The resulting toxification effected an inhibition of the acetoclastic methanogenic community, explaining the ongoing accumulation of acetate in both reactors after day 200. In contrast, the observed time courses of the pH values and VFA

of the mesophilic regime (R1, R2) were not affected (Figure 1). Finally, for the thermophilic regimes (R3, R4), a comparison of the pH values and the concentrations of acetate and propionate revealed clear differences (Figure 2). The temperature shift caused an initial accumulation of acetate and propionate in R4 from less than 5 mM to more than 10 mM that dropped to less than 5 mM shortly after reaching 56°C. A similar behavior was described before by Ziembińska-Buczyńska et al. [43]. In reactor R3, however, concentrations of both VFA remained high, resulting in lower pH values than in its biological replicate R4. For unknown reasons, the microbial community of R4 was able to adapt to the temperature shift and to reduce the accumulated VFA gradually while levels of both remained higher (>10 mM) for R3. Both acetate [44] and propionate [45] were reported previously as markers for an instable AD process, but critical concentrations were not reached here. Apparently, the process data already hint at differences regarding the composition of the microbial communities in R3 and R4. Nonetheless, from an operator point of view, the mesophilic and the thermophilic regimes as well as the first 200 days of the high-N regimes, showed a stable performance and biogas production (Figure 1-3).

Validation of proteotyping

Reproducible quantification is still a challenge in proteomics of pure cultures. Labelling techniques with tags containing stable isotopes are widely applied but not applicable to samples containing contaminations from the inoculum [46]. Thus, spectral counting was applied in this study for quantification. However, replicated MS/MS runs of the same sample using data-dependent data acquisition (top number of precursor selections) [47] are known to show a low overlap of identified peptides interfering with subsequent statistical analysis. Nevertheless, the fact that nearly all technical replicates formed distinct branches in the cluster tree (Figure 1, suffixes 1 and 2) whereas the different process regimes were clearly separated validated the applied experimental workflow.

Proteotyping indicates multiple steady-states of microbial communities

The six biogas reactors running in parallel under well-defined conditions, which were inoculated with the same starting material, showed a similar behavior for the first 93 d regarding the production of biogas, pH value, and the content of VFA (Figure 1-3). However, SDS-PAGE of proteins (Figure 4), LC-MS/MS of tryptic peptides (Figure 8), and T-RFLP of 16S rRNA genes suggested clear differences in community composition (Figure 5-6). As a time

period of 93 d is rather short for a full adaption of microbial communities from the inoculum to laboratory conditions, the latter results might represent intermediate states of enrichment only. However, the different SDS-PAGE and T-RFLP patterns of R1 and R2 (controls) after long adaption (until day 261, >3 HRT) suggest that indeed different stable communities with similar performance evolved. This is also indicated by principal component analysis based on metaprotein identifications: samples of R1 at day 93 and day 261 showed a high distance to each other in the plot, indicating an ongoing adaption (Figure 9). However, samples of R2 at day 93 and day 261 were plotted close to each other, indicating a stable community composition. In comparison to the inoculum, the various replicates were characterized by a larger spread of the abundance of the identified taxa (Figure 8). Apparently, the spectral abundances of the orders *Methanosarcinales*, *Methanomicrobiales* or *Bacillales* can vary without having a significant impact on reactor performance — for example production of biogas.

Microbial communities of full-scale agricultural BGP showed compositional variations, although biogas production was not affected [48, 49]. The fact that replicates inoculated with identical biomass and cultivated under well-defined laboratory conditions using a defined medium can result in microbial communities with different composition is challenging regarding the interpretation of laboratory experiments for process optimization and the analysis of data from BGP. So far, differences or contradictory findings in similar studies [46, 50, 51] were considered to be mainly due to methodical biases during sample lysis or the extraction of DNA and proteins [52], but did not take into account multiple biological steady-states. However, and most obvious for the thermophilic regimes (R3 and R4) in the present study, samples taken from full-scale BGP operated under similar process conditions (e.g. temperature, design, substrate, and biogas production) might yield different microbial communities representing multiple steady-states. Accordingly, the presence of different microbial communities should be considered for interpretation of data and problems regarding the definition of core microbiomes for AD [5, 17, 49, 53].

Regime and reactor-specific metaproteomic analyses

Principal component analysis was performed to identify regime-specific biological processes (Figure 9, upper figure). ‘Transport’ and ‘Methanogenesis’ were found in the respective loading plot (Figure 9, lower figure) to be the characteristic biological processes of the

inoculum. This is in accordance with a proteotyping study performed by Heyer et al. [17], where these functions also dominated full-scale BGP. The replicates of the controls (R1, R2) and thermophilic regimes (R3, R4) after the shift were not closely located. The control R1 was separated from the main group of samples including its replicate R2. It was mainly characterized by a high abundance of the biological processes 'Glycolysis' and 'Amino-acid biosynthesis'. These functions should be common for all process regimes, but since positions in the loading plot (Figure 9, lower figure) result from an overlay of all biological processes this could be explained by a lower abundance of 'Transport' and 'Methanogenesis'. Moreover, 'Polysaccharide degradation' and 'Cellulose degradation' were the most abundant functions that separated the thermophilic regime R3 from its replicate R4 and all other regimes confirming its distinct branch in the cluster tree. 'Cellulose degradation' was expected for all BGP with cellulose as main carbon source, but this biological process was not identified in a previous metaproteomic study [54]. This is probably due to the reduced complexity of the enriched thermophilic communities investigated here: microorganisms preferring mesophilic conditions are outcompeted by those that are able to adapt to higher temperatures. Due to the reduced sample complexity (a lower number of microorganisms), the resolution of MS measurements is enhanced and the detection limit for biological processes lowered [46]. Accordingly, it is more likely to identify low abundant proteins, i.e. 'Cellulose degradation'. Nevertheless, the major biological functions were not significantly changed suggesting the presence of structurally different but functional similar microbial communities. This is in accordance with the very similar reactor performance of the replicated enrichments. Furthermore, the different behaviour of the replicates on the functional level after the shift demonstrates that two biological replicates are not sufficient to identify key functions for the different process regimes. Most likely, small differences in the composition of microbial communities direct their development into different states after changes in process conditions as shown for the thermophilic laboratory-scale regimes (R3, R4) in the following.

Considering the level of metaproteins instead groups of biological functions, methanogenic metaproteins of archaeal *Methanosarcina* were decreased in R3 as a result of the shift to thermophilic conditions. This corresponds to findings of earlier studies showing decreased acetoclastic methanogenesis [55, 56]. According to the results of biological processes (Figure 9), cellulolytic enzymes like cellulose 1,4-beta-cellobiosidase and cellulosomal-scaffolding

proteins of *Clostridium thermocellum* were highly abundant in the final stage. To some extent these enzymes were also detected in R4, which was confirmed by principal component analysis, too. In contrast to R3, the methanogenic acetyl-CoA decarbonylase/synthase complex subunits of acetoclastic *Methanosarcina* were highly increased in the thermophilic R4 at day 261. This tendency towards acetoclastic methanogenesis at higher temperatures was also observed earlier [17]. *Methanosarcina* present in R4 would be also a plausible explanation for the transient acetate accumulation after the shift in R4 whereas acetate remained at higher levels in R3 for a longer time period (above 10 mM, Figure 2). Nonetheless, in the context of this study, the different behaviour of the thermophilic reactors R3 and R4, e.g. pH values and VFA, demonstrates very clearly that diverse complex microbial communities can evolve under comparable conditions.

For the high-N regimes R5 and R6, the sets of identified metaproteins showed similar trends. However, the increased abundance of acetyl-CoA decarbonylase/synthase subunits of acetoclastic *Methanosarcina* cannot be easily explained by ammonia toxification. Apparently, the initial intention to enrich a syntrophic acetate oxidation (SAO) community in the high-N regime failed. Although a slight increase in metaproteins indicating hydrogenothrophic methanogenesis was detected, the key enzyme for SAO — formyltetrahydrofolate synthetase [50] — was not detected. Additionally, no other metaproteins indicating the presence of species capable of performing SAO, e.g. *Pseudothermotoga lettingae*, *Clostridium ultunense*, or *Tepidanaerobacter acetatoxydans*, were identified. It might be argued that the applied metagenomes for database search did not cover genes of SAO species present in these enrichments. However, the same metagenomes allowed the successful identifications of candidate proteins and species for SAO in a previous proteotyping study [17]. Therefore, it is more reasonable that the acetate is consumed by ammonia tolerant acetoclastic methanogenesis instead SAO [57]. In the late phase (after the shift), the accumulation of acetate and the decrease in methane production indicated that acetoclastic methanogenesis was more and more inhibited by ammonia concentrations exceeding 1 $gNH_3\ L^{-1}$ (1HRT: hydraulic retention time.

Table 2).

Overall, proteotyping on metaprotein levels seems to be more powerful than using biological functions only. It enabled a clear differentiation of the two stable microbial communities of the thermophilic regimes, and the establishment of hypotheses regarding their discrepancy on structural and functional level.

Conclusions

548 Proteotyping is a sensitive tool for characterization of microbial communities in laboratory-
549 scale reactors and full-scale BGP. In addition, proteotyping provides valuable taxonomic and
550 functional data. However, even for parallel cultivations in well controlled cultivations using
551 the same inoculum, the high compositional and functional variances of microbiomes after
552 enrichment does not enable the identification of specific markers for very different process
553 conditions. Interestingly, besides all differences between microbial communities their
554 process performance was very similar. This highlights the flexibility of microbial communities
555 regarding the efficient use of substrates even under defined laboratory conditions. In
556 addition, the results help to understand contradictory findings of different research groups
557 regarding taxonomic abundancies in full-scale BGP. As a high number of multiple steady-
558 state can evolve even under controlled cultivation conditions, conclusions from laboratory
559 experiments regarding the operation of full-scale BGP should be handled with great caution.
560 Using enriched communities as model systems, the number of replicates should be
561 increased in future studies to properly cover the wide range of biological variation, and to
562 allow for more general conclusions regarding the design and optimization of full-scale BGP
563 [58].

564 **Acknowledgment**

565 The authors acknowledge excellent support of C. Siewert, S. Fischer (both Max Planck
566 Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany) and C. Best
567 (Otto von Guericke University, Magdeburg, Germany) in the laboratory. Furthermore the
568 support of S. Theuerl (Leibniz Institute for Agricultural Engineering Potsdam-Bornim,
569 Potsdam, Germany) was very valuable for setting up the T-RFLP methods. Finally, the
570 authors want to thank M. Leifheit and J. Greiser (Gesellschaft zur Förderung von Medizin-
571 Bio- und Umwelttechnologie e.V., Halle, Germany) for the modification of the GärOnA
572 system to allow monitoring of the reactors.

573 **Funding**

574 R. Heyer was supported by a grant of the Federal Ministry of Food, Agriculture and
575 Consumer Protection (BMELV) communicated by the Agency for Renewable Resources
576 (FNR), grant no. 22404115 (Biogas-Messprogramm III). R. Kottler and E. Rapp acknowledge
577 support by the European Union (EC) under the project “HighGlycan” (grant no. 278535).

References

[1] M.J. McInerney, M.P. Bryant. Basic principles of bioconversions in anaerobic digestion and methanogenesis. Biomass Conversion Processes for Energy and Fuels: Springer; 1981, p. 277-96.

[2] B. Demirel. Major pathway of methane formation from energy crops in agricultural biogas digesters. Critical Reviews in Environmental Science and Technology 44 (2014) 199-222.

[3] P. Weiland. Biogas production: current state and perspectives. Applied microbiology and biotechnology 85 (2010) 849-60.

[4] I. Angelidaki, K. Boe, L. Ellegaard. Effect of operating conditions and reactor configuration on efficiency of full-scale biogas plants. Water Science and Technology 52 (2005) 189-94.

[5] A. Shade, J. Handelsman. Beyond the Venn diagram: the hunt for a core microbiome. Environmental microbiology 14 (2012) 4-12.

[6] A. Hanreich, U. Schimpf, M. Zakrzewski, A. Schlüter, D. Benndorf, R. Heyer, et al. Metagenome and metaproteome analyses of microbial communities in mesophilic biogas-producing anaerobic batch fermentations indicate concerted plant carbohydrate degradation. Systematic and applied microbiology 36 (2013) 330-8.

[7] H. Pobeheim, B. Munk, J. Johansson, G.M. Guebitz. Influence of trace elements on methane formation from a synthetic model substrate for maize silage. Bioresource technology 101 (2010) 836-9.

[8] J. Liu, G. Olsson, B. Mattiasson. A volumetric meter for monitoring of low gas flow rate from laboratory-scale biogas reactors. Sensors and Actuators B: Chemical 97 (2004) 369-72.

[9] M.H. Gerardi. The microbiology of anaerobic digesters: John Wiley & Sons; 2003.

[10] C. Herrmann, M. Heiermann, C. Idler. Effects of ensiling, silage additives and storage period on methane formation of biogas crops. Bioresource Technology 102 (2011) 5153-61.

[11] S. Weiß, M. Tauber, W. Somitsch, R. Meincke, H. Müller, G. Berg, et al. Enhancement of biogas production by addition of hemicellulolytic bacteria immobilised on activated zeolite. Water research 44 (2010) 1970-80.

[12] A. Schlüter, T. Bekel, N.N. Diaz, M. Dondrup, R. Eichenlaub, K.-H. Gartemann, et al. The metagenome of a biogas-producing microbial community of a production-scale biogas plant fermenter analysed by the 454-pyrosequencing technology. Journal of biotechnology 136 (2008) 77-90.

[13] B.G. Clement, L.E. Kehl, K.L. DeBord, C.L. Kitts. Terminal restriction fragment patterns (TRFPs), a rapid, PCR-based method for the comparison of complex bacterial communities. Journal of microbiological methods 31 (1998) 135-42.

[14] P. Wilmes, P.L. Bond. Metaproteomics: studying functional gene expression in microbial ecosystems. Trends in microbiology 14 (2006) 92-7.

[15] R.L. Graham, C. Graham, G. McMullan. Microbial proteomics: a mass spectrometry primer for biologists. Microbial cell factories 6 (2007) 26.

[16] T. Muth, B.Y. Renard, L. Martens. Metaproteomic data analysis at a glance: advances in computational microbial community proteomics. Expert review of proteomics 13 (2016) 757-69.

[17] R. Heyer, D. Benndorf, F. Kohrs, J. Vrieze, N. Boon, M. Hoffmann, et al. Proteotyping of biogas plant microbiomes separates biogas plants according to process temperature and reactor type. Biotechnology for biofuels 9 (2016) 155.

[18] P. Seng, M. Drancourt, F. Gouriet, B. La Scola, P.-E. Fournier, J.M. Rolain, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clinical infectious diseases 49 (2009) 543-51.

[19] R. Karlsson, L. Gonzales-Siles, F. Boulund, L. Svensson-Stadler, S. Skovbjerg, A. Karlsson, et al. Proteotyping: Proteomic characterization, classification and identification of microorganisms–A prospectus. Systematic and applied microbiology 38 (2015) 246-57.

[20] A. Bensmann, R. Hanke-Rauschenbach, R. Heyer, F. Kohrs, D. Benndorf, R. Kausmann, et al. Diagnostic concept for dynamically operated biogas production plants. Renewable energy 96 (2016) 479-89.

[21] N. Popov, M. Schmitt, S. Schulzeck, H. Matthies. Eine störungsfreie Mikromethode zur bestimmung des Proteingehaltes in Gewebehomogenaten. Acta biol med ger 34 (1975) 1441-6.

[22] U.K. Laemmli. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227 (1970) 680-5.

[23] F. Kohrs, R. Heyer, A. Magnussen, D. Benndorf, T. Muth, A. Behne, et al. Sample prefractionation with liquid isoelectric focusing enables in depth microbial metaproteome analysis of mesophilic and thermophilic biogas plants. Anaerobe 29 (2014) 59-67.

[24] V. Neuhoff, N. Arold, D. Taube, W. Ehrhardt. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. Electrophoresis 9 (1988) 255-62.

[25] J.S. Cottrell, U. London. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20 (1999) 3551-67.

[26] T. Muth, A. Behne, R. Heyer, F. Kohrs, D. Benndorf, M. Hoffmann, et al. The MetaProteomeAnalyzer: a powerful open-source software suite for metaproteomics data analysis and interpretation. Journal of proteome research 14 (2015) 1557-65.

[27] A. Shevchenko, S. Sunyaev, A. Loboda, A. Shevchenko, P. Bork, W. Ens, et al. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. Analytical chemistry 73 (2001) 1917-26.

[28] B.E. Suzek, H. Huang, P. McGarvey, R. Mazumder, C.H. Wu. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 23 (2007) 1282-8.

[29] R. Großkopf, P.H. Janssen, W. Liesack. Diversity and structure of the methanogenic community in anoxic rice paddy soil microcosms as examined by cultivation and direct 16S rRNA gene sequence retrieval. Applied and environmental microbiology 64 (1998) 960-9.

660 [30] T. Lueders, M. Friedrich. Archaeal population dynamics during sequential reduction
661 processes in rice field soil. Applied and Environmental Microbiology 66 (2000) 2732-42.

662 [31] D. Lane. 16S/23S rRNA sequencing. Nucleic acid techniques in bacterial systematics
663 (1991) 125-75.

664 [32] W.G. Weisburg, S.M. Barns, D.A. Pelletier, D.J. Lane. 16S ribosomal DNA amplification
665 for phylogenetic study. Journal of bacteriology 173 (1991) 697-703.

666 [33] V. Després, J. Nowoisky, M. Klose, R. Conrad, M. Andreae, U. Pöschl. Characterization
667 of primary biogenic aerosol particles in urban, rural, and high-alpine air by DNA sequence
668 and restriction fragment analysis of ribosomal RNA genes. Biogeosciences 4 (2007) 1127-41.

669 [34] A. Rademacher, C. Nolte, M. Schönberg, M. Klocke. Temperature increases from 55 to
670 75 C in a two-phase biogas reactor result in fundamental alterations within the bacterial and
671 archaeal community structure. Applied microbiology and biotechnology 96 (2012) 565-76.

672 [35] R. Hennig, E. Rapp, R. Kottler, S. Cajic, M. Borowiak, U. Reichl. N-glycosylation
673 fingerprinting of viral glycoproteins by xCGE-LIF. Carbohydrate-based vaccines: methods and
674 protocols (2015) 123-43.

675 [36] J. Dunbar, L.O. Ticknor, C.R. Kuske. Phylogenetic specificity and reproducibility and
676 new method for analysis of terminal restriction fragment profiles of 16S rRNA genes from
677 bacterial communities. Applied and environmental microbiology 67 (2001) 190-7.

678 [37] K.H. Hansen, I. Angelidaki, B.K. Ahring. Anaerobic digestion of swine manure:
679 inhibition by ammonia. Water research 32 (1998) 5-12.

680 [38] M. Westerholm, B. Müller, S. Isaksson, A. Schnürer. Trace element and temperature
681 effects on microbial communities and links to biogas digester performance at high ammonia
682 levels. Biotechnology for biofuels 8 (2015) 1.

683 [39] U. Consortium. Reorganizing the protein space at the Universal Protein Resource
684 (UniProt). Nucleic acids research (2011) gkr981.

685 [40] W. Boyle. Energy recovery from sanitary landfills—a review. Microbial energy
686 conversion (Ed HG Schlegel and J Barnea) pergamon press, Oxford, UK (1976) 119-38.

687 [41] I. Angelidaki, B. Ahring. Thermophilic anaerobic digestion of livestock waste: the
688 effect of ammonia. Applied microbiology and biotechnology 38 (1993) 560-4.

689 [42] L. De Baere, M. Devocht, P. Van Assche, W. Verstraete. Influence of high NaCl and NH
690 4 Cl salt levels on methanogenic associations. Water research 18 (1984) 543-8.

691 [43] A. Ziembińska-Buczyńska, A. Banach, T. Bacza, M. Pieczykolan. Diversity and
692 variability of methanogens during the shift from mesophilic to thermophilic conditions while
693 biogas production. World journal of microbiology and biotechnology 30 (2014) 3047-53.

694 [44] R. Heyer, F. Kohrs, D. Benndorf, E. Rapp, R. Kausmann, M. Heiermann, et al.
695 Metaproteome analysis of the microbial communities in agricultural biogas plants. New
696 biotechnology 30 (2013) 614-22.

697 [45] F. Blume, I. Bergmann, E. Nettmann, H. Schelle, G. Rehde, K. Mundt, et al.
698 Methanogenic population dynamics during semi-continuous biogas fermentation and
699 acidification by overloading. Journal of applied microbiology 109 (2010) 441-50.

700 [46] R. Heyer, F. Kohrs, U. Reichl, D. Benndorf. Metaproteomics of complex microbial
701 communities in biogas plants. Microbial biotechnology 8 (2015) 749-63.

702 [47] D.L. Tabb, C.G. Fernando, M.C. Chambers. MyriMatch: highly accurate tandem mass
703 spectral peptide identification by multivariate hypergeometric analysis. Journal of proteome
704 research 6 (2007) 654-61.

705 [48] S. Theuerl, F. Kohrs, D. Benndorf, I. Maus, D. Wibberg, A. Schlüter, et al. Community
706 shifts in a well-operating agricultural biogas plant: how process variations are handled by the
707 microbiome. Applied microbiology and biotechnology 99 (2015) 7791-803.

708 [49] J. De Vrieze, L. Raport, H. Roume, R. Vilchez-Vargas, R. Jáuregui, D.H. Pieper, et al.
709 The full-scale anaerobic digestion microbiome is represented by specific marker populations.
710 Water research 104 (2016) 101-10.

711 [50] B. Müller, L. Sun, M. Westerholm, A. Schnürer. Bacterial community composition and
712 fhs profiles of low-and high-ammonia biogas digesters reveal novel syntrophic acetate-
713 oxidising bacteria. Biotechnology for biofuels 9 (2016) 1.

[51] J.J. Werner, M.L. Garcia, S.D. Perkins, K.E. Yarasheski, S.R. Smith, B.D. Muegge, et al. Microbial community dynamics and stability during an ammonia-induced shift to syntrophic acetate oxidation. Applied and environmental microbiology 80 (2014) 3375-83.

[52] A. Weiss, V. Jérôme, R. Freitag. Comparison of strategies for the isolation of PCR-compatible, genomic DNA from a municipal biogas plants. Journal of chromatography B 853 (2007) 190-7.

[53] J. De Vrieze, A.M. Saunders, Y. He, J. Fang, P.H. Nielsen, W. Verstraete, et al. Ammonia and temperature determine potential clustering in the anaerobic digestion microbiome. Water research 75 (2015) 312-23.

[54] F. Kohrs, S. Wolter, D. Benndorf, R. Heyer, M. Hoffmann, E. Rapp, et al. Fractionation of biogas plant sludge material improves metaproteomic characterization to investigate metabolic activity of microbial communities. Proteomics 15 (2015) 3585-9.

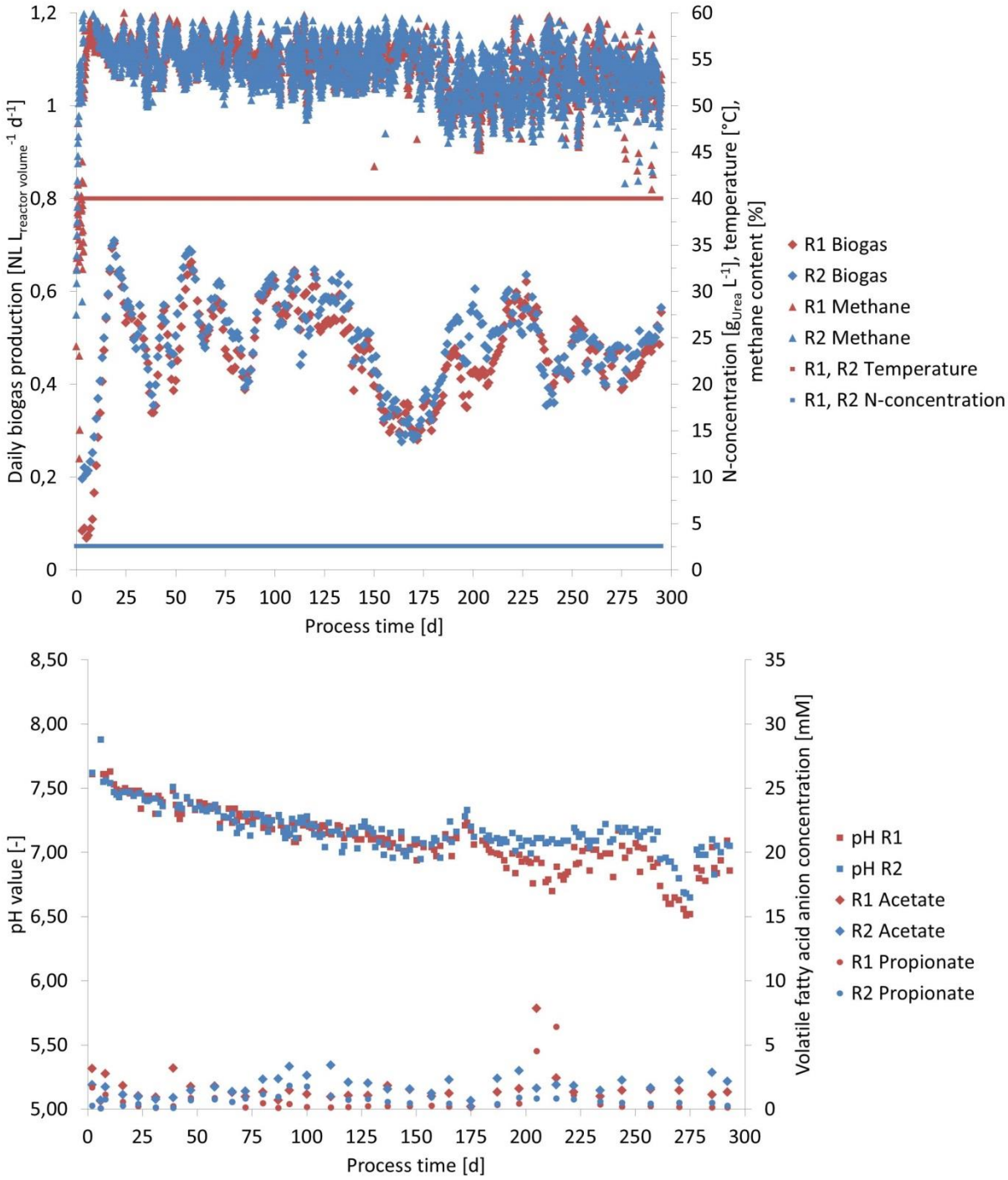
[55] N. Krakat, A. Westphal, K. Satke, S. Schmidt, P. Scherer. The microcosm of a biogas fermenter: comparison of moderate hyperthermophilic (60 C) with thermophilic (55 C) conditions. Engineering in Life Sciences 10 (2010) 520-7.

[56] A.N. Nozhevnikova, V. Nekrasova, A. Ammann, A.J. Zehnder, B. Wehrli, C. Holliger. Influence of temperature and high acetate concentrations on methanogenesis in lake sediment slurries. FEMS microbiology ecology 62 (2007) 336-44.

[57] A. Schnürer, Å. Nordberg. Ammonia, a selective agent for methane production by syntrophic acetate oxidation at mesophilic temperature. Water science and technology 57 (2008) 735-40.

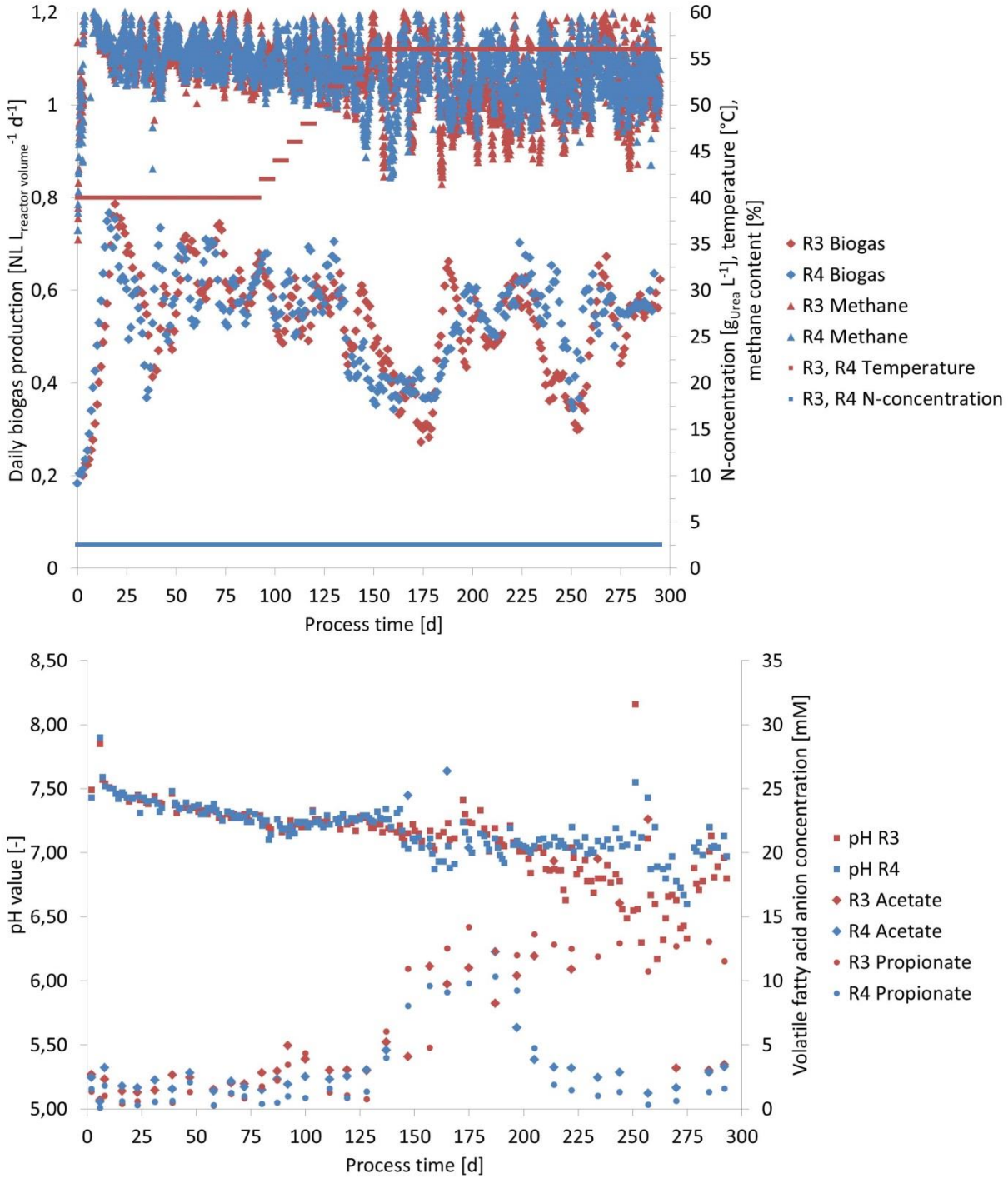
[58] S. Koch, D. Benndorf, K. Fronk, U. Reichl, S. Klamt. Predicting compositions of microbial communities from stoichiometric models with applications for the biogas process. Biotechnology for biofuels 9 (2016) 1.

Figures



739

740 **Figure 1.** Abiotic process data of the mesophilic regimes R1 and R2. Top: temperature, N-
741 concentration in the respective medium, methane content and daily biogas production for
742 reactors R1 and R2. The latter is given in NL $L_{reactor\ volume}^{-1}$ d^{-1} (gas volume at standard
743 conditions for temperature and pressure). Biogas production: unweighted centric moving
744 average with a width of seven days; methane content: values exceeding mean ± 2-fold the
745 standard deviation were defined as outliers and removed. Bottom: pH value, acetate
746 concentration, and propionate concentration.



747

748 **Figure 2.** Abiotic process data of the thermophilic regimes R3 and R4. Top: temperature, N-
749 concentration in the respective medium, methane content and daily biogas production for
750 reactors R3 and R4. The latter is given in NL $L_{reactor\ volume}^{-1}$ d^{-1} (gas volume at standard
751 conditions for temperature and pressure). Biogas production: unweighted centric moving
752 average with a width of seven days; methane content: values exceeding mean ± 2-fold the
753 standard deviation were defined as outliers and removed. Bottom: pH value, acetate
754 concentration, and propionate concentration.

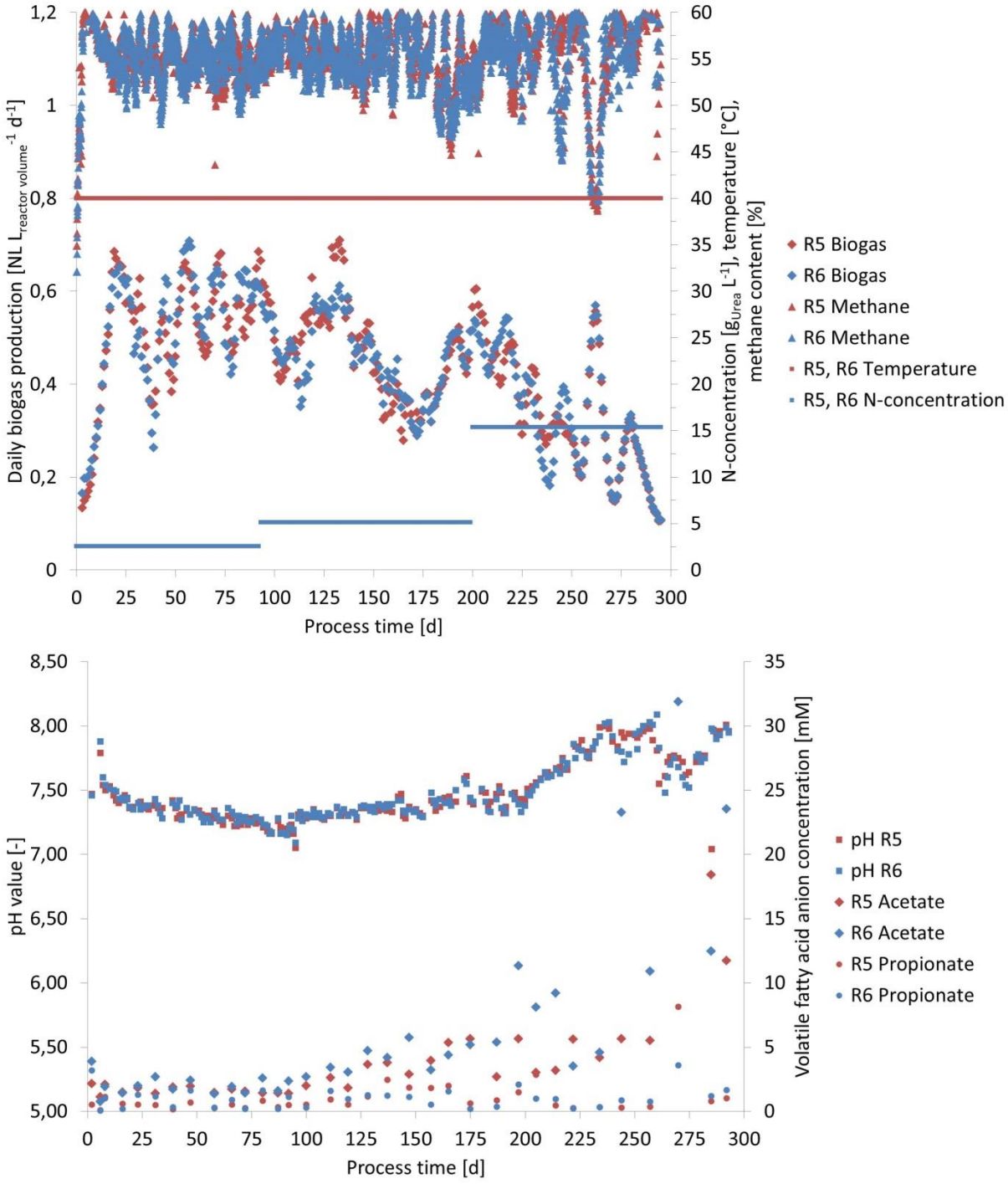


Figure 3. Abiotic process data of the high-N regimes R5 and R6. Top: temperature, N-concentration in the respective medium, methane content and daily biogas production for reactors R5 and R6. The latter is given in NL $L_{reactor\ volume}^{-1}$ d^{-1} (gas volume at standard conditions for temperature and pressure). Biogas production: unweighted centric moving average with a width of seven days; methane content: values exceeding mean ± 2-fold the standard deviation were defined as outliers and removed. Bottom: pH value, acetate concentration, and propionate concentration.

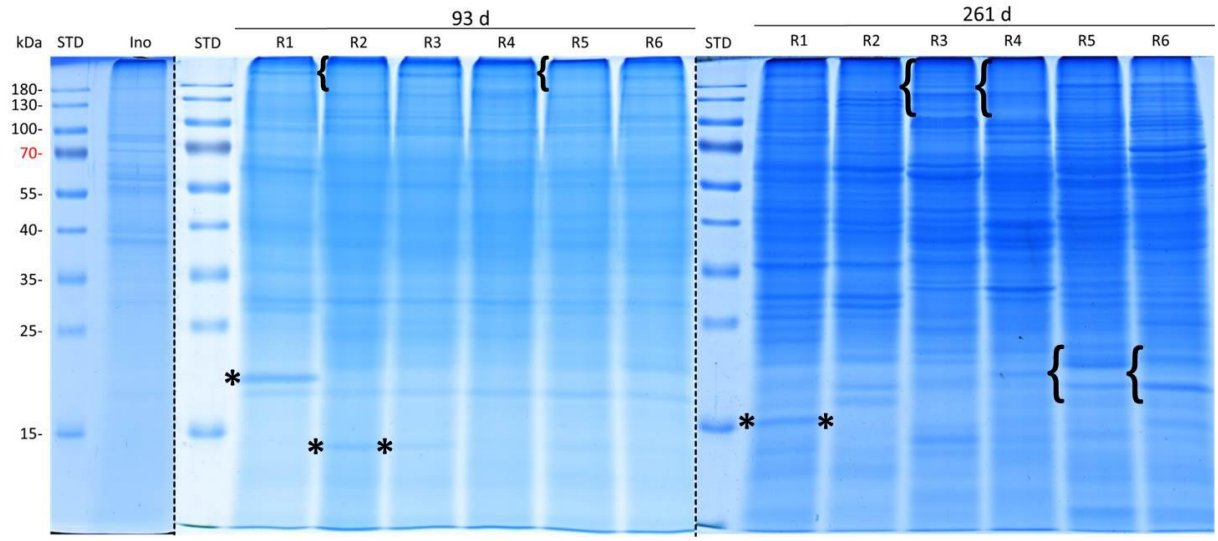


Figure 4. SDS-PAGE profiles of protein extracts from the inoculum and reactor R1 to R6 after 93 d and 261 d of operation. Protein patterns of the inoculum (Ino), mesophilic reactors (R1 and R2), thermophilic reactors (R3 and R4) and high-N reactors (R5 and R6); 25 µg of protein were loaded per lane and sample. Areas revealing differences highlighted by * and {.

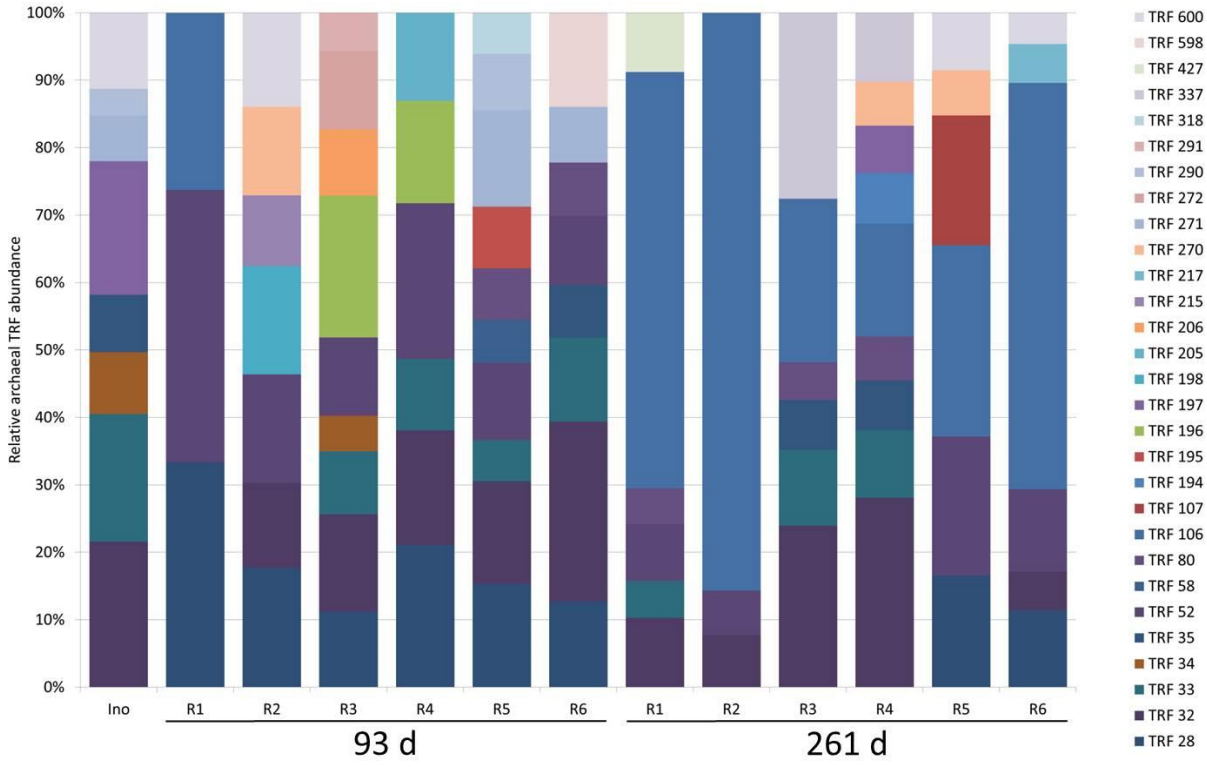


Figure 5. Archaeal terminal restriction fragments (TRF) derived from the inoculum and reactor R1 to R6 after 93 d and 261 d of operation. The TRF number on the right corresponds to the respective fragment base pair size. From each sample, three independent DNA extractions followed by two PCR assays were performed resulting in six electropherograms.

TRF were considered only when identified in more than three of the six electropherograms. After normalization to total intensities, only amplicons with a total relative abundance of at least 3% were taken into account.

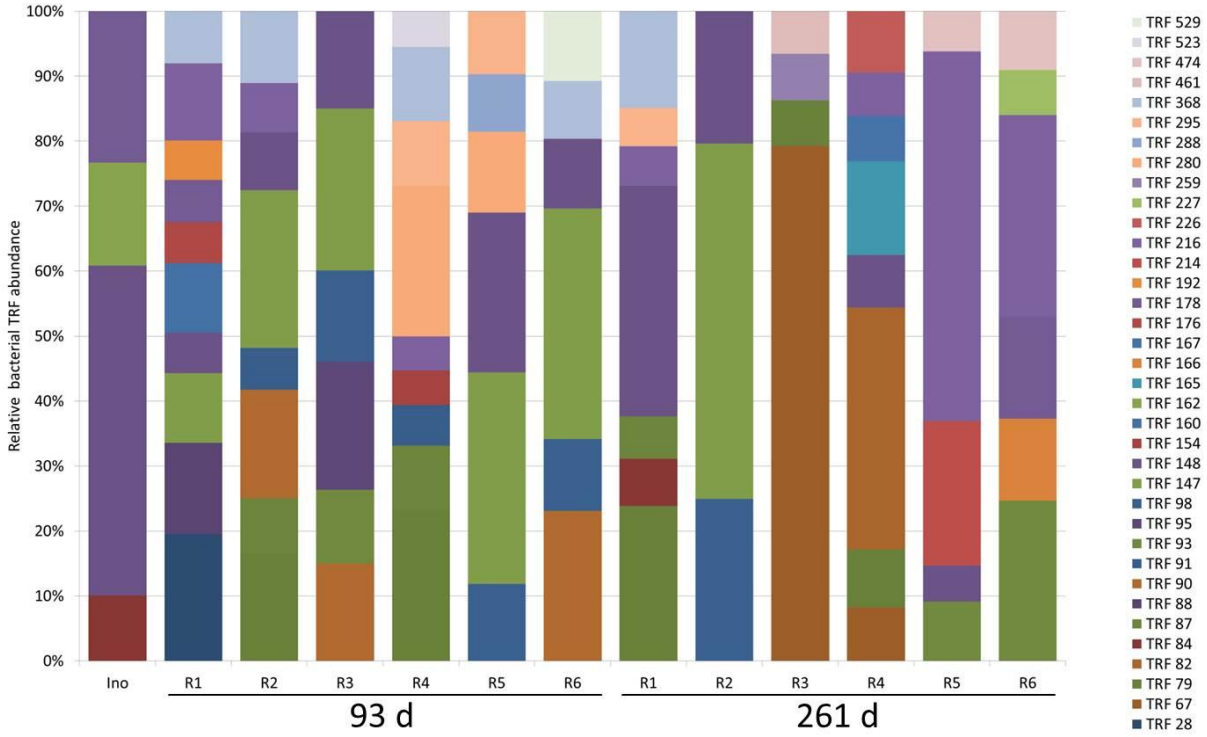
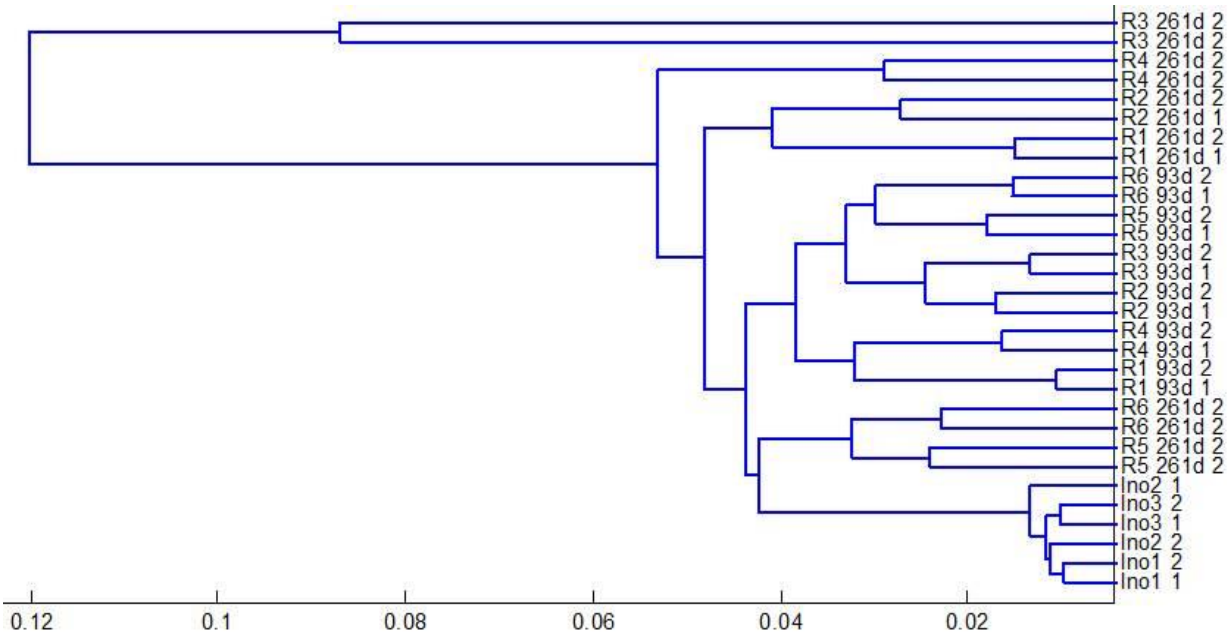


Figure 6. Bacterial terminal restriction fragments (TRF) derived from the inoculum and reactor R1 to R6 after 93 d and 261 d of operation. The TRF number on the right corresponds to the respective fragment base pair size. From each sample, three independent DNA extractions followed by two PCR assays were performed resulting in six electropherograms. TRF were considered only when identified in more than three of the six electropherograms. After normalization to total intensities, only amplicons with a total relative abundance of at least 3% were taken into account.



784

785 **Figure 7.** Hierarchical clustering of UniRef50 metaproteins from the inoculum and the
786 reactor samples after 93 d and 261 d of operation. Technical MS/MS replicates (suffix 1 or 2)
787 of samples at given time points for all reactor regimes (R1, R2: mesophilic; R3, R4:
788 thermophilic; R5, R6: high nitrogen load) were plotted to investigate reproducibility. Proteins
789 from the inoculum (Ino) were additionally digested after SDS-PAGE as technical triplicates
790 (Ino1-3). After normalization to the individual total spectral counts, only metaproteins with a
791 total abundance of at least 1% of a single sample overall spectral count were taken into
792 account.

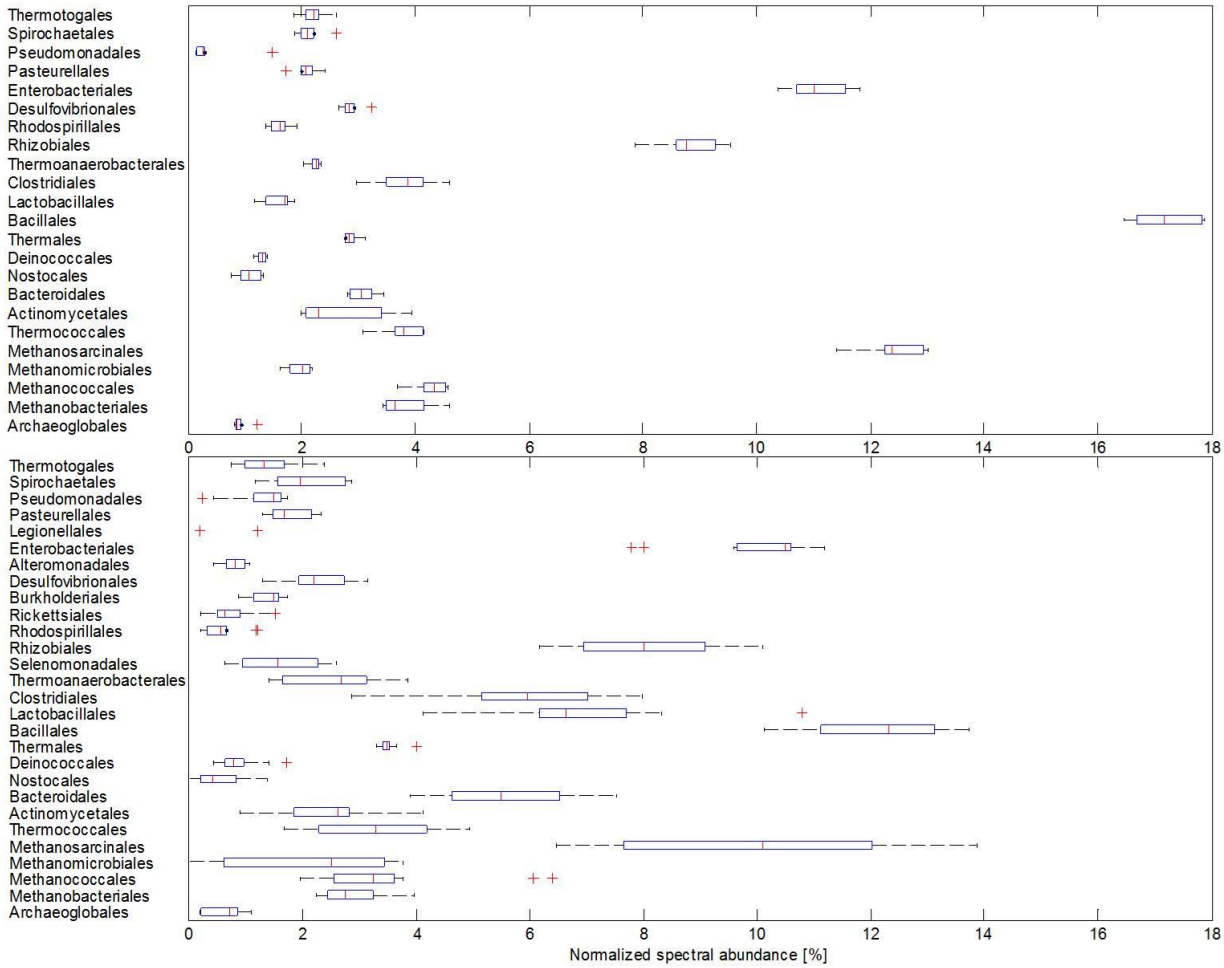
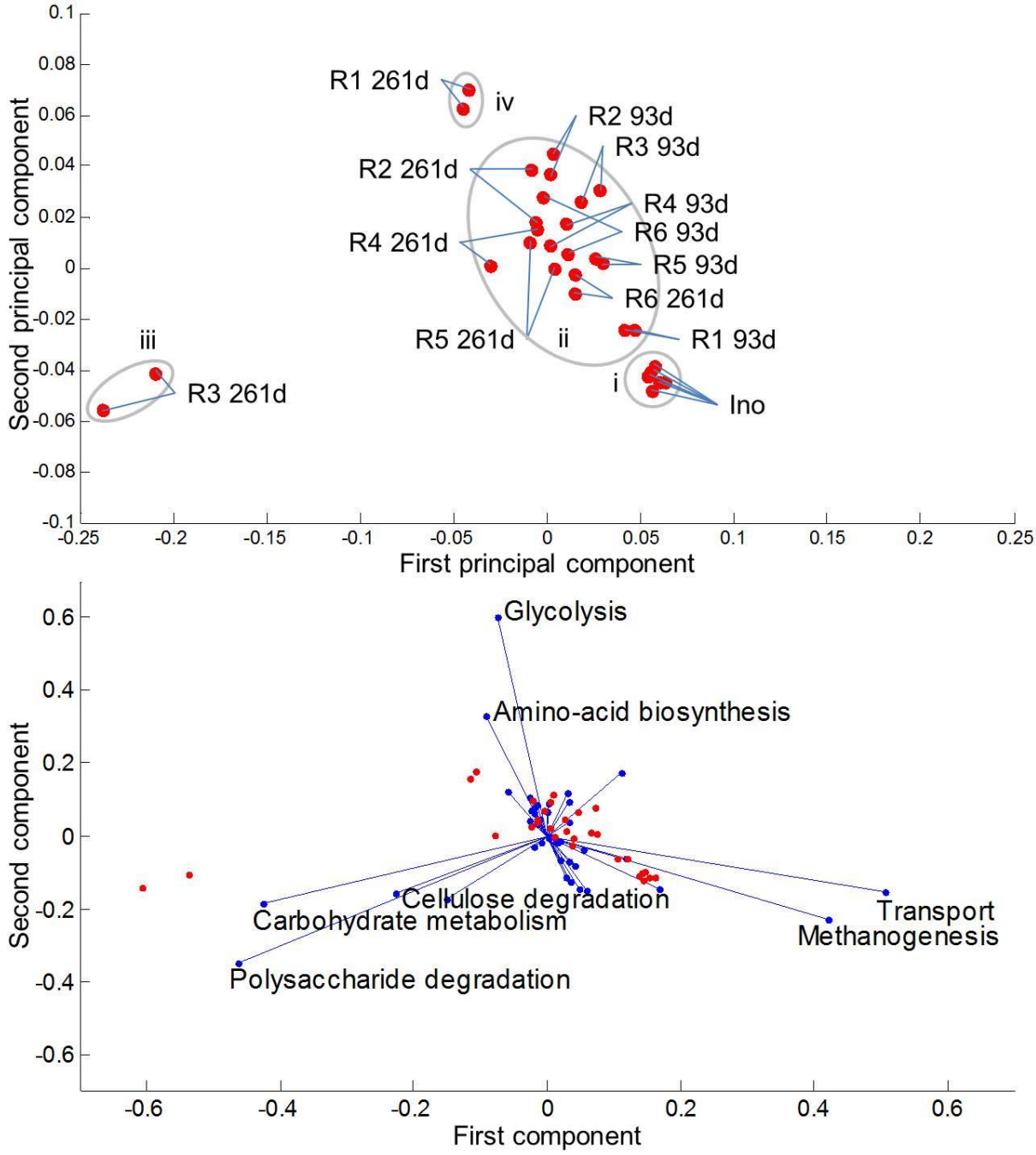


Figure 8. Box plots of taxonomies identified in R1-R6. Top: Samples of the inoculum, bottom: samples after 93 d of cultivation. Blue box: median 50% of data points, vertical dash: median value, left antenna: lowest value, right antenna: highest value, and cross: outlier. Only taxonomic orders of at least 1% of the overall spectral count of the individual sample after normalization were taken into account.



799

800 **Figure 9.** Principal component analysis of biological processes identified in R1-R6. Only
801 biological processes of at least 1% of the overall spectral count of the individual sample after
802 normalization were taken into account. Upper figure: Samples of the inoculum (Ino), after
803 93 d, and after 261 d of cultivation were plotted to investigate similarity (R1, R2: mesophilic;
804 R3, R4: thermophilic; R5, R6: high nitrogen load; two technical replicates per sample; i-iv
805 correspond to groups described in the text). Lower figure: loading plot of the underlying
806 data, indicating how biological processes (blue dots) contribute to a data point's position in
807 the main plot (red dots correspond to the sample points labelled in the upper figure).

Tables

Table 1. Medium and feeding characteristics for the different process regimes of R1-R6.

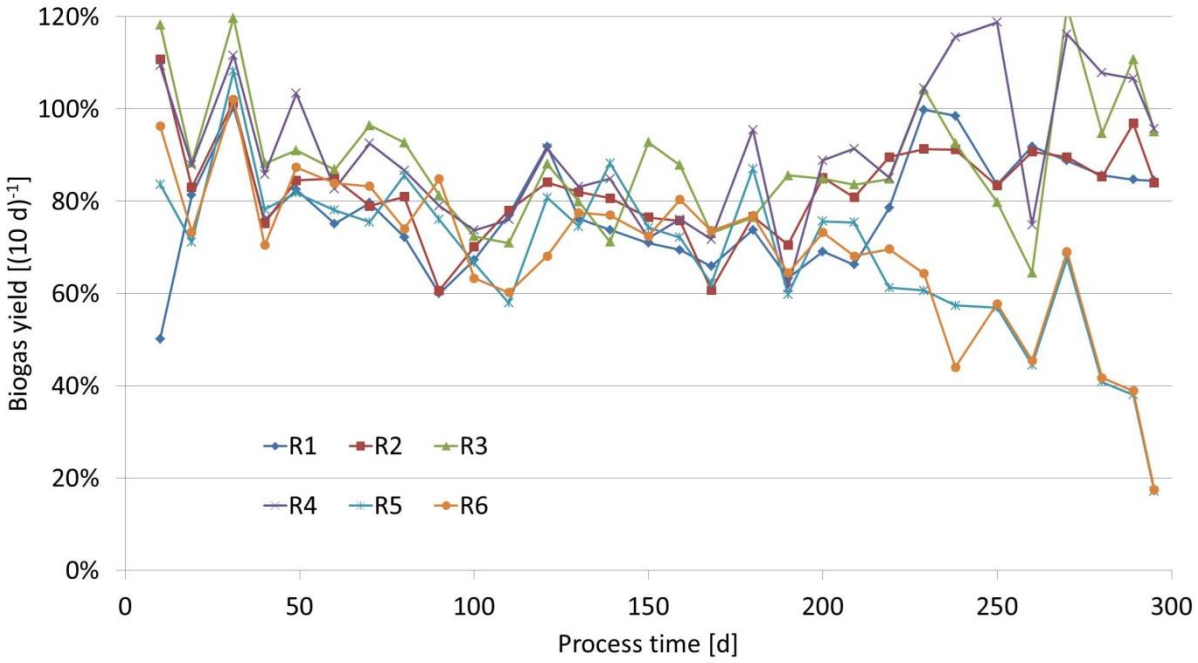
Process time [d]	Substrate [mL]	Urea in medium [g L^{-1}]		Molar C:N:P ratio of medium		HRT[1] [d]
		R1 to R4	R5, R6	R1 to R4	R5, R6	
0-10	5	2.56	2.56	105:7:1	105:7:1	200
10-14	7.5					71
14-93	10					62
93-200			5.13		108:14:1	
200-298			15.38		122:41:1	

[1]HRT: hydraulic retention time.

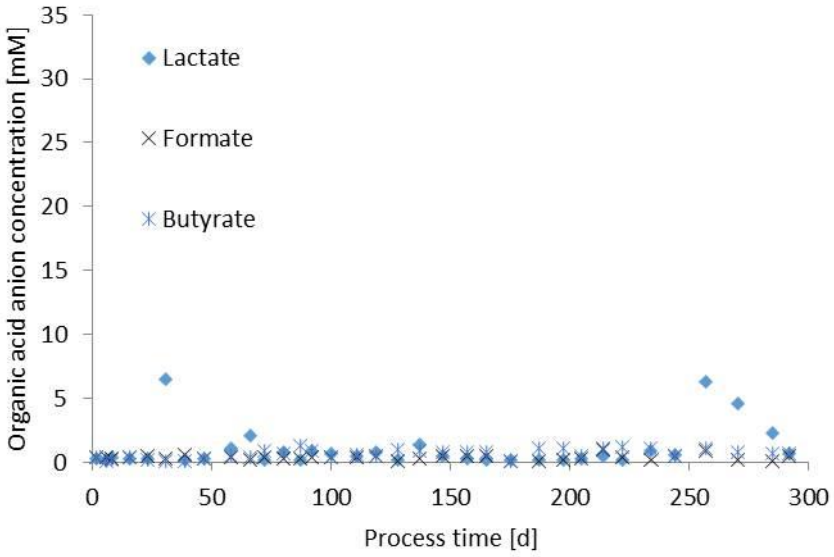
Table 2. Ammonium and ammonia concentrations of R1-R6 at given time points.

Process time [d]	N [g L^{-1}]	R1	R2	R3	R4	R5	R6
143	NH_4^+	0.93	1.03	0.94	1.16	1.76	1.78
	NH_3	0.02	0.02	0.06	0.09	0.08	0.07
183	NH_4^+	1.11	1.21	1.23	1.05	1.83	1.83
	NH_3	0.02	0.03	0.07	0.05	0.07	0.06
240	NH_4^+	0.86	1.08	0.83	0.70	5.05	5.07
	NH_3	0.01	0.02	0.02	0.03	0.60	0.66
292	NH_4^+	1.21	1.31	0.86	0.98	6.68	6.61
	NH_3	0.02	0.02	0.03	0.05	1.06	1.01

Supplementary

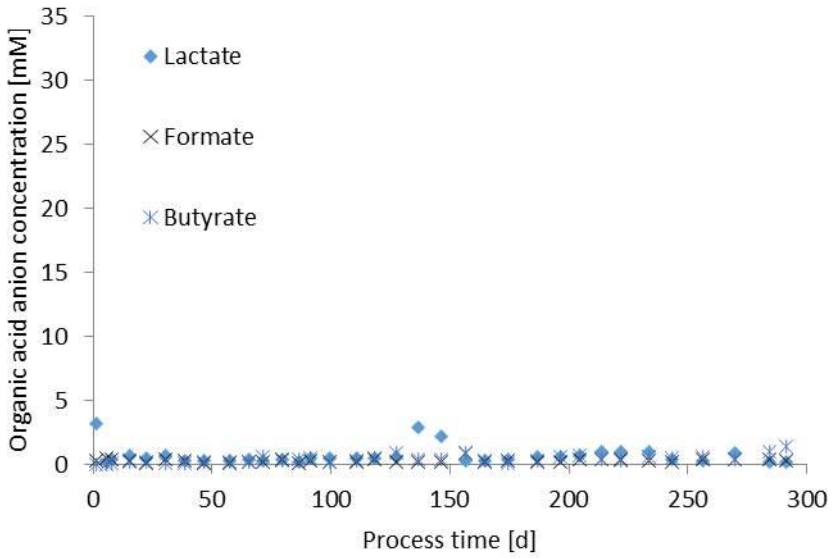


Supplementary_Figure 1. Biogas yield estimations for R1-R6. The respective yield was determined as the ratio of measured biogas volume to the potential biogas volume; For the latter, the total amount of all medium components was calculated for every ten days of cultivation (Table 1) according to Boyle et al. [40]. Values above 100% result from medium components with conversion covering two subsequent time intervals. Data points were connected for the sake of clarity.



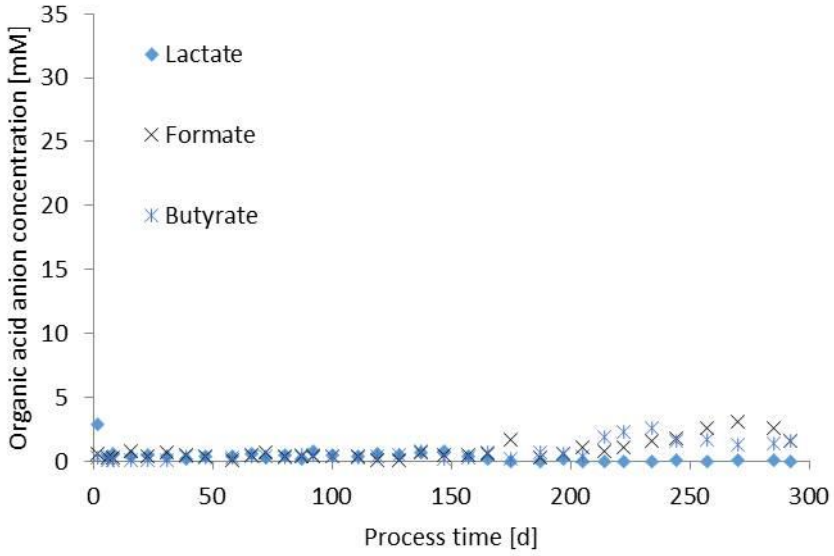
820

821 **Supplementary_Figure 2.** Lactate, formate and butyrate concentrations of R1.

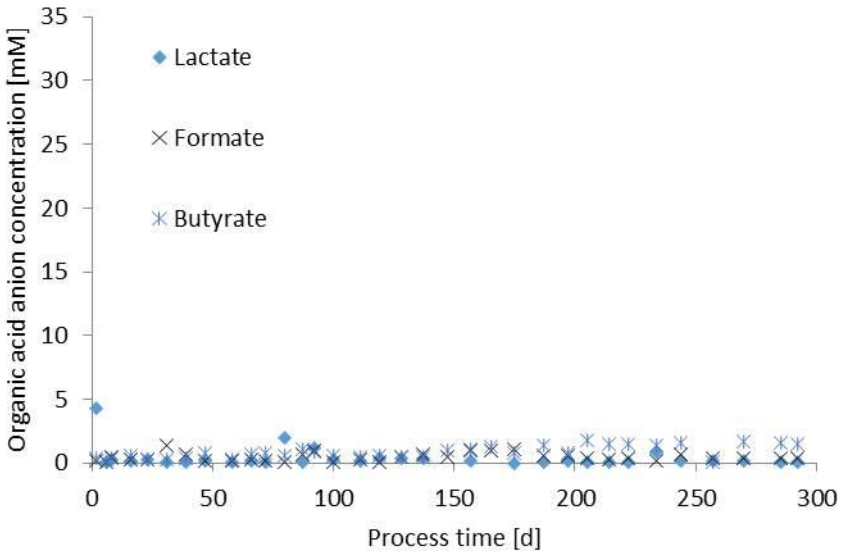


822

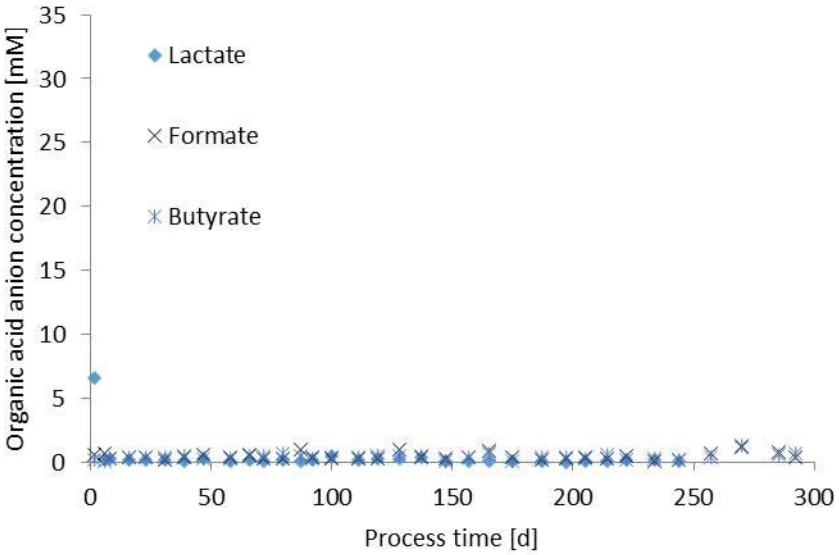
Supplementary_Figure 3. Lactate, formate and butyrate concentrations of R2.



Supplementary_Figure 4. Lactate, formate and butyrate concentrations of R3.

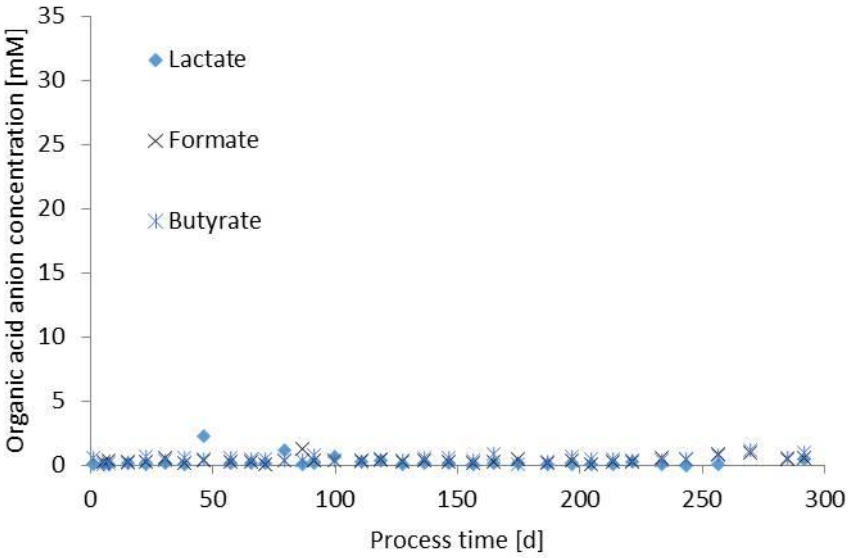


Supplementary_Figure 5. Lactate, formate and butyrate concentrations of R4.



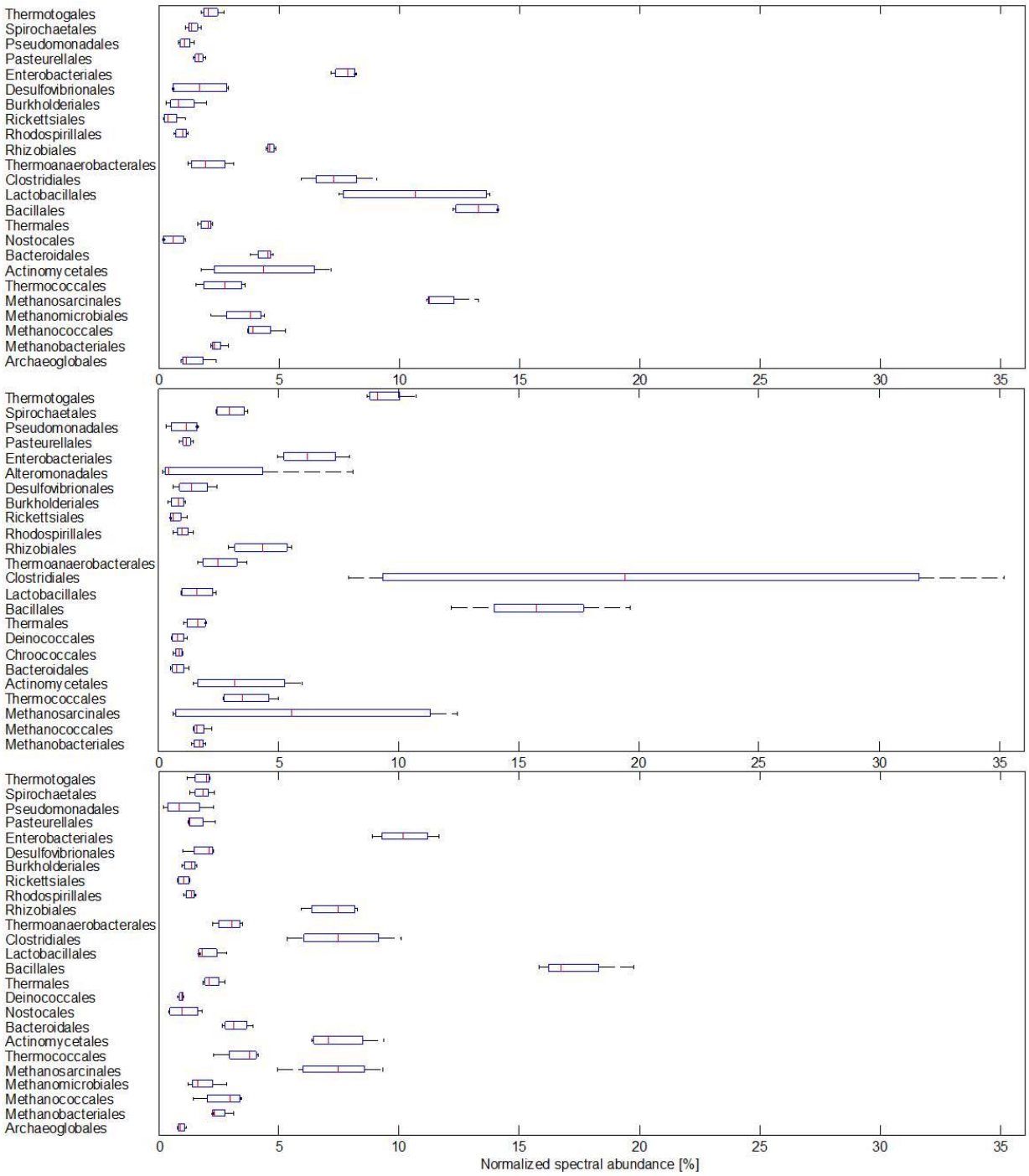
828

829 **Supplementary_Figure 6.** Lactate, formate and butyrate concentrations of R5.



830

831 **Supplementary_Figure 7.** Lactate, formate and butyrate concentrations of R6.



Supplementary_Figure 8. Box plots of taxonomies identified in R1-R6 at day 261. Top: Samples of R1-R2, middle: samples of R3-R4, bottom: samples of R5-R5. Blue box: median 50% of data points, vertical dash: median value, left antenna: lowest value, and right antenna: highest value, and cross: outlier. Only taxonomic orders of at least 1% of the overall spectral count of the individual sample after normalization were taken into account.

Supplementary_Note 1. Setup, process conditions and substrates of the BGP sampled for inoculations.

Supplementary_Note 2. Reactor-specific investigations of metaproteins. Identified metaproteins of each reactor were compared pairwise between day 93 and day 261. Only metaproteins identified with more than one spectrum each in both technical replicates of a sample were considered. Metaproteins considered as potential marker for an individual reactor if the sum of the spectral counts was either twice as high or halve for consecutive sampling points.

Supplementary_Note 3. Detailed description of methods used for metaproteomics and community fingerprinting using T-RFLP.