



Article

# A Novel Primer Mixture for GH48 Genes: Quantification and Identification of Truly Cellulolytic Bacteria in Biogas Fermenters

Regina Rettenmaier <sup>1</sup>, Yat Kei Lo <sup>1</sup>, Larissa Schmidt <sup>1</sup>, Bernhard Munk <sup>2</sup>, Ilias Lagkourdos <sup>3</sup>, Klaus Neuhaus <sup>3</sup>, Wolfgang Schwarz <sup>4</sup>, Wolfgang Liebl <sup>1</sup> and Vladimir Zverlov <sup>1,5,\*</sup>

<sup>1</sup> Chair of Microbiology, Technical University of Munich, Emil-Ramann-Str. 4, 85354 Freising, Germany; regina.rettenmaier@tum.de (R.R.); kei.lo@campus.lmu.de (Y.K.L.); schmidt\_larissa@mail.ru (L.S.); wliebl@wzw.tum.de (W.L.)

<sup>2</sup> Bavarian State Research Center for Agriculture, Central Department for Quality Assurance and Analytics, Lange Point 6, 85354 Freising, Germany; bernhard.munk@lfl.bayern.de

<sup>3</sup> ZIEL—Core Facility Microbiome, Technical University of Munich, Weißenstephaner Berg 3, 85354 Freising, Germany; ilias.lagkourdos@tum.de (I.L.); neuhaus@tum.de (K.N.)

<sup>4</sup> Aspratis GmbH, Munich, Germany, Hübnerstr. 11, 80637 Munich, Germany; schwarz@outlook.de

<sup>5</sup> Institute of Molecular Genetics of National Research Centre (Kurchatov Institute), Kurchatov Sq. 2, 123182 Moscow, Russia

\* Correspondence: vladimir.zverlov@tum.de

Received: 13 July 2020; Accepted: 23 August 2020; Published: 25 August 2020



**Abstract:** Genomic studies revealed the glycoside hydrolases of family 48 (GH48) as a powerful marker for the identification of truly cellulolytic bacteria. Here we report an improved method for detecting cellulolytic bacteria in lab-scale biogas fermenters by using GH48 genes as a molecular marker in DNA and RNA samples. We developed a mixture of primers for the specific amplification of a GH48 gene region in a broad range of bacteria. Additionally, we built a manually curated reference database containing GH48 gene sequences directly linked to the corresponding taxonomic information. Phylogenetic correlation analysis of GH48 to 16S rRNA gene sequences revealed that GH48 gene sequences with 94% identity belong with high confidence to the same genus. Applying this analysis, GH48 amplicon reads revealed that at mesophilic fermenter conditions, 50–99% of the OTUs appear to belong to novel taxa. In contrast, at thermophilic conditions, GH48 gene sequences from the genus *Hungateiclostridium* dominated with 60–91% relative abundance. The novel primer combinations enabled detection and relative quantification of a wide spectrum of GH48 genes in cellulolytic microbial communities. Deep phylogenetic correlation analysis and a simplified taxonomic identification with the novel database facilitate identification of cellulolytic organisms, including the detection of novel taxa in biogas fermenters.

**Keywords:** amplicon sequencing; qPCR; biomass hydrolysis; taxonomy; DNA; RNA

## 1. Introduction

One of the primary targets in bioeconomy is the production of biofuels such as biodiesel, bioethanol and biogas [1]. Crops like maize (whole plant) have been widely utilized for the production of biogas via anaerobic digestion. Lignocellulose—the fibrous material in the plant cell wall—accounts for 60% of total biomass on earth. Depending on its source, plant leaves and stems consist of lignocellulose: 35–50% cellulose, 20–35% hemicellulose, 10–25% lignin and small amounts of other components [2]. In biogas fermenters, the lignocellulosic substrate passes through several steps: (i) saccharolytic bacteria hydrolyze the polysaccharide components of biomass to soluble oligo- and monosaccharides;

(ii) acidogenic and acetogenic bacteria use the sugars and products from primary fermentation to produce mainly acetate, hydrogen and carbon dioxide; (iii) methanogenic archaea produce methane and carbon dioxide [3]. The first step in biomass fermentation, which is mainly driven by carbohydrate active enzymes (CAZymes), can be a bottleneck due to the recalcitrant parts of biomass such as highly lignified material or cellulose fibers with a high degree of crystallinity [4–6].

The application of high-throughput sequencing technologies enabled a deeper insight into the biogas microbiota [7]. In general, 16S rRNA gene sequences are used for the characterization of microbial communities [8–10], but the resolution of this method is limited especially when only shortened sequences (i.e., amplicons) of the gene are targeted [11]. In addition, 16S rRNA genes are inferior when functional groups of organisms are to be analyzed. Due to considerable functional variability within the taxa, taxonomic identification per se does not provide reliable information about the metabolic properties of an organism [12–16]. For this purpose, characterization of functional genes is advantageous allowing inferences about the function of interest. To this end, genes or transcripts of key enzymes of metabolic pathways can serve as a representative marker for a pathway of interest and the linked physiological features.

For the solubilization of crystalline cellulose, in bacteria the glycoside hydrolases of the hydrolytic enzyme families 5, 6, 8, 9 and 48 are crucial [17]. Remarkably, as far as studied to date, all organisms possessing at least one glycoside hydrolase of family 48 (GH48) are truly cellulolytic, i.e., can substantially degrade crystalline cellulose [17,18]. In contrast, most bacteria producing GH5 and GH9 enzymes can hydrolyze soluble (artificial) cellulose substrates such as carboxymethyl cellulose. However, they are unable to degrade crystalline cellulose. This makes GH48 a suitable molecular marker for the characterization of truly cellulolytic bacteria especially in anaerobic environments [14,16,19].

However, previous studies on GH48 gene sequences are based on genes known at their time. The steadily growing availability of metagenomes and an increasing number of bacterial genome sequences has led to a dramatic increase of identified GH48 gene sequences in public databases [20,21]. Despite this, there are only a limited number of open-access databases focusing on carbohydrate-active enzymes (CAZymes) like GHs. The CAZy database (<http://www.cazy.org/>) [21] is currently the most complete database documenting over 100,000 non-redundant CAZymes. Nevertheless, this database only provides protein sequences of GH48, which hampers gene analysis. The GenBank database at the National Center for Biotechnology Information (NCBI) contains about 26 bacterial GH48 enzymes from 9 bacterial genera with links to the corresponding nucleotide sequences. Yet, GH48 is known to be expressed by at least 50 genera [21,22]. Thus, an analysis solely based on the latter genes does not reflect the phylogenetic diversity of GH48, not to mention the genes in hitherto unknown species.

To fill this gap, a dedicated GH48 database composed of GH48 coding nucleotide sequences, the corresponding protein sequences, as well as the affiliated 16S rRNA gene sequences and corresponding taxonomy was constructed. So far, it proved not possible to develop a nucleotide primer pair for detecting and amplifying a broad range of GH48 genes [14,16,23]. To monitor as many GH48 gene sequences in a given metagenomic or metatranscriptomic sample as possible, the primers used for amplification should cover a broad range of variants. To this end, we improved GH48 as a molecular marker system for the identification of cellulolytic key players via GH48-amplicon sequencing. The developed novel primers were used to amplify GH48 gene sequences from DNA and RNA isolated from lab-scale biogas fermenters at four different process conditions. The identification of cellulolytic organisms in such fermenters helps to shed light on the microbial diversity at the specific fermenter stage and performance and thereby to come to a better understanding of the ‘black box’—biogas microbiome [7,24,25]. In addition, these four process conditions were analyzed to study the relative abundance of truly cellulolytic bacteria compared to the total copy number of bacterial 16S rRNA genes by using qPCR for quantification of the GH48 and 16S rRNA genes. The quantification of the latter genes can help to assess the importance of the first step in the overall process of microbial conversion of biomass to biogas.

## 2. Materials and Methods

### 2.1. Primer Development

Based on 87 bacterial GH48 gene sequences (Supplementary Material Table S1) and previous primers [14], a new mixture of 16 forward and nine reverse primers, designated cel48-Mix2, was developed. The GH48 gene sequences were aligned to the sequence of *Hungateiclostridium thermocellum* (ABN53296.1), formerly known as *Clostridium thermocellum*, to identify possible primer positions in each sequence. Chosen primer binding sites were aligned again individually and resulting groups were manually combined by introducing degenerate bases or deoxy-inosine [I]. Degenerate positions were used only if this increased number of combined species. The primer mix cel48-Mix2 and cel48\_490F\_I & cel48\_920R\_I (modified after Pereyra et al., 2010 [14]; see Tables 1 and 2) were tested for specificity and phylogenetic diversity. Amplicons were cloned using TOPO TA Cloning® pCR4 in *Escherichia coli* TOP10 (both Invitrogen, Carlsbad, CA, USA) and sequenced (Eurofins MWG). Resulting sequences, without primer, were aligned (ClustalW in MEGA7 [26]). To assess the phylogenetic diversity, Maximum-likelihood trees based on the Tamura–Nei model [27] with 1000 bootstraps were calculated for diversity assessment and sequences were compared to the NCBI database using BLASTx.

**Table 1.** Primer used for the amplification of target genes.

| Primer-ID    | Target Gene        | Nucleotide Sequence          | Source  |
|--------------|--------------------|------------------------------|---|
| Cel48_490F_I | GH48               | T[I]ATGGTTGAAGCTCCDGAYTAYGG  | Pereyra et al. 2010 [14], modified within this work         |
| Cel48_920R_I | GH48               | CCAAA[I]CCRTACCAGTTRTCAACRTC |   |
| F919mod      | bacterial 16S rRNA | GAATTGACGGGGRYCCGCACAAG      | Rudi et al. 1997 [28] modified by Lebuhn et al. 2014 [12]   |
| R1378mod     | bacterial 16S rRNA | CGGTGTGTACAAGRCCCGRGAACG     | Costa et al. 2006 [29], modified by Lebuhn et al. 2014 [12] |
| Cel48-Mix2F  | GH48               | Table 2                      | this work   |
| Cel48-Mix2R  | GH48               | Table 2                      |   |

Degenerative positions following International Union of Pure and Applied Chemistry (IUPAC) nucleotide code: D—AGT; R—AG; Y—CT.

**Table 2.** Summary of primers to amplify glycoside hydrolases of family 48 (GH48) gene sequences.

| ID   | Sequence                         | Length (bp) | V (μL) | N (Degeneration) |
|------|----------------------------------|-------------|--------|------------------|
| F1   | TCTTGAGTGAAGCTCCAGACTA           | 22          | 1      | 0                |
| F2   | TGCTTAGCGA[AG]GCGCCCAA           | 19          | 2      | 1                |
| F3   | GAGCGA[AG]GC[I]CC[I]GATTAC       | 19          | 2      | 1                |
| F4   | TGAGCGAGGCTCC[GT]GAC[CT]A        | 19          | 4      | 2                |
| F5   | TGATATGCGAAGCACCTGA[CT]TATG      | 24          | 2      | 1                |
| F6   | TCATATG[CT]GAAGC[GA]CC[TG]GATTAC | 23          | 8      | 3                |
| F7   | TGATATGCGAGGCACCGGATTA           | 22          | 1      | 0                |
| F8   | TGAACTGTGAAGCTCCTGATCAA          | 23          | 1      | 0                |
| F9   | TGTGTGTTGAAGCGCCTGATTAC          | 23          | 1      | 0                |
| F10  | TGATTGT[I]GAAGCTCC[TG]GA[CT]TATG | 24          | 4      | 2                |
| F11  | ATCGTCGAAGCGCCTGACCA             | 20          | 1      | 0                |
| F12  | ATCGACGAGGCGCCCGA                | 17          | 1      | 0                |
| F13  | TTTTGGTGGAAAGCTCCGGACTAT         | 23          | 1      | 0                |
| F14  | TAATGGTTGAAGCAC[II]GAC[CT]AT     | 23          | 2      | 1                |
| F15  | TCATCGTCGA[GA]GC[I]CC[I]GA[CT]TA | 22          | 4      | 2                |
| F16n | ATCGTCGAGGC[I]CC[GC]GACC         | 19          | 2      | 1                |
| R1   | CCGAAGCCGTAGACGTTGTC             | 20          | 1      | 0                |
| R2   | CCGTAACCGTAGATGTTGTC[I]A         | 22          | 1      | 0                |

Table 2. Cont.

| ID | Sequence                         | Length (bp) | V (μL) | N (Degeneration) |
|----|----------------------------------|-------------|--------|------------------|
| R3 | CCG[AT]AGCCGTA[GC][GC]TGTTGTC    | 20          | 8      | 3                |
| R4 | CCGTATCCATACCAGTTATCCGTA         | 24          | 1      | 0                |
| R5 | CCAAATCCATACCAGTTATCACAGT        | 25          | 1      | 0                |
| R6 | CCAAAACCATACCAGTTATCAACATC       | 26          | 1      | 0                |
| R7 | CCATCCATACCAGTT[AG]TC[I]AC[AG]TC | 24          | 4      | 2                |
| R8 | CCR[AT]A[I]CCGTAGACGTTGTC        | 20          | 4      | 1                |
| R9 | CCGAAGCCGTACCA[AG]TT[AG]TC       | 20          | 4      | 2                |

[I]—degenerative base.

## 2.2. Nucleic Acid Extraction from Biogas Fermenter Samples

Total DNA and RNA was extracted from digestate (GR from German: *Gärrest*) of lab-scale biogas fermenters fed with maize silage and operated at 38 °C (mesophilic, M) or 52–55 °C (thermophilic, T) at the Bavarian State Research Center for Agriculture (Freising, Germany). At either mesophilic or thermophilic reactor operation, different fermentation stabilities and performances were tested: ‘stable’ with no accumulation of volatile fatty acids and methane yield of 350–400 L<sub>STP</sub> CH<sub>4</sub> × kg<sub>VS</sub><sup>−1</sup> or ‘unstable’ with acidification due to the accumulation of volatile fatty acids and reduced methane yield. For the fermenter stages analyzed (mesophilic-stable, MS; mesophilic-unstable, MU; thermophilic-stable, TS; thermophilic-unstable, TU), see Supplementary Material Table S2. Next, *in sacco* (IS) enrichment, originally developed by Mohamed et al. [30], was performed to enrich cellulolytic bacteria that adhere to the cellulose fibers directly in their natural habitat. *In sacco* was conducted at each fermenter stage as described previously [31] with slight modifications: Briefly, nylon bags (ANKOM Technology, Macedon, NY, USA) with a size of 5 × 10 cm and a pore width of 50 μm were filled with 2 g shredded filter paper (Whatman No. 1), fixed on a custom-made sample support device and directly submerged into the lab-scale biogas fermenter. Two biological replicates were incubated for 5 days. Each *in sacco* sample was carefully washed with isotonic NaCl solution (0.9% (w/v)) to remove remaining digestate. The washed remains were used for DNA and RNA extraction with the EURx GeneMATRIX DNA + RNA + protein purification kit (Roboklon, Berlin, Germany) with cell lysis modifications as described in Rettenmaier et al., 2020 [32]. In brief, cell lysis was achieved by adding 300 μL ROTI-phenol (Carl-Roth, Karlsruhe, Germany) to 300 mg sample in a lysing tube with lysing matrix E (MP Biomedicals, Eschwege Germany) and bead-beating for 2 × 20 s at 5 m/s followed by centrifugation for 10 min at 10,000× g. Hundred microliter of clear supernatant were mixed with 200 μL Lyse ALL buffer and 300 μL DRP buffer and applied to the DNA-binding column (Roboklon). Subsequent purification of nucleic acids was performed following the manufacturer’s instructions. The DNA was eluted in 40 μL elution buffer, RNA was eluted in 40 μL RNase-free water.

## 2.3. (Quantitative) Polymerase Chain Reaction (qPCR)

Primers used are summarized in Table 1. Performing PCR with degenerate primers poses certain challenges for DNA polymerases. For successful amplification of GH48 gene sequences with cel48-Mix2, it was essential to use platinum *Taq* DNA polymerase (Thermo Fisher, Waltham, MA, USA). (q)PCRs were found specific and efficient when using 6-mM MgCl<sub>2</sub> and 2-μM primer mix each. Each 25-μL reaction (adjusted with water) contained 2.5 μL 10×*Taq* Buffer; 3 μL MgCl<sub>2</sub> (50 mM); 0.5 μL dNTPs (10 mM each); 0.5 μL cel48-Mix2F (100 pmol/μL); 0.5 μL cel48-Mix2R (100 pmol/μL); 0.2 μL Platinum *Taq* DNA polymerase. The amount of template varied from 5–50 ng. For qPCR, 1 μL of EvaGreen® (Biotium, Hayward, CA, USA) was added. The PCR was conducted as follows: initial denaturation for 10 min at 95 °C, followed by 26 or 40 cycles (for NGS or qPCR, respectively) of 95 °C for 20 s; 60 °C for 20 s and 72 °C for 20 s followed by a final elongation step for 2 min at 72 °C. For qPCR of 16S rRNA genes, a two-step PCR was applied (40 cycles 15 s at 95 °C and 1 min at 70 °C). To avoid quantification of primer-dimers in (RT)-qPCR of both 16S rRNA and GH48 genes, an additional step for 10 s at 83 °C after elongation was applied. Quantification by fluorescence was measured immediately after this step

(Supplementary Material Figure S1). Dissociation curves of amplicons were measured using EvaGreen as follows: 1 min at 95 °C, 30 s at 55 °C and 30 s at 95 °C using increments of 1 °C up to 95 °C.

#### 2.4. Reverse Transcription (RT)

Total RNA was cleaned from residual DNA with DNase (Turbo DNA-free™ kit, Ambion, Austin, TX, USA) following manufacturer's instructions. Gene-specific RT was performed using 5 µL cleaned RNA, 0.5 µL reverse primer (10 pmol/µL) and adjusted to 8 µL with DEPC-water (Invitrogen). The mixture was incubated 5 min at 65 °C and 10 min on ice. For GH48 genes, primer cel48\_920R\_I [14] or cel48-Mix2R and for 16S rRNA genes, primer R1378\_mod [33] were used. RT was performed by adding 2 µL dNTP-mix (10 mM of each nucleotide), 2 µL DTT, 2 µL buffer and 1 µL reverse transcriptase (Affinity Script Multi Temp, Agilent, Santa Clara, CA, USA) and incubation for 1 h at 45 °C, followed by 15 min at 70 °C. Negative RT-qPCR controls without reverse transcriptase were performed for each RNA extraction.

#### 2.5. DNA Standard for Absolute Quantification

For the quantification of GH48 gene sequences, a plasmid DNA standard with *cel48S* of *H. thermocellum* cloned into pET24c was chosen (TU Munich, München, Germany). Plasmids were extracted from *E. coli* DH10β cells via the QIAGEN Plasmid Midi Kit, following the manufacturer's instructions and linearized with *NdeI* (Thermo Fisher). Genomic DNA of *E. coli* DH10β, extracted from a cell pellet of a 50 mL overnight culture, was used as background DNA in the qPCR standard. For extraction of genomic DNA, the cell pellet was resuspended in 5 mL lysing buffer (200-mM Tris/HCl; pH 7.5, 35 mM EDTA, 75 mM NaCl, 1 mg/mL lysozyme) and incubated 1 h at 37 °C followed by 2 h at room temperature after addition of 500 µL SDS (10% *w/v*) and 500 µL proteinase K (1 mg/mL; AppliChem, Darmstadt, Germany). For removal of proteins, 2 mL of 5 M NaCl and 1/3 of the total volume chloroform/isoamyl alcohol (24:1, Carl-Roth) were added. The mixture was shaken (20 min, 180 rpm, room temperature) and centrifuged (10 min, 5000× *g*, room temperature). The purified DNA was precipitated from the supernatant with one volume isopropanol, centrifuged, washed with ethanol (70% *v/v*) and finally dissolved in 200 µL ddH<sub>2</sub>O. The final qPCR standard consisted of 50 ng/µL gDNA of *E. coli* with a serial dilution of 10<sup>3</sup>–10<sup>9</sup> GH48 gene copies/µL. The number of GH48 gene copies was calculated from the concentration of the linearized plasmid DNA as well as the molecular weight based on the nucleotide sequence (Available online: [http://www.bioinformatics.org/sms2/dna\\_mw.html](http://www.bioinformatics.org/sms2/dna_mw.html)). For the quantification of 16SrRNA gene sequences, genomic DNA of *E. coli* was used in serial dilutions as standard. The number of 16S rRNA gene sequences in the standard reaction was assessed via MPN-qPCR.

#### 2.6. GH48 Reference Database Construction

To link the GH48 sequences to taxonomic units, nucleotide sequences of GH48 and 16S rRNA gene sequences from the same species were required. Thus, to generate the database, 'glycosyl hydrolase family 48' was searched in NCBI RefSeq. All entries found, together with a GenPept file (i.e., full protein description), were downloaded. The taxonomy identification (txid) was retrieved from the GenPept files using a custom script (GenPept2txid.py, Supplementary Material Script S1). Genome assemblies (complete genome, chromosome, scaffold and contig) with RefSeq annotations were downloaded from the NCBI Assembly database using the txid with Batch Entrez [34]. From the GenBank files, features (species name, nucleotide sequence, protein sequences, accession number and gene product name) were extracted using the custom script Genbank2Features.py (Supplementary Material Script S2), which searches for genome assemblies with one of the protein accession numbers of interest. 16S rRNA gene sequences were extracted from the same GenBank using the custom script Genbank216S.py (Supplementary Material Script S3). Nucleotide sequences of the 16S rRNA genes are extracted from the GenBank files annotated as '16S ribosomal RNA'. For species with more than one 16S rRNA gene, the first found was chosen. Taxonomy was retrieved using the txid with taxonomizr (Available

online: <https://github.com/sherrillmix/taxonomizr/>). All information was combined using R (Available online: <https://www.r-project.org/>). For species and strains with more than one genome assemblies, only information extracted from one of the assemblies was retained. Identical sequences, GH48 modules or amplicons, in between different strains, species or genera were identified by clustering sequences at 100% identity using USEARCH [35]. The final database ('combined database') was exported as an Excel file (Supplementary Material File S1). All GH48-module sequences of this database are available as FASTA file (Supplementary Material File S2).

### 2.7. Phylogenetic Correlation of 16S rRNA Genes to GH48 Gene Sequences

GH48 genes are used as a molecular marker either full length (1800–2100 bp) or using a 344-bp amplicon [14]. For identity thresholds based on GH48, 16S rRNA gene identities were taken as base and correlated to GH48. For GH48-amplicon thresholds, either the genus identification or operational taxonomic unit (OTU) clustering at genus level was used. Sequences from the 'Combined database' were extracted into GH48-16S and GH48-16S-C datasets. For the former, GH48 amplicon and the corresponding full-length 16S rRNA sequences of any source were combined into the GH48-16S dataset; for the latter, GH48 amplicon and full-length 16S rRNA sequences extracted only from complete genomes were combined into the GH48-16S-C dataset. The amplicon sequences were extracted from full-length sequences in the 'Combined database' by aligning reference amplicon sequences using MUSCLE [36] and extracting the aligned region using MEGA7 [26]. Since the 'Combined database' was dominated by *Streptomyces*, such without valid species name were excluded from GH48-16S, but not from GH48-16S-C, since for the latter *Streptomyces* accounted for a smaller proportion. From both datasets, hypothetical proteins were also excluded.

To model the sequence identity correlation, GH48 amplicon sequences extracted and full-length 16S rRNA gene sequences were aligned separately with MUSCLE [36] and distance matrices of the alignment were calculated using MOTHUR [37]. For each dataset, distance matrixes were combined using tidyverse (Available online: <https://tidyverse.tidyverse.org/>) and reshape2 (Available online: <https://github.com/hadley/reshape>) and plotted in R with ggplot2 (Available online: <https://github.com/tidyverse/ggplot2>). The corresponding GH48 sequence similarities to 16S rRNA similarities were determined by fitting a polynomial model and a linear model. To evaluate the modelling, adjusted  $R^2$  ( $R_{adj}^2$ ), root-mean-square error (RMSE), mean absolute error (MAE) and predicted residual error sum of squares (PRESS) was computed in R using caret (Available online: <https://github.com/topepo/caret>) and DAAG (Available online: <https://CRAN.R-project.org/package=DAAG>).

Given that there is no universal agreement on 16S rRNA thresholds for species or genus classification, GH48 thresholds corresponding to 99%, 97% or 95% 16S rRNA gene sequence similarity, respectively, as proposed from several authors in the past years [11,38–40], may not accurately cluster reads from the same species or genus into the same OTU. To improve the reliability of the GH48 threshold, another threshold was computed based on taxonomic annotations in NCBI. To this end, GH48 amplicons and full-length module sequences were clustered at a range of sequence identity at 1% increment using USEARCH [35]. A clustering agreement was computed for each percentage of identity. The clustering agreement was defined as percentage of sequences belonging to the same genus or species (according to the taxonomic annotation in NCBI) that were being clustered into the same OTU. A 100% agreement indicated that all sequences were clustered into the correct genus or species.

### 2.8. Next-Generation-Sequencing (NGS) and Data Analysis

For sequencing with Illumina-MiSeq, adaptors were added to each primer within cel48-Mix2. Adaptor sequences are TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG (forward) and GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG (reverse). PCR reactions were cleaned using Ampure beads in accordance with the manufacturer's instructions and amplified in a second PCR reaction to attach sequencing adapters following Berry et al. [41]. Final libraries were pooled equimolarly and sequenced on a MiSeq system as recommended by the manufacturer (e.g., Reitmeier et al. [42]).

Data analysis of raw sequence reads was performed with an UPARSE-based analysis pipeline [43]. Demultiplexing was performed by `demultiplexor_v3.pl` (Perl script, available on request). Pairing, quality filtering and OTU clustering (94% identity) was done by USEARCH 8.0 [35]. Taxonomic classification was performed manually by pairwise sequence comparison via BLASTn with the database constructed within this work (see '2.5 GH48 reference database construction'). Sequence alignment was done by MUSCLE [36] and treeing by Fasttree [44]. The number of allowed mismatches in the barcode was 1. Minimal fastq quality score for trimming of unpaired reads was 20. Length for paired reads was set between 320–500 bp with a maximal rate of expected errors of 0.02 in paired sequences. Trimming at the forward and reverse side of the sequence was 10 bp. Statistical analysis was performed via Rhea [45] in R.

### 3. Results

#### 3.1. Primer Development

In silico GH48 primer development resulted in 16 different, partly degenerated forward primers in combination with 9 different, partly degenerative reverse primers. The best composition of primers was experimentally verified (see '2.1 Primer development' in Materials and Methods section) and resulted in the use of certain primer mixtures (Tables 1 and 2). The calculation of the mixture composition was based on the amount of degenerative positions within a single primer. A degenerative position was defined as a position with two possible bases. The introduction of the non-natural base deoxy-inosine [I], which was preferred rather than introducing a degenerative position with three or four bases, was not regarded as a degenerative position in this context. The fraction of a single primer was calculated with the following formula:  $V = 2^N$ , with N being the number of degenerations, see Table 2 for the detailed fractions of cel48-Mix2.

Comparison of the phylogenetic diversity of amplified GH48 gene sequences obtained with cel48-Mix2 or cel48\_490F\_I + cel48\_920R\_I [14] revealed strong overlaps. With both primer sets, nine different clusters of GH48 gene sequences were detected (Supplementary Material Figure S2 and Table S3). Although the primer design was only based on GH48 gene sequences of organisms with validly published species names, GH48 gene sequences with high homology to hitherto unknown bacteria were also amplified. Amplification via cel48-Mix2 led to the detection of new GH48 gene sequences with homology to *Paenibacillus* sp., which were not detected using cel48\_490F\_I and cel48\_920R\_I. Besides, for qPCR the cel48-Mix2 had also advantages compared to the formerly developed primers and a higher PCR efficiency was achieved (see '3.3 Quantification of GH48 and 16S rRNA gene sequences' below). Subsequently, amplification of GH48 gene sequences via cel48-Mix2 was adjusted for the Illumina MiSeq sequencer to increase the sequencing depth of the molecular marker (see '3.6 GH48 amplicon sequencing' below).

#### 3.2. Isolation of Metagenomic DNA and Metatranscriptomic RNA

Metagenomic DNA and metatranscriptomic RNA was isolated from two different fermenters, operated at either 38 or 52–55 °C, at four or two different stages (i.e., volumetric loads), respectively (Supplementary Material Table S2). Extraction of nucleic acids was performed either from raw digestate (GR) or from *in sacco* (IS) enrichment samples (see '2.2 Nucleic acid extraction from biogas fermenter samples' in Materials and Methods section), each in technical triplicates. Nucleic acids of the technical triplicates were pooled for further analysis (Supplementary Material Tables S4 and S5).

#### 3.3. Quantification of GH48 and 16S rRNA Gene Sequences

Criteria for trustworthy absolute quantification are a sufficient PCR efficiency (85–125%) as well as a sufficient y-intercept ( $C_t$  values ranging from 35–40) [12]. For the quantification of GH48 gene sequences amplified with cel48-Mix2, a PCR efficiency of 93% was achieved (Supplementary Material Figure S3). Furthermore, the number of GH48 gene copies should not be over- or underestimated

when using a plasmid-based GH48-gene standard as determined via the y-intercept (y-intercept: 37.7; Supplementary Material Figure S3). The PCR efficiency was lower (<90%) when the formerly published primers cel48\_490F\_I + cel48\_920R\_I [14] were used (data not shown). GH48 genes in the metagenome of 12 different samples derived from either GR or IS samples of the mesophilic or thermophilic biogas process (Supplementary Material Table S2) ranged from  $9 \times 10^4$  to  $2 \times 10^7$  gene copies/ $\mu\text{L}$ . Compared to the total amount of 16S rRNA genes, the overall GH48 gene copy numbers were low, ranging from 3–8 and 12–37 GH48 genes per 1000 rRNA genes in the DNA of GR or IS samples, respectively (Tables 3 and 4 for GR or IS samples, respectively).

**Table 3.** GH48 and 16S rRNA gene copy numbers quantified in DNA isolated from raw digestate samples of biogas fermenters.

| Sample | GH48<br>(Gene Copies/ $\mu\text{L}$ ) | Standard<br>Deviation | 16S rRNA<br>(Gene Copies/ $\mu\text{L}$ ) | Standard<br>Deviation | Ratio GH48/16S<br>rRNA (%) |
|--------|---------------------------------------|-----------------------|---|-----------------------|----------------------------|
| MS1    | $5.7 \times 10^6$                     | $\pm 1.1 \times 10^6$ | $7.1 \times 10^8$                         | $\pm 1.8 \times 10^8$ | 0.8                        |
| MS2    | $4.8 \times 10^6$                     | $\pm 3.1 \times 10^5$ | $1.5 \times 10^9$                         | $\pm 3.6 \times 10^8$ | 0.3                        |
| MS3    | $3.6 \times 10^6$                     | $\pm 4.7 \times 10^5$ | $1.2 \times 10^9$                         | $\pm 1.5 \times 10^8$ | 0.3                        |
| MI     | $5.4 \times 10^6$                     | $\pm 4.0 \times 10^5$ | $6.7 \times 10^8$                         | $\pm 1.5 \times 10^7$ | 0.8                        |
| TS     | $2.9 \times 10^6$                     | $\pm 2.9 \times 10^5$ | $8.0 \times 10^8$                         | $\pm 2.6 \times 10^8$ | 0.4                        |
| TI     | $7.9 \times 10^6$                     | $\pm 8.1 \times 10^5$ | $1.2 \times 10^9$                         | $\pm 6.8 \times 10^7$ | 0.7                        |

Absolute gene copy numbers calculated from the standard curve measured within the same qPCR. Standard deviation calculated from three technical replicates.

**Table 4.** GH48 and 16S rRNA gene copy numbers quantified in DNA isolated from *in sacco* samples of biogas fermenters.

| Sample | GH48<br>(Gene Copies/ $\mu\text{L}$ ) | Standard<br>Deviation | 16S rRNA<br>(Gene Copies/ $\mu\text{L}$ ) | Standard<br>Deviation | Ratio GH48/16S<br>rRNA (%) |
|--------|---------------------------------------|-----------------------|---|-----------------------|----------------------------|
| MS1    | $6.8 \times 10^6$                     | $\pm 8.7 \times 10^5$ | $2.2 \times 10^8$                         | $\pm 5.1 \times 10^7$ | 3.1                        |
| MS2    | $2.4 \times 10^7$                     | $\pm 9.6 \times 10^6$ | $6.4 \times 10^8$                         | $\pm 1.1 \times 10^7$ | 3.7                        |
| MS3    | $3.8 \times 10^6$                     | $\pm 1.4 \times 10^5$ | $1.9 \times 10^8$                         | $\pm 2.0 \times 10^7$ | 2.0                        |
| MI     | $9.1 \times 10^4$                     | $\pm 1.3 \times 10^4$ | $7.8 \times 10^6$                         | $\pm 8.4 \times 10^5$ | 1.2                        |
| TS     | $3.4 \times 10^6$                     | $\pm 4.0 \times 10^5$ | $2.4 \times 10^8$                         | $\pm 4.1 \times 10^7$ | 1.4                        |
| TI     | $5.8 \times 10^6$                     | $\pm 1.5 \times 10^6$ | $1.7 \times 10^8$                         | $\pm 1.4 \times 10^7$ | 3.5                        |

Absolute gene copy numbers calculated from the standard curve measured within the same qPCR. Standard deviation calculated from three technical replicates.

Interestingly, compared to the GH48 gene copy numbers obtained from analysis of the DNA samples, the sequence copy numbers were even lower in the RNA of the same fermenters. In some cases, no valid quantification for RNA samples was possible, as the copy numbers were below the linear detection limit of  $10^3$  GH48 gene copies/ $\mu\text{L}$  (data not shown). This was true for the mesophilic (M) samples derived from stable (S) or instable (I) fermenter stages MS1-GR, MS2-GR and MI-IS or the thermophilic (T) sample, TS-GR. The ratio of GH48 genes to 16S rRNA genes quantified in cDNA of MS1-IS, MS2-IS, MS3-IS, MS3-GR, MI-GR, TS-IS, TI-IS and TI-GR ranged from 6 GH48 genes per 1,000,000 16S rRNA genes to 3 GH48 genes per 10,000 16S rRNA genes and was highest in experiment MS3-GR and lowest in experiment MS3-IS, respectively (Supplementary Material Table S6). For a detailed description of all fermenter-stage analyses, see ‘2.2 Nucleic acid extraction from biogas fermenter samples’ in Materials and Methods section as well as Supplementary Material Table S2.

### 3.4. Reference Database Construction

The NCBI RefSeq protein database contains about 1200 bacterial GH48-module containing proteins, excluding proteins annotated as “multispecies”. All entries with a definite taxonomic identifier were downloaded from the NCBI Assembly database (146 complete genome assemblies, 22 chromosome assemblies, 402 scaffold assemblies and 484 contig assemblies, state 20-05-2019). The final ‘combined’ database for GH48 genes contains 944 entries. Each entry consists of the taxonomic identifier (txid), organism name, nucleotide accession number, nucleotide sequence of the GH48-module containing

protein, protein accession number, protein sequence, product name of the protein, 16S rRNA gene sequence and the expanded taxonomy of the organism. The coding sequences of GH48 modules range from 1429–7032 bp in our database.

The 944 entries mentioned above originate from 813 different species and strains belonging to 104 different genera of 42 bacterial families, while most the entries belongs to the genera *Streptomyces* and *Paenibacillus*. For a complete list of genera, see Supplementary Material Table S7. Out of 944 entries, 580 entries (61%) contained full-length 16S rRNA gene sequences (about 1500 bp). Two full-length 16S rRNA gene sequences contained ambiguous bases. Forty-five or 109 pairs with identical GH48-module nucleotide sequences or identical theoretical GH48 amplicon sequences (average length: 344 bp) were observed, respectively. Almost all pairs of identical GH48-module sequences belonged to strains of the same species, i.e., *Hungateiclostridium thermocellum* (former *Clostridium thermocellum*) DSM 1313 (ADU75731.1) and *H. thermocellum* AD2 (ALX09761.1) with an exception for *Phytoactinopolyspora* sp. YIM 96934 (WP\_112256909.1) and *Actinobacteria* bacterium YIM 96077 (AYY11565.1), which probably belong to the same species as an analysis of the total genome sequence suggests (data not shown). Only in rare cases, identical amplicon sequences classified as GH48 were observed between different species of one genus, i.e., *Streptomyces malaysiensis* (WP\_099013847.1) and *S. autolyticus* (AQA12037.1). The parameters of the database are summarized in Table 5, detailed information about the amplicon length distribution is shown as Supplementary Material Figure S4.

**Table 5.** Parameters of the GH48 database.

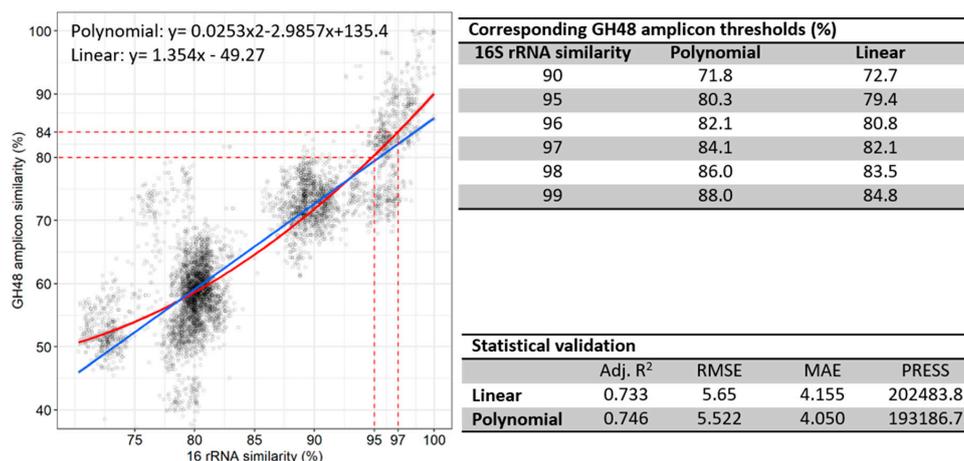
| Entries Summary                      |                | Sequence Length Distribution                       |      |      |      |
|--------------------------------------|----------------|--|------|------|------|
| No. of entries                       | 944            |  | Min  | Mean | Max  |
| No. of genera                        | 104            | GH48 module containing protein (nucleotide)        | 1429 | 2926 | 7032 |
| No. of families                      | 42             | GH48 module containing protein (amino acid)        | 475  | 975  | 2343 |
| Entries with full-length 16S         | 580<br>(61%)   | GH48 amplicon length (nucleotide, without primers) | 291  | 344  | 379  |
| Species or strains with >1 GH48      | 82<br>(8.7%)   | 16S rRNA gene sequence (nucleotide)                | 72   | 1459 | 1834 |
| Entries with identical GH48 module   | 78<br>(8.3%)   |  |      |      |      |
| Entries with identical GH48 amplicon | 184<br>(19.5%) |  |      |      |      |

### 3.5. Correlation of 16S rRNA to GH48 Gene Sequences

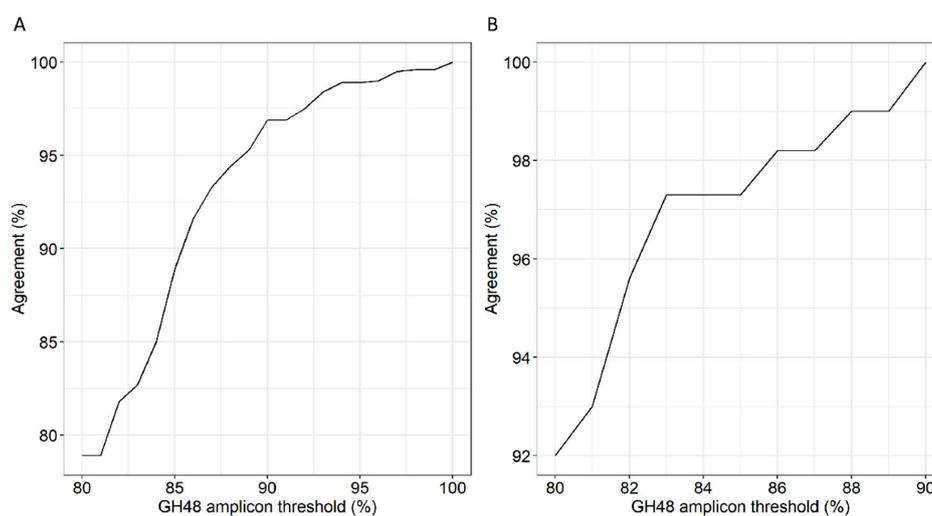
16S rRNA sequences are widely utilized for OTU clustering and species identification based on specific sequence similarity [7,38]. To establish GH48 as a specific molecular marker for identification and taxonomic classification of truly cellulolytic bacteria, GH48 sequence similarity corresponding to certain 16S rRNA gene sequence similarity was computed using the subset GH48-16S-C (see: ‘2.5 GH48 database construction’ Materials and Methods section). This dataset contains GH48 amplicons and full-length 16S rRNA gene sequences from complete genome assemblies only, excluding hypothetical genes. The GH48 amplicon thresholds obtained from the GH48-16S-C dataset corresponding to 95%, 97% and 99% rRNA sequence identity were estimated to be 79%, 82% and 85% sequence identity, respectively, according to the linear model with adjusted  $R^2 = 0.733$ , RMSE = 5.654, MAE = 4.155 and PRESS = 2,024,834 or 80%, 84% and 88% sequence identity, respectively, according to polynomial model with adjusted  $R^2 = 0.746$ , RMSE = 5.522, MAE = 4.050 and PRESS = 193,187 (Figure 1). For simplification, data for the subset GH48-16S (see Materials and Methods: ‘2.5 GH48 database construction’), which resulted in similar values, are not shown.

To ensure that the GH48 thresholds defined from 16S rRNA similarity approximated genus boundaries, GH48 amplicons from the subsets GH48-16S and GH48-16S-C were clustered using a range of sequence identity values. Thresholds, which enabled clustering the GH48 sequences at genus and species level according to their taxonomic annotation, were identified. Based on the GH48-16S dataset, 100% amplicon sequence identity was required to cluster all sequences correctly into different genera (Figure 2A). However, only 85% of the amplicons were correctly clustered at a threshold of 84%

for the amplicon sequence identity, which should correspond to the genus boundary at 97% 16S rRNA sequence identity (determined by means of pairwise alignments). In contrast, for GH48 amplicons extracted from complete genomes only (subset GH48-16S-C), 97.3% of the sequences were correctly clustered at the proposed genus boundary of 84%. Further, all sequences were correctly clustered into their correct genera at 90% for the GH48 amplicon sequence identity (Figure 2B). Sequence pairs misplaced while clustering shared 94.2% to 98.2% 16S rRNA gene sequence identity. Assuming that these species may be misclassified in the database, these outliers were excluded. Thus, 94% amplicon sequence identity was required to classify all amplicon sequences correctly and this was chosen as threshold for any subsequent analysis.



**Figure 1.** GH48 thresholds corresponding to 16S rRNA gene sequence similarity determined from 113 amplicons and full-length 16S rRNA gene sequences extracted from complete genome assemblies (GH48-16S-C subdataset). The corresponding GH48 sequence similarity to 16S rRNA similarities data were plotted by fitting a polynomial model (red) and a linear model (blue). To evaluate the modeling, adjusted R<sup>2</sup>, root-mean-square error (RMSE), mean absolute error (MAE) and predicted residual error sum of squares (PRESS) are shown.

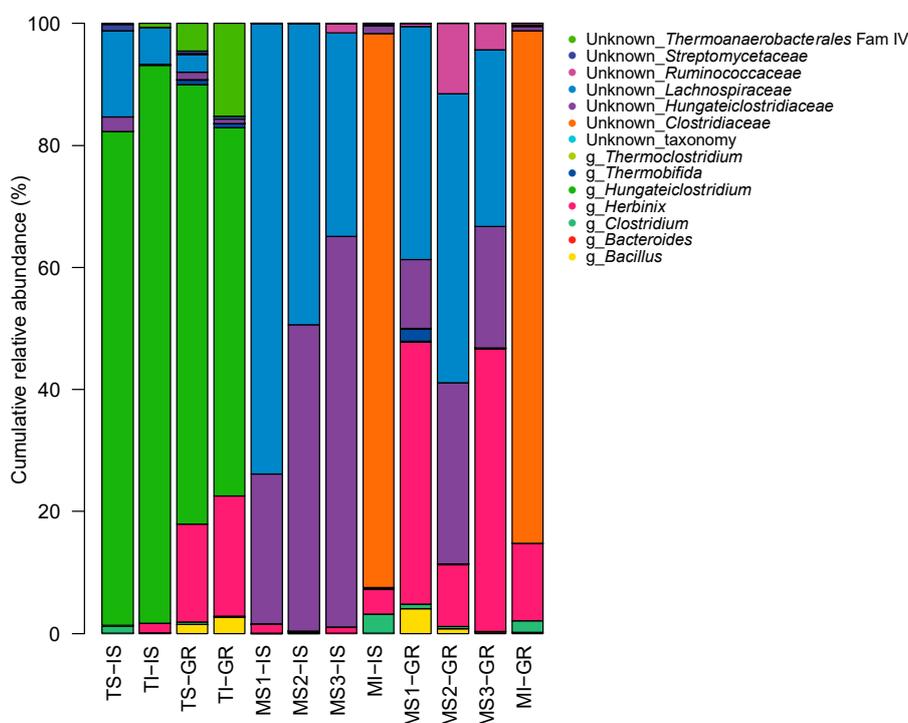


**Figure 2.** GH48 sequences clustering agreement with National Center for Biotechnology Information (NCBI) taxonomy. GH48 nucleotide sequences were clustered at a range of sequence identity at 1% increment using USEARCH [35]. (A) Clustering agreement of amplicon sequences based on GH48-16S dataset. All sequences were correctly clustered at genus level at 100% sequence identity; (B) Clustering agreement of amplicon sequences based on GH48-16S-C dataset. All sequences were correctly clustered at genus level at 90% sequence identity.

### 3.6. GH48 Gene Amplicon Sequencing

To demonstrate the applicability of GH48 sequencing for species identification, the composition of true cellulolytic bacteria in the metagenome of lab-scale biogas fermenters was investigated. Sequencing reads of GH48 gene amplicons, generated with the Illumina MiSeq sequencer, were clustered into different OTUs with 94% amplicon sequence identity as explained above.

Based on the GH48 classification outlined above, 17 different genera were identified on average in metagenomic DNA in all analyzed fermenter stages (Figure 3). Thereof, 50–99% belonged to unknown genera in mesophilic biogas fermentation. In contrast, only 7–18% of the GH48 gene sequences in the thermophilically operated biogas fermentation were unknown. In detail, metagenomic DNA of mesophilic fermenter samples MS1, MS2 and MS3 consisted in large proportions of GH48 gene sequences from unknown *Lachnospiraceae* (73.9%, 49.3% and 33.4% in IS and 38.1%, 47.4% and 29.0% in GR, respectively) and unknown *Hungateiclostridiaceae* (24.5%, 50.2% and 64.0% in IS and 11.3%, 29.7% and 20.0% in GR, respectively). Interestingly, the GH48 gene sequences in metagenomic DNA of instable, mesophilic fermenter stages (MI), either IS or GR, differed from the other samples, e.g., GH48 gene sequences of unknown *Clostridiaceae* were predominant (90.7% or 84.0%, respectively). In the thermophilic samples TS-IS, TI-IS, TS-GR and TI-GR, the GH48 gene sequences of *Hungateiclostridium* were most abundant (80.9%, 91.5%, 72.0% and 60.3%, respectively). Furthermore, GH48 gene sequences of unknown *Lachnospiraceae* were present in TS-IS (14.2%) and, in smaller parts, in TI-IS and TS-GR (6.0% and 2.9%, respectively), but were very low in TI-GR (0.4%). In contrast, GH48 gene sequences of *Herbinix* occurred in TS-GR and TI-GR (16.0% and 19.7%, respectively), less in TI-IS (1.5%) and very little in TS-IS (0.1%). A summary of the relative abundances for the different genera is given in Supplementary Material Table S9. A detailed analysis of the  $\alpha$ -diversity based on the GH48 amplicons of all samples is summarized in Supplementary Material Table S10. The relative abundance of the individual OTUs in each sample, along with the accession number and corresponding identities [%] to the best hit in the NCBI database refseq\_protein, identified via BLASTx, is given as Supplementary Material Table S11.



**Figure 3.** Taxonomic binning of GH48 gene sequences amplified in DNA of lab-scale biogas fermenters at genus level. TS—thermophilic stable; TI—thermophilic instable; MS—mesophilic stable; MI—mesophilic instable; IS—*in sacco*; GR—raw digestate.

#### 4. Discussion

With very few exceptions, the expression of GH48 module containing genes has been found to be a precondition of making bacteria truly cellulolytic [14,16,17]. This makes GH48 a valuable tool for identifying cellulolytic bacteria and to analyze their role in plant degrading ecosystems, a task that cannot be performed by using the 16S rDNA sequence alone. The novel primer mixture, consisting of 16 forward and 9 reverse primers, is a tool to quantify and identify cellulolytic bacteria via the amplification of GH48 gene sequences in DNA or RNA. Instead of using a single primer pair as performed by Pereyra et al., 2010 [14], the mixture is flexible in respect to expand the primer mixture with more individual primers for possibly novel evolving GH48 gene sequences in the future. In addition and in direct comparison to the primers of Pereyra et al., 2010 [14], a higher qPCR efficiency was achieved.

For the assignment of GH48 gene sequences to known organisms, a reliable reference database is necessary. Current GH48 gene containing databases comprise the dedicated CAZy database and general databases. However, despite CAZy.org [21] is a dedicated database on carbohydrate metabolic enzymes and documenting the protein name, organism name, protein accession number and experimental protein characterization data, this database lacks the nucleotide sequence allowing gene-based analyses. General databases like Uniprot ([www.uniprot.org](http://www.uniprot.org)) and PDB (Protein Data Bank, [www.rcsb.org](http://www.rcsb.org)) contain 375 entries from taxonomically classified bacteria [accessed 20-05-2019] and provide protein identification numbers and links to the gene for each GH48 enzyme. However, these databases also lack the ability to comprehensively search all GH48 genes. Thus, a comprehensive GH48 database having an emphasis on taxonomy was compiled in this study, allowing gene-based primer-finding and easy taxonomic classification. This novel database contains a taxonomic identifier (txid), the full taxonomic classification, the nucleotide sequences of GH48 and the 16S rRNA gene sequences of the corresponding organism (Supplementary Material File S1).

Similar to the 16S rRNA gene, in this work it is shown that the GH48 gene can be used as a molecular marker for genus (or even species) identification and OTU clustering using the full-length nucleotide sequence (about 2000 bp) or a highly conserved amplicon (about 344 bp). The choice of using the full-length sequence or the amplicon depends on the goal of the study. For species identification, use of the full-length GH48 nucleotide sequence is required as more identical amplicon sequences than identical full-length GH48 module nucleotide sequences were observed. On the other hand, identical full-length sequences were unlikely to be found in different species. However, modern sequencers for high-throughput (e.g., Illumina MiSeq, HiSeq 3000 and ThermoFisher Ion PGM) can only accurately sequence up to 150–400 bp per read [46], while the full-length gene sequences by far exceed this sequencing limit. Thus, for microbiota analysis, gene-specific amplicon sequencing is commonly used to assess microbial diversity. For biogas fermentation, the number and species of cellulolytic bacteria (i.e., based on GH48) are even more interesting than the number and species of bacterial organisms in general, based on 16S rRNA genes. However, a combination of both will be most informative.

GH48 thresholds for approximation of taxonomic boundaries were defined according to the corresponding gene sequence identities at certain 16S rRNA gene sequence identities using different subdatasets derived from the 'Combined database'. Modeling the correlation of GH48 gene sequence to 16S rRNA gene sequence identities revealed that RMSE, MAE and PRESS of the polynomial model were lower than that of the linear model (Figure 1). As expected, both amplicon (Figure 1) and module sequence identities did not reach 100% at 100% 16S rRNA gene sequence identity, because some species (although rarely) contain more than one GH48 (Table 5). For instance, *Hungateiclostridium thermocellum* ATCC27405<sup>T</sup> possesses two GH48: ABN53296.1 is a cellulosomal protein comprised of a GH48 module along with a dockerin type I domain (docI), whereas ABN51312.1 is a non-cellulosomal protein comprised of a GH48 module along with a carbohydrate binding module type 3 (CBM3). The GH48 modules of these two proteins share 54.48% identity on amino acid sequence level found using BLASTp. This also indicates that two proteins with redundant catalytic function, but distinct sites

of action (i.e., soluble in the extracellular space or bound to the outer cell membrane assembled within a multi-enzyme protein complex like the cellulosome, respectively) can occur within one species.

Assuming that 97% 16S rRNA gene sequence identity correspond to the species boundary [47], sequencing reads can be classified into one species at 84% GH48 amplicon identity (Figure 1). To cross-validate this, another method that, surprisingly, needed 100% sequence identity to classify reads correctly was applied (Figure 2). However, excluding outliers, 94% sequence identity was needed. The disagreement between the two methods, calculating a distance-matrix or a clustering agreement, is substantial, but probably due to horizontal gene transfer of GH48 [48], such that species from different genera possess similar GH48 sequences (Supplementary Material Table S7). In fact, a high probability of GH48 horizontal gene transfer is observed among *Actinobacteria* and *Firmicutes* [49]. Consistently, most of the genera with similar GH48 amplicon sequence identity but low full-length 16S rRNA gene sequence identity belonged to *Actinobacteria* and *Firmicutes* indeed (Supplementary Material Table S8).

Clustering of GH48-gene amplicon sequencing reads at the proposed threshold of 94% identity enabled a comprehensive analysis of the cellulolytic microbiota at genus level. The advantage of combining the reads at genus level is a simplification of the comparison of the different fermenter stages tested, since the amplified sequence segment does not allow for identification at species level. In any case, in all samples, an average of 17 genera was identified indicating a very low diversity of cellulolytic organisms compared to the overall diversity of the microbial community in biogas fermenters [7]: e.g., 95 genera of Bacteria and Archaea in 21 full-scale biogas digesters [50] or 1641 bacterial 16S rRNA OTUs identified in 43 household biogas digesters [51].

Despite the low diversity of just 17 genera of cellulolytic genera in each experiment, a large number of these cellulolytic taxa are found to be novel, especially at mesophilic conditions, when using high-throughput sequencing of GH48 amplicons and the novel primer mix for amplification. The previous studies from Pereyra et al. 2010 and Izquierdo et al. 2010 compare either different feeding of sulfate-reducing bioreactors [14] or different thermophilic, cellulolytic enrichment cultures [16], respectively and used cloning of PCR products for gene sequencing instead of next-generation sequencing methods as presented in this study. Another more current study highlights that analyzing the genetic diversity, here using GH48 genes in soil [52], of different environmental samples can subsequently lead to the identification and characterization of novel enzymes, which may exhibit an improved function of interest.

The results obtained within this study clearly indicate that at mesophilic fermenter operation conditions most of the detected GH48 gene sequences cannot be assigned to a known organism. Thus, there is a high potential of identifying novel enzymes, indicating potentially new cellulolytic bacterial taxa. Hence, the results stress the urgent need of further isolation, characterization and genome sequencing of novel cellulolytic organisms. Detailed taxonomic identification of metagenomic data and subsequent characterization of enzymes of interest is only possible with a number of closely related reference species [53]. Only the recent characterization of organisms within the genus *Herbinix*, isolated in 2015 [54] and 2016 [55], enabled the assignment of amplicon reads to this genus in this study and also made the enzymatic characterization of the arabinoxylan-degrading enzymes of *H. hemicellulosilytica* possible, as described by Mechelke et al., 2017 [56]. Further, the recently isolated novel member of the genus *Hungateiclostridium*, *H. mesophilum* [57], confirms that there are still novel organisms within the cellulolytic biogas microbiota to be analyzed for gaining deeper insights into substrate degradation during the microbial biogas process.

Evaluating cellulolytic activity of hydrolytic bacteria at different fermenter stages using sequencing and absolute numbers of GH48 gene copies may be biased for each experimental condition due to differences in the quality of the extracted nucleic acids. Contamination of PCR templates, e.g., with humic acids, have a great impact on amplification efficiencies [58]. Therefore, the cellulolytic potential of a fermenter condition was defined as the ratio of bacterial GH48 gene copies to the total amount of bacterial 16S rRNA gene sequence copies (GH48/16S), quantified within the same nucleic

acid sample. Admittedly, the ratio can be biased due to the different copy numbers of ribosomal RNA operons or possession of one or more GH48 genes in each strain. Interestingly, soil bacteria have an average of 1.4 or 5.5 copies of 16S rRNA genes in slow or fast growing organisms, respectively [15]. Consequently, the ‘true’ cellulolytic potential may be underestimated by a simple GH48 amplicon over 16S rRNA amplicon ratio. However, this ratio gives a hint into a relative cellulolytic potential at a certain time point and enables comparison of different samples or fermenter stages as shown in this study.

In general, higher amounts of GH48 genes in the metagenome, and in almost all samples of the metatranscriptome (except MS3), was observed in the *in sacco* enrichments (Table 4) compared to the GR samples of the same time point (Table 3). This was to be expected, since *in sacco* is a well-established method to enrich for cellulolytic organisms in their natural environment [19,30,31,59]. Nevertheless, to assess the cellulolytic potential of a certain biogas fermenter stage, the overall GH48 to 16S rRNA gene ratio in the digestate samples (GR) is more important as it represents the total bacterial community involved in the biogas process.

Interestingly, the observed values indicate higher ratios at the inefficient state MS1 as well as the instable state MI and TI compared to the stable and more efficient MS2, MS3 and TS (Tables 3 and 4). Therefore, we conclude that the overall number of cellulolytic organisms increases with a higher volumetric load of the fermenters. Subsequently, this can lead to stress in the overall biogas process [60]. In contrast, the GH48-to-16S rRNA ratios in the metatranscriptome reflects the efficiency of the biogas fermenters: ratios are highly reduced in the inefficient states compared to the more efficient ones, although keeping in mind that quantification of GH48 gene copies is below a valid detection limit for some of the very inefficient fermenter states (Supplementary Material Table S6). These results indicate the importance of preferably analyzing the metatranscriptome instead or in addition to the metagenome to monitor the active community in a certain sample, but also reflect the emerging challenges, e.g., low extraction yields, while working with RNA samples.

## 5. Conclusions

The novel primer combination cel48-Mix2 can be applied efficiently for qPCR and RT-qPCR as well as for amplicon sequencing to quantify and identify cellulolytic bacteria in biogas fermenters and, presumably, also in other complex habitats. Since the novel primer mix covers a wide spectrum of GH48 gene sequences, a more accurate comparison of different environmental samples is possible. In addition, the novel database for GH48 genes facilitates the taxonomic identification of cellulolytic organisms via GH48 gene sequences. The proposed thresholds enable fast OTU clustering based on GH48 sequences, as well as direct taxonomic identification of novel metagenomic and metatranscriptomic sequencing data by comparison with the provided database. Moreover, sequencing of GH48 amplicons enabled identifying novel cellulolytic taxa, as was demonstrated in this study, in particular for biogas fermentations operated under mesophilic conditions. Therefore, for these samples, targeted isolation and characterization of the latter organisms may be important to elucidate the biogas microbiota in more detail. The novel primer mixture cel48-Mix2, applied to other environmental samples, will help identifying a greater variety of unknown cellulolytic organisms and, subsequently, aiding in their isolation.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2076-2607/8/9/1297/s1>, Table S1: GH48 gene sequences used for the design of cel48-Mix2, Table S2: Lab-scale biogas fermenters, Table S3: BLASTx results to the NCBI refseq\_protein database of GH48 gene sequences amplified via cel48\_490F\_I + cel48\_920R\_I [4] or cel48-Mix2, Table S4: Nucleic acid extraction from lab-scale fermenter operated at 38 °C, Table S5: Nucleic acid extraction from lab-scale fermenter operated at 55/52 °C, Table S6: GH48 and 16S rRNA gene copy numbers quantified in cDNA of biogas fermenters, Table S7: List of genera in the ‘Combined’ database, Table S8: Outliers in GH48-16S dataset, Table S9: Relative abundances [%] of GH48 gene sequences in the metagenome at different fermenter stages, summarized at genus level, Table S10:  $\alpha$ -diversity of GH48 amplicon sequencing in metagenomic DNA, Table S11: Relative abundance of GH48 OTUs from metagenomic DNA of biogas fermenter samples in addition to the BLASTx result (highest score) in the NCBI refseq\_protein database, Figure S1: Dissociation curve of PCR products amplified with cel48-Mix2 in the qPCR setup, Figure S2:

Phylogenetic tree of GH48 gene sequences amplified from the metagenome of a mesophilically operated lab-scale biogas fermenter using cel48\_490F-I+cel48\_920R-I or cel48-Mix2 (highlighted in blue), Figure S3: Standard for the absolute quantification of GH48 gene sequences with linearized pET24c:cel48S and gDNA of *E. coli* as background, Figure S4: Amplicon length distribution, Script S1: GenPept2txid.py, Script S2: GenBank2Features.py, Script S3: GenPept216S.py. File S1: GH48\_database.xlsx, File S2: GH48\_modules.fasta.

**Author Contributions:** Conceptualization, R.R.; W.S., V.Z. and W.L.; methodology, R.R., Y.K.L., L.S. and B.M.; software, R.R., Y.K.L., K.N. and I.L.; validation, R.R., Y.K.L. and V.Z.; formal analysis, R.R., Y.K.L., L.S. and V.Z.; investigation, R.R., Y.K.L., L.S. and V.Z.; resources, B.M., K.N., I.L., V.Z. and W.L.; data curation, R.R., Y.K.L., K.N. and I.L.; writing—original draft preparation, R.R., Y.K.L., W.S. and V.Z.; writing—review and editing, K.N., B.M., I.L. and W.L.; visualization, R.R. and Y.K.L.; supervision, V.Z. and W.L.; project administration, V.Z. and W.L.; funding acquisition, V.Z. and W.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the German Federal Ministry of Food and Agriculture, Grant Number 22021715 and 22003513.

**Acknowledgments:** The authors thank Bernward Klocke (Intrexon Bioinformatics, Germany) for his help in NGS data analysis, Caroline Ziegler (ZIEL—Core Facility Microbiome, TU Munich, Freising, Germany) for NGS library preparation and Georgina Mikacevic (Bavarian State Research Center for Agriculture, Freising, Germany) for technical maintenance of the biogas fermenters.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Data Availability:** The NGS raw sequencing reads of the GH48 gene sequences in metagenomic DNA of the analyzed biogas fermenters are deposited at the Sequence Read Archive, NCBI, available under the accession numbers SRR11545268-79, BioProject no. PRJNA625239 and the corresponding BioSamples nos. SAMN14595678-89.

## Abbreviations

|                    |  |
|--------------------|--|
| CAZy(mes)          | carbohydrate-active enzyme(s)            |
| ddH <sub>2</sub> O | doubled deionized water                  |
| GH48               | glycoside hydrolase of family 48         |
| GR                 | digestate (from German: <i>Gärrest</i> ) |
| IS                 | <i>in sacco</i>                          |
| MAE                | mean absolute error                      |
| MI                 | mesophilic-instable                      |
| MPN                | most probable number                     |
| MS                 | mesophilic-stable                        |
| NGS                | next-generation sequencing               |
| NTC                | non-template control                     |
| OTU                | operational taxonomic unit               |
| PRESS              | predicted residual error sum of squares  |
| qPCR               | quantitative polymerase chain reaction   |
| RMSE               | root-mean-square error                   |
| RT                 | reverse transcription                    |
| TI                 | thermophilic-instable                    |
| TS                 | thermophilic-stable                      |
| txid               | taxonomic identification number          |

## References

1. Naik, S.; Goud, V.V.; Rout, P.K.; Dalai, A.K. Production of first and second generation biofuels: A comprehensive review. *Renew. Sustain. Energy Rev.* **2010**, *14*, 578–597. [[CrossRef](#)]
2. Mood, S.H.; Golfeshan, A.H.; Tabatabaei, M.; Jouzani, G.S.; Najafi, G.; Gholami, M.; Ardjmand, M. Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment. *Renew. Sustain. Energy Rev.* **2013**, *27*, 77–93. [[CrossRef](#)]
3. Zverlov, V.V.; Köck, D.E.; Schwarz, W.H. The role of cellulose-hydrolyzing bacteria in the production of biogas from plant biomass. In *Microorganisms in Biorefineries*; Kamm, B., Ed.; Springer: Berlin/Heidelberg, Germany, 2015; pp. 335–361. ISBN 978-3-662-45208-0.
4. Antoni, D.; Zverlov, V.V.; Schwarz, W.H. Biofuels from microbes. *Appl. Microbiol. Biotechnol.* **2007**, *77*, 23–35. [[CrossRef](#)] [[PubMed](#)]

5. Álvarez, C.; Sosa, F.M.R.; García, B.D. Enzymatic hydrolysis of biomass from wood. *Microb. Biotechnol.* **2016**, *9*, 149–156. [[CrossRef](#)]
6. Berlemont, R.; Martiny, A.C. Genomic potential for polysaccharide deconstruction in bacteria. *Appl. Environ. Microbiol.* **2015**, *81*, 1513–1519. [[CrossRef](#)] [[PubMed](#)]
7. Hassa, J.; Maus, I.; Off, S.; Pühler, A.; Scherer, P.; Klocke, M.; Schlüter, A. Metagenome, metatranscriptome, and metaproteome approaches unraveled compositions and functional relationships of microbial communities residing in biogas plants. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 5045–5063. [[CrossRef](#)]
8. Hugenholtz, P.; Goebel, B.M.; Pace, N.R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **1998**, *180*, 4765–4774. [[CrossRef](#)]
9. Lane, D.J.; Pace, B.; Olsen, G.J.; Stahl, D.A.; Sogin, M.L.; Pace, N.R. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 6955–6959. [[CrossRef](#)]
10. Ward, D.M.; Weller, R.; Bateson, M.M. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* **1990**, *345*, 63–65. [[CrossRef](#)]
11. Johnson, J.S.; Spakowicz, D.J.; Hong, B.-Y.; Petersen, L.M.; Demkowicz, P.; Chen, L.; Leopold, S.R.; Hanson, B.M.; Agresta, H.O.; Gerstein, M.; et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **2019**, *10*, 5029. [[CrossRef](#)]
12. Lebuhn, M.; Hanreich, A.; Klocke, M.; Schlüter, A.; Bauer, C.; Pérez, C.M. Towards molecular biomarkers for biogas production from lignocellulose-rich substrates. *Anaerobe* **2014**, *29*, 10–21. [[CrossRef](#)] [[PubMed](#)]
13. Vos, M.; Quince, C.; Pijl, A.S.; De Hollander, M.; Kowalchuk, G.A. A comparison of rpoB and 16S rRNA as markers in pyrosequencing studies of bacterial diversity. *PLoS ONE* **2012**, *7*, e30600. [[CrossRef](#)] [[PubMed](#)]
14. Pereyra, L.P.; Hiibel, S.R.; Riquelme, M.V.P.; Reardon, K.F.; Pruden, A. Detection and quantification of functional genes of cellulose-degrading, fermentative, and sulfate-reducing bacteria and methanogenic archaea. *Appl. Environ. Microbiol.* **2010**, *76*, 2192–2202. [[CrossRef](#)] [[PubMed](#)]
15. Klappenbach, J.A.; Dunbar, J.M.; Schmidt, T.M. rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* **2000**, *66*, 1328–1333. [[CrossRef](#)]
16. Izquierdo, J.A.; Sizova, M.V.; Lynd, L.R. Diversity of bacteria and glycosyl hydrolase family 48 genes in cellulolytic consortia enriched from thermophilic biocompost. *Appl. Environ. Microbiol.* **2010**, *76*, 3545–3553. [[CrossRef](#)] [[PubMed](#)]
17. E Koeck, D.; Pechtl, A.; Zverlov, V.V.; Schwarz, W.H. Genomics of cellulolytic bacteria. *Curr. Opin. Biotechnol.* **2014**, *29*, 171–183. [[CrossRef](#)]
18. Zverlov, V.V.; Hiegl, W.; Köck, D.E.; Kellermann, J.; Köllmeier, T.; Schwarz, W.H. Hydrolytic bacteria in mesophilic and thermophilic degradation of plant biomass. *Eng. Life Sci.* **2010**, *10*, 528–536. [[CrossRef](#)]
19. Dai, X.; Tian, Y.; Li, J.; Su, X.; Wang, X.; Zhao, S.; Liu, L.; Luo, Y.; Liu, D.; Zheng, H.; et al. Metatranscriptomic analyses of plant cell wall polysaccharide degradation by microorganisms in the cow rumen. *Appl. Environ. Microbiol.* **2015**, *81*, 1375–1386. [[CrossRef](#)]
20. Henrissat, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **1991**, *280 Pt 2*, 309–316. [[CrossRef](#)]
21. Lombard, V.; Ramulu, H.G.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2013**, *42*, D490–D495. [[CrossRef](#)] [[PubMed](#)]
22. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [[CrossRef](#)] [[PubMed](#)]
23. Li, J.; Su, X.; Tian, Y.; Dong, Z.; Hu, S.; Huang, L.; Dai, X. Gene diversity of the bacterial 48 family glycoside hydrolase (GH48) in rumen environment. *Wei Sheng Wu Xue Bao = Acta Microbiol. Sin.* **2014**, *54*, 53–61.
24. Maus, I.; Koeck, D.E.; Cibis, K.G.; Hahnke, S.; Kim, Y.S.; Langer, T.; Kreubel, J.; Erhard, M.; Bremges, A.; Off, S.; et al. Unraveling the microbiome of a thermophilic biogas plant by metagenome and metatranscriptome analysis complemented by characterization of bacterial and archaeal isolates. *Biotechnol. Biofuels* **2016**, *9*, 171. [[CrossRef](#)] [[PubMed](#)]
25. De Vrieze, J. The next frontier of the anaerobic digestion microbiome: From ecology to process control. *Environ. Sci. Ecotechnol.* **2020**, *3*, 100032. [[CrossRef](#)]
26. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)] [[PubMed](#)]

27. Tamura, K.; Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Boil. Evol.* **1993**, *10*, 512–526. [[CrossRef](#)]
28. Rudi, K.; Skulberg, O.M.; Larsen, F.; Jakobsen, K.S. Strain characterization and classification of oxyphotobacteria in clone cultures on the basis of 16S rRNA sequences from the variable regions V6, V7, and V8. *Appl. Environ. Microbiol.* **1997**, *63*, 2593–2599. [[CrossRef](#)]
29. Costa, R.; Götz, M.; Mrotzek, N.; Lottmann, J.; Berg, G.; Smalla, K. Effects of site and plant species on rhizosphere community structure as revealed by molecular analysis of microbial guilds. *FEMS Microbiol. Ecol.* **2006**, *56*, 236–249. [[CrossRef](#)]
30. Mohamed, R.; Chaudhry, A.S. Methods to study degradation of ruminant feeds. *Nutr. Res. Rev.* **2008**, *21*, 68–81. [[CrossRef](#)]
31. Koeck, D.E.; Mechelke, M.; Zverlov, V.V.; Liebl, W.; Schwarz, W.H. *Herbivorax saccincola* gen. nov., sp. nov., a cellulolytic, anaerobic, thermophilic bacterium isolated via *in sacco* enrichments from a lab-scale biogas reactor. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 4458–4463. [[CrossRef](#)]
32. Rettenmaier, R.; Schneider, M.; Munk, B.; Lebuhn, M.; Jünemann, S.; Sczyrba, A.; Maus, I.; Zverlov, V.V.; Liebl, W. Importance of *Defluviitalea raffinosedens* for hydrolytic biomass degradation in co-culture with *Hungateiclostridium thermocellum*. *Microorganisms* **2020**, *8*, 915. [[CrossRef](#)] [[PubMed](#)]
33. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExpASY: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788. [[CrossRef](#)] [[PubMed](#)]
34. Sayers, E.W.; Barrett, T.; Benson, D.A.; Bryant, S.H.; Canese, K.; Chetvernin, V.; Church, D.M.; DiCuccio, M.; Edgar, R.; Federhen, S.; et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **2009**, *37*, D5–D15. [[CrossRef](#)] [[PubMed](#)]
35. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461. [[CrossRef](#)] [[PubMed](#)]
36. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)] [[PubMed](#)]
37. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **2009**, *75*, 7537–7541. [[CrossRef](#)]
38. Edgar, R.C. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics* **2018**, *34*, 2371–2375. [[CrossRef](#)]
39. Nguyen, N.-P.; Warnow, T.; Pop, M.; White, B.A. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *NPJ Biofilms Microbiomes* **2016**, *2*, 16004. [[CrossRef](#)]
40. Yarza, P.; Yilmaz, P.; Pruesse, E.; Glöckner, F.O.; Ludwig, W.; Schleifer, K.-H.; Whitman, W.B.; Euzéby, J.; Amann, R.; Rosselló-Mora, R. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Genet.* **2014**, *12*, 635–645. [[CrossRef](#)]
41. Berry, D.; Ben Mahfoudh, K.; Wagner, M.; Loy, A. Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Appl. Environ. Microbiol.* **2011**, *77*, 7846–7849. [[CrossRef](#)]
42. Reitmeier, S.; Kiessling, S.; Clavel, T.; List, M.; Almeida, E.L.; Ghosh, T.; Neuhaus, K.; Grallert, H.; Troll, M.; Rathmann, W.; et al. Arrhythmic gut microbiome signatures for risk profiling of Type-2 diabetes. *SSRN Electron. J.* **2019**. [[CrossRef](#)]
43. Edgar, R.C. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **2013**, *10*, 996–998. [[CrossRef](#)] [[PubMed](#)]
44. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [[CrossRef](#)] [[PubMed](#)]
45. Lagkouvardos, I.; Fischer, S.; Kumar, N.; Clavel, T. Rhea: A transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons. *PeerJ* **2017**, *5*, e2836. [[CrossRef](#)]
46. Deurenberg, R.H.; Bathoorn, E.; Chlebowicz, M.A.; Couto, N.; Ferdous, M.; García-Cobos, S.; Kooistra-Smid, A.M.; Raangs, E.C.; Rosema, S.; Veloo, A.C.; et al. Application of next generation sequencing in clinical microbiology and infection prevention. *J. Biotechnol.* **2017**, *243*, 16–24. [[CrossRef](#)]
47. Tindall, B.J.; Busse, H.-J.; Ludwig, W.; Rosselló-Mora, R.; Kämpfer, P. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* **2010**, *60*, 249–266. [[CrossRef](#)]

48. Koonin, E.V.; Makarova, K.S.; Aravind, L. Horizontal Gene Transfer in Prokaryotes: Quantification and Classification. *Annu. Rev. Microbiol.* **2001**, *55*, 709–742. [[CrossRef](#)]
49. Sukharnikov, L.O.; Alahuhta, M.; Brunecky, R.; Upadhyay, A.; Himmel, M.E.; Lunin, V.V.; Zhulin, I.B. Sequence, structure, and evolution of cellulases in glycoside hydrolase family 48. *J. Boil. Chem.* **2012**, *287*, 41068–41077. [[CrossRef](#)] [[PubMed](#)]
50. Sundberg, C.; Abu Al-Soud, W.; Larsson, M.; Alm, E.; Yekta, S.S.; Svensson, B.H.; Sørensen, S.J.; Karlsson, A. 454 pyrosequencing analyses of bacterial and archaeal richness in 21 full-scale biogas digesters. *FEMS Microbiol. Ecol.* **2013**, *85*, 612–626. [[CrossRef](#)]
51. Rui, J.; Li, J.; Zhang, S.; Yan, X.; Wang, Y.; Li, X. The core populations and co-occurrence patterns of prokaryotic communities in household biogas digesters. *Biotechnol. Biofuels* **2015**, *8*, 1–15. [[CrossRef](#)]
52. Hua, M.; Yu, S.; Ma, Y.; Chen, S.; Li, F. Genetic diversity detection and gene discovery of novel glycoside hydrolase family 48 from soil environmental genomic DNA. *Ann. Microbiol.* **2018**, *68*, 163–174. [[CrossRef](#)]
53. Maus, I.; Bremges, A.; Stolze, Y.; Hahnke, S.; Cibis, K.G.; Koeck, D.E.; Kim, Y.S.; Kreubel, J.; Hassa, J.; Wibberg, D.; et al. Genomics and prevalence of bacterial and archaeal isolates from biogas-producing microbiomes. *Biotechnol. Biofuels* **2017**, *10*, 264. [[CrossRef](#)] [[PubMed](#)]
54. Koeck, D.E.; Ludwig, W.; Wanner, G.; Zverlov, V.V.; Liebl, W.; Schwarz, W.H. *Herbinix hemicellulosilytica* gen. nov., sp. nov., a thermophilic cellulose-degrading bacterium isolated from a thermophilic biogas reactor. *Int. J. Syst. Evol. Microbiol.* **2015**, *65*, 2365–2371. [[CrossRef](#)] [[PubMed](#)]
55. Koeck, D.E.; Hahnke, S.; Zverlov, V.V. *Herbinix luporum* sp. nov., a thermophilic cellulose-degrading bacterium isolated from a thermophilic biogas reactor. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 4132–4137. [[CrossRef](#)]
56. Mechelke, M.; Koeck, D.; Broeker, J.; Roessler, B.; Krabichler, F.; Schwarz, W.H.; Zverlov, V.V.; Liebl, W. Characterization of the arabinoxylan-degrading machinery of the thermophilic bacterium *Herbinix hemicellulosilytica*—Six new xylanases, three arabinofuranosidases and one xylosidase. *J. Biotechnol.* **2017**, *257*, 122–130. [[CrossRef](#)]
57. Rettenmaier, R.; Gerbault, M.; Liebl, W.; Zverlov, V.V. *Hungateiclostridium mesophilum* sp. nov., a mesophilic, cellulolytic and spore-forming bacterium isolated from a biogas fermenter fed with maize silage. *Int. J. Syst. Evol. Microbiol.* **2019**, *69*, 3567–3573. [[CrossRef](#)]
58. Lebuhn, M.; Effenberger, M.; Gronauer, A.; Wilderer, P.; Wuertz, S. Using quantitative real-time PCR to determine the hygienic status of cattle manure. *Water Sci. Technol.* **2003**, *48*, 97–103. [[CrossRef](#)]
59. Rettenmaier, R.; Duerr, C.; Neuhaus, K.; Liebl, W.; Zverlov, V.V. Comparison of sampling techniques and different media for the enrichment and isolation of cellulolytic organisms from biogas fermenters. *Syst. Appl. Microbiol.* **2019**, *42*, 481–487. [[CrossRef](#)]
60. Orhororo, E.K.; Ebunilo, P.O.; Sadjere, G.E. Effect of organic loading rate (OLR) on biogas yield using a single and three-stages continuous anaerobic digestion reactors. *Int. J. Eng. Res. Afr.* **2018**, *39*, 147–155. [[CrossRef](#)]

