Deep metagenome and metatranscriptome analyses of plant biomass-degrading microbial communities to improve lignocellulose conversion

Dissertation with the aim of achieving a doctoral degree at the Faculty of Mathematics,

Informatics and Natural Sciences

Department of Biology

of Universität Hamburg

Submitted by M.Sc. Simon Güllert

2016 in Hamburg

The following evaluators recommended the admission of the dissertation:

Professor Dr. Wolfgang Streit

Professor Dr. Wolfgang Liebl

Date of oral defense: 23.01.2017

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Declaration on oath

I hereby declare, on oath, that I have written the present dissertation by my own and have not used any other than the acknowledged resources and aids.

Hamburg, 13.10.2016

_____

Simon Güllert

Contributions to the quoted articles

Güllert S., Fischer M. A., Turaev D., Noebauer B., Ilmberger N., Wemheuer B., Alawi M., Rattei T., Daniel R., Schmitz R. A., Grundhoff A. and Streit W. R. (2016). "Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies." Biotechnology for Biofuels 9(1): 1-20. doi: 10.1186/s13068-016-0534-x

- Design of the study and writing of the manuscript
- Extraction of DNA and RNA from biogas samples
- Analysis of assembled metagenome datasets including data evaluation and interpretation
- Comparative analyses between biogas fermenter, elephant feces, and cow rumen samples
- Processing of RNA-Seq data and analysis of transcript levels
- Analysis of the microbial community in the industrial biogas fermenter based on metagenomics marker genes and 16S rRNA gene amplicon sequences

Ilmberger N., Güllert S., Dannenberg J., Rabausch U., Torres J., et al. (2014) "A Comparative Metagenome Survey of the Fecal Microbiota of a Breast- and a Plant-Fed Asian Elephant Reveals an Unexpectedly High Diversity of Glycoside Hydrolase Family Enzymes." PLoS ONE 9(9): e106707. doi: 10.1371/journal.pone.0106707

- Planning of 16S rRNA gene amplicon sequencing and data processing
- Rarefaction analysis of 16S rRNA gene amplicons to assess species richness in the fecal samples
- Analysis of the microbial community in the elephant feces samples based on 16S rRNA gene amplicon sequences

Fischer M. A., Güllert S., Neulinger S. C., Streit W. R. & Schmitz R. A. (2016) "Evaluation of 16S rRNA Gene Primer Pairs for Monitoring Microbial Community Structures Showed High Reproducibility within and Low Comparability between Datasets Generated with Multiple Archaeal and Bacterial Primer Pairs." Frontiers in Microbiology 7. doi:10.3389/fmicb.2016.01297

- Participation in the design of the study
- Providing of a biogas fermenter shotgun metagenome

Table of Contents

1 Abstract

Recent advances in sequencing technologies have given rise to deep metagenome- and metatranscriptome-based analyses of complex microbiomes allowing a previously unprecedented insight into the community structure, the genetic potential, and the metabolic activity of the microorganisms in different habitats. In this study, these techniques were used to investigate the microbial communities associated with a herbivorous hindgut fermentation system (gut of an Asian Elephant), a herbivorous foregut fermentation system (rumen of a Holstein Cow), and a technical biomass fermentation system (agricultural biogas fermenter). The three systems were compared with regard to differences in the composition of cellulolytic organisms, the diversity and abundance of genes encoding lignocellulolytic enzymes, the transcription of these genes, and the overall cellulolytic enzyme activity. The main goal was to achieve a better understanding of the processes involved in the hydrolysis of plant biomass in natural and technical systems.

The results obtained in this study show that a typical industrial biogas reactor fed with maize silage, cow manure, and chicken manure has relatively lower cellulose hydrolysis rates compared to feces samples from various herbivores. In addition, a distinct difference in the microbial community structure with regard to the ratio of the *Firmicutes* versus the *Bacteroidetes* between multiple biogas fermenters and various microbiomes associated with herbivorous animals was revealed. The *Bacteroidetes* were usually 4 to 6 times less prominent than the *Firmicutes* in biogas fermenters compared to natural plant biomass degrading communities. Further, it could be shown that the relatively lower abundance of members of the *Bacteroidetes* was associated with a decreased richness of predicted lignocellulolytic enzymes in the biogas fermenter. On average, 2.5 genes encoding cellulolytic glycoside hydrolases per megabase pair of assembled metagenomic DNA were identified in the biogas fermenter compared to 3.8 in the elephant feces and 3.2 in the cow rumen derived metagenomes. In addition, data obtained from metatranscriptome sequencing indicated that highly transcribed cellulases in the biogas fermenter were four times more often affiliated with the *Firmicutes* compared to the *Bacteroidetes*, while an equal distribution of these enzymes was observed in an elephant feces sample.

Based on the findings in this study, it is likely that the deficiency of lignocellulolytic enzymes affiliated with the *Bacteroidetes* implies a potentially important limitation in the biogas

fermenter with regard to the hydrolysis of biomass. It can further be hypothesized that the importance of the phylum *Bacteroidetes* for the degradation of lignocellulosic biomass was underestimated in the past and that increasing the members of the *Bacteroidetes* in biogas fermenters will most likely result in an improved hydrolytic performance. However, due to the not fully understood better adaption of the *Bacteroidetes* to natural habitats, further research is required to achieve this increase in biogas plants.

2 Introduction

2.1 Lignocellulosic plant biomass and its potential as a renewable energy resource

Due to increasing efforts to limit greenhouse gas emissions, the production of renewable plant-based energy resources like bioethanol and biogas gained increasing attention during the last decade. Today, the main fermentation substrate for the production of bioethanol is starch, which is mainly derived from corn in the US and sugarcane in Brazil (Horn, Vaaje-Kolstad et al. 2012). Large fractions of the agricultural areas in these countries have been diverted for the production of energy crops. Therefore, these areas are no longer available for food production, either for humans or livestock feed. This competition in land use is referred to as the "Food vs. Fuel" debate and the consequences are controversially discussed (Tenenbaum 2008, Thompson 2012). In addition, the large scale conversion in land use has indirect effects which negatively impair the CO_2 record of the energy crops. These effects are described by the indirect land use change (ILUC) factor (Ahlgren and Di Lucia 2016). In general, the same set of problems applies when energy crops are used for the production of biogas in biogas plants. In Europe, and especially Germany, virtually all agricultural biogas plants use energy crops as the main fermentation substrate. Due to its high starch content, maize silage is the most commonly used substrate for biogas production with more than 70% mass input (Herrmann, Idler et al. 2015). A high starch content allows for a more efficient fermentation and results in higher methane yield rates compared to lignocellulose rich substrates such as grass silage.

However, for a sustainable and less controversial use of plant-based energy resources, a conversion to non-food fermentation substrates like the lignocellulosic fraction of plant biomass is crucial. A mayor limitation for this conversion is the recalcitrant nature of lignocellulosic substrates, and as a result, the low fermentation efficiency for the industrial production of bioethanol or biogas (Zheng, Zhao et al. 2014). A prerequisite for the fermentation of plant biomass to either biogas or biofuels is the hydrolysis of the polysaccharide polymers in the plant cell-wall. Basically, this breakdown can be achieved via an enzymatic or chemical treatment and with our without a mechanical or chemical pretreatment (Limayem and Ricke 2012). In biogas plants for example, the initial

degradation of the biomass is carried out by microorganisms which produce hydrolytic enzymes. Yet, this enzymatic breakdown of lignocellulosic biomass is slow and challenging due to its complex cross-linked structure and requires the interaction of various hydrolytic enzymes. Lignocellulose is mainly made up of cellulose, hemicellulose, and lignin in varying ratios (Figure 1).

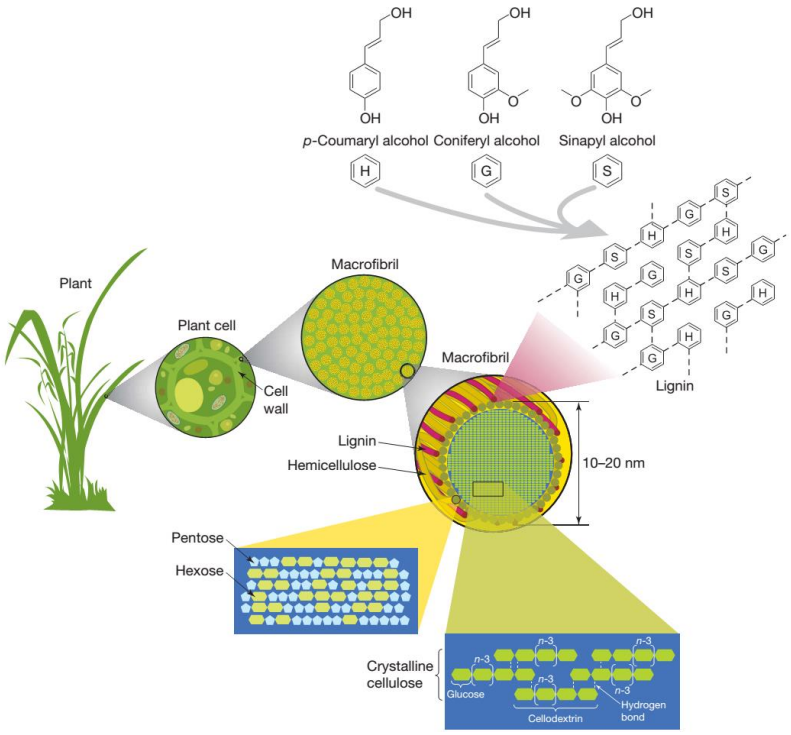

Figure 1. Structure of lignocellulosic biomass. Lignocellulose consists of three main biopolymers: cellulose, hemicellulose, and lignin in varying ratios depending on the type of plant (figure reprinted from Rubin 2008).

2.2 Degradation of lignocellulosic plant biomass

The key element of the plant cell wall is cellulose. Cellulose is made up of D-glucose molecules which are connected to chains via beta 1,4-glycosidic bonds. Multiple cellulose chains are linked to microfibrils by hydrogen bonds. To degrade this crystalline fibril structure, at least three different classes of enzymes are needed: exoglucanases (EC 3.2.1.91), endoglucanases (EC 3.2.1.4), and β-glucosidases (EC 3.2.1.21) (Lynd, Weimer et al. 2002). The cellulose microfibrils are also interspersed with hemicellulose and lignin components (Rubin 2008). Hemicelluloses are a heterogeneous mixture of different hexoses (e.g. D-mannose and D-galactose) and pentoses (e.g. D-xylose and L-arabinose) which form

an amorphous structure. One of the most abundant types of hemicellulose is xylan, also called arabinoxylan. The backbone of arabinoxylan is made up of beta 1,4-linked D-xylose units. In addition, several of these units can be substituted with L-arabinose molecules (Scheller and Ulvskov 2010). The amount of hemicellulose in the plant cell wall, as well as its composition and branching patterns, differs widely between different plant species. Therefore, a variety of different enzymes is involved in the degradation of hemicelluloses including: Acetyl-xylan-esterases (EC 3.1.1.72), α-l-Arabinofuranosidases (EC 3.2.1.55) Endo-xylanases (EC 3.2.1.8), α-Glucuronidases (EC 3.2.1.139), and others (Shallom and Shoham 2003).

The lignin fraction of the plant cell wall is composed of three phenylpropanoid monolignols: p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol. These lignin building units are polymerized by oxidative enzymes in a free-radical reaction to a rigid aromatic polymer. The lignin polymer primarily provides stiffness and mechanicals strength to the plant cell wall. Lignin also increases the resistance of the cell wall to degradation by blocking the access to cellulose and hemicellulose for hydrolytic enzymes. Depending on the plant, the lignin content of the cell wall varies. The highest amounts of lignin are found in softwoods and hardwoods. Grasses and other gramineous plants usually contain less lignin (Limayem and Ricke 2012). Lignin can be depolymerized by bacteria and fungi via peroxidases and laccases (Brown and Chang 2014).

Another important factor which contributes to the high recalcitrance of lignocellulose is the cross-linking of the different fractions. Ferulic acid, for example, cross-links the arabinoxylan chains of the hemicellulose fraction via ester bonds and also cross-links the hemicellulose to lignin via ether bonds (Wong, Chan et al. 2011). Ferulic acid can account for up to 3% of the dry weight of gramineous plant cell walls (Wong 2006). Feruloyl esterases can cleave the ferulic acid cross-links and by this improve the biodegradation of the lignocellulose into fermentable sugars. Therefore, feruloyl esterases are important accessory enzymes for the degradation of plant cell walls because of their synergistic effect with lignocellulolytic enzymes (Braga, Delabona Pda et al. 2014).

In summary, the degradation of plant biomass is a slow and complex process. A better understanding of the bacterial organisms which are able to degrade lignocellulose and how they accomplish this task is crucial to improve the fermentation efficiency for the production of biogas or bioethanol and thus, for the widespread use of lignocellulose biomass as a

renewable energy resource. The most important known cellulolytic bacterial groups and their enzyme systems will be introduced in the next chapter.

2.2.1 Degradation of cellulose by bacteria

Heterotrophic bacteria utilize organic molecules as energy and carbon sources. Some of these heterotrophs have developed systems to degrade the most abundant organic polymer on earth, cellulose. The polysaccharide degrading systems of these organisms range from simple free enzyme systems to sophisticated multi enzyme complexes called cellulosomes. In addition, the importance of certain bacterial groups for the degradation of cellulose, like e.g. the *Bacteroidetes*, is only gradually becoming apparent (Naas, Mackenzie et al. 2014).

The simplest but arguably also most inefficient way for bacteria to degrade cellulose, is to secrete cellulases into their environment. Here, these enzymes degrade the cellulose polymer and the breakdown products are afterwards transported into the cell where they are oxidized to generate energy. Various aerobic organisms rely on this strategy to degrade cellulose, for example, bacteria belonging to the genus *Cellvibrio* of the phylum *Proteobacteria* or to the genus *Cellulomonas* of the phylum *Actinobacteria* (Lynd, Weimer et al. 2002). The secreted cellulases found in these organisms often exhibit a modular structure including a catalytic domain and one or multiple carbohydrate-binding modules (CBMs). These CBMs typically possess a high binding affinity for crystalline or amorphous cellulose while the catalytic domains most often encompass exoglucanase (EC 3.2.1.91) and endoglucanase (EC 3.2.1.4) activity (Christopherson, Suen et al. 2013).

As described in the previous chapter, cellulose is usually embedded in a hemicellulose-lignin complex in the plant cell wall. To degrade this complex, bacteria have to secrete multiple lignocellulolytic enzymes. However, in the surrounding of the cell, the binding of the secreted enzymes to their individual substrates occurs randomly and therefore not in an optimal ratio and order. This is why especially anaerobic bacteria have developed sophisticated and more efficient systems to degrade plant cell wall polysaccharides. These systems are named cellulosomes. Until now, cellulosomes have only been identified in anaerobic bacteria which belong to the class *Clostridia* of the phylum *Firmicutes.* It is assumed that an anaerobic environment inflicts a, currently not fully understood, evolutionary pressure on the bacteria which has led to the development of these more efficient systems. It is likely, that energetic constrains of the anaerobic metabolism were

important factors contributing to the formation of cellulosomes (Fontes and Gilbert 2010). Cellulosomes are multi-enzyme complexes which are attached to the outer cell membrane via an anchoring protein and can differ in their complexity and makeup. The currently best studied cellulosome is found in the thermophilic anaerobic bacterium *Clostridium thermocellum*. This bacterial species has been thoroughly investigated due to its potential application as a bioethanol production strain (Demain, Newcomb et al. 2005). The *C. thermocellum* cellulosome is attached to the cell membrane via three S-layer homology (SLH) modules in the anchoring protein (Figure 2A).

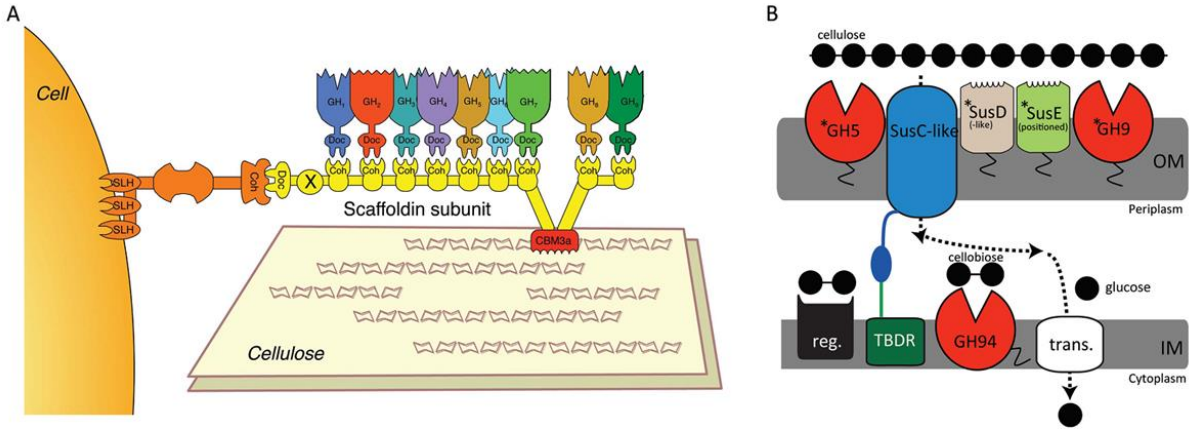

Figure 2. A) Model of the clostridial *C. thermocellum* cellulosome (figure reprinted from Yaniv, Fichman et al. 2014). **B)** Proposed model of a cellulolytic bacteroidetal PUL system (figure reprinted from Naas, Mackenzie et al. 2014).

The largest subunit of the cellulosome, the scaffolding protein, is attached to the cell membrane via a dockerin module which interacts with a cohesion module in the anchoring protein. This cohesin-dockerin affinity interaction is one of the strongest protein-protein interactions known in nature and a central feature required for the assembly of the cellulosome (Mechaly, Fierobe et al. 2001).

The scaffolding protein contains additional cohesion modules which interact with dockerin modules present in the actual enzymatic subunits. In most cases, these subunits are different glycoside hydrolases (GHs) i.e. enzymes that hydrolyze glycosidic bonds. GHs can be classified by EC numbers based on the reaction they catalyze, or they can be grouped in GH families based on sequence similarity which allows a more differentiated distinction. A comprehensive repository of GH families is the Carbohydrate Active enZyme (CAZy) database (Lombard, Ramulu et al. 2014). In this database, currently more than 130 GH families are described encompassing a variety of different carbohydrate modifying activities

and substrate specificities. Several of the CAZy GH families comprise hydrolytic activities which are required for the depolymerization of plant derived biomass. In the *C. thermocellum* cellulosome, subunits with cellobiohydrolase (GH9), chitinase (GH 18), endoglucanase (GH5, GH8, GH9, and GH44), exoglucanase (GH9, GH48), xylanase (GH10, GH11), and other activities were identified. In addition to GH subunits, carbohydrate esterases (CEs) with specificity to acetyl xylans (CE4) and feruloyl esters (CE1) are present in the *C. thermocellum* cellulosome (Blum, Kataeva et al. 2000, Doi and Kosugi 2004). The assembly of all of these subunits in close proximity, the optimal ratio and right order, results in a considerable synergistic effect for the degradation of plant biomass (Gefen, Anbar et al. 2012). Another important factor contributing to the high processivity of the cellulosomal enzymes is the presence of CBMs in the complex. In case of the *C. thermocellum* cellulosome, a cellulose specific CBM is integrated in the scaffolding protein. In general, cellulosomal systems are highly diverse in their makeup and can contain additional CBMs as well as multiple scaffoldin subunits.

While the *Clostridia* and their cellulosomes are subject to intensive studies for already more than thirty years, recently the involvement of another bacterial group in the degradation of plant biomass has attracted increasing attention. Bacteria of the phylum *Bacteroidetes* are frequently found in major amounts in digestive organs of herbivorous animals. Next to the *Firmicutes*, they have been identified as the main bacterial group in various gut, rumen, and fecal samples (Pope, Mackenzie et al. 2012, He, Ivanova et al. 2013, Henderson, Cox et al. 2013, Ilmberger, Güllert et al. 2014, Roggenbuck, Sauer et al. 2014, Zeng, Han et al. 2015). However, most of these *Bacteroidetes* are currently uncultured in the lab and therefore poorly characterized. The difficulties in culturing many bacteroidetal organisms significantly limit the understanding of the contribution of this phylum to the degradation of lignocellulose. Opposing to some *Clostridia*, bacteria of the phylum *Bacteroidetes* do not produce cellulosomes. Still, evidence is mounting that they might utilize an alternative system for the degradation of cellulose (Naas, Mackenzie et al. 2014). In the genomes of various *Bacteroidetes* species, specific Polysaccharide Utilization Loci (PULs) have been identified. These PULs probably provide an advantage for the uptake of polysaccharides in densely populated and competitive habitats (Thomas, Hehemann et al. 2011).

A PUL can be described as a set of genes coding for polysaccharide-degrading enzymes in proximity to a *susC*-like and s*usD*-like gene pair. Potential PULs have been predicted in the

genomes of more than 70 *Bacteroidetes* species and are listed in the Polysaccharide-Utilization Loci DataBase (PULDB) (Terrapon, Lombard et al. 2015). While PULs were initially described as systems for the breakdown of starch, the range of PUL target substrates has been continuously extended during the last years and now includes mucin glycans, pectin, alginate, and various hemicelluloses (Dodd, Moon et al. 2010, Sonnenburg, Zheng et al. 2010, Martens, Lowe et al. 2011, Hehemann, Kelly et al. 2012, Kabisch, Otto et al. 2014, Larsbrink, Rogers et al. 2014, Despres, Forano et al. 2016).

Recently, it was shown that the outer membrane associated SusD protein is able to bind to crystalline cellulose and also interacts with plant cell walls (Mackenzie, Pope et al. 2012). In addition, various PULs which encompass cellulolytic GH family enzymes were identified in different metagenomic datasets (Wang, Hatem et al. 2013, Ilmberger, Güllert et al. 2014, Naas, Mackenzie et al. 2014, Güllert, Fischer et al. 2016). These findings strongly suggest that bacteroidetal PULs might also be involved in lignocellulose degradation. A hypothetical model of a cellulolytic PUL system was proposed by Naas and colleagues (Naas, Mackenzie et al. 2014) based on a binned draft genome sequence of an uncultured *Bacteroidetes* species which was obtained from a cow rumen metagenome (Figure 2B). In this model, the SusD and SusE lipoproteins are embedded in the outer membrane of the cell and bind to the cellulose substrate. Membrane associated endo- and exoglucanases (GH families 5 and 9) hydrolyze the cellulose to cellobiose disaccharides. The cellobiose is subsequently transported to the periplasm via a SusC-like (TonB-dependent) transporter in the outer membrane. In the periplasm, an inner membrane-bound cellobiose phosphorylase (GH family 94) degrades the cellobiose to monomeric sugars which are then transported to the cytoplasm. Furthermore, a regulatory system which controls the gene transcription of a specific PUL in response to the presence of the “target” polysaccharide is proposed in the model. A transcriptional regulation of PUL genes has already been shown for two *Bacteroides* species (Martens, Lowe et al. 2011).

In conclusion, several bacteria have developed effective systems and associated enzymes for the degradation of cellulosic biomass. While certain cellulolytic bacterial groups like the *Clostridia* have been intensively studied during the last years, the contribution of the *Bacteroidetes* to the degradation of plant biomass remains mostly elusive. In this context, recent findings of high abundances of mostly uncultured *Bacteroidetes* in many natural lignocellulosic biomass-degrading systems have raised the interest in this phylum.

Consequently, further research is needed to elucidate the relevance and potential of these bacteria for the hydrolysis of plant biomass.

2.3 Anaerobic digestion in agricultural biogas plants

Agricultural biogas plants are technical systems designed to convert plant biomass to biogas in a microbial process called anaerobic digestion (Figure 3A and 3B). This process is usually divided into four steps: hydrolysis, acidogenesis, acetogenesis, and methanogenesis (Weiland 2010).

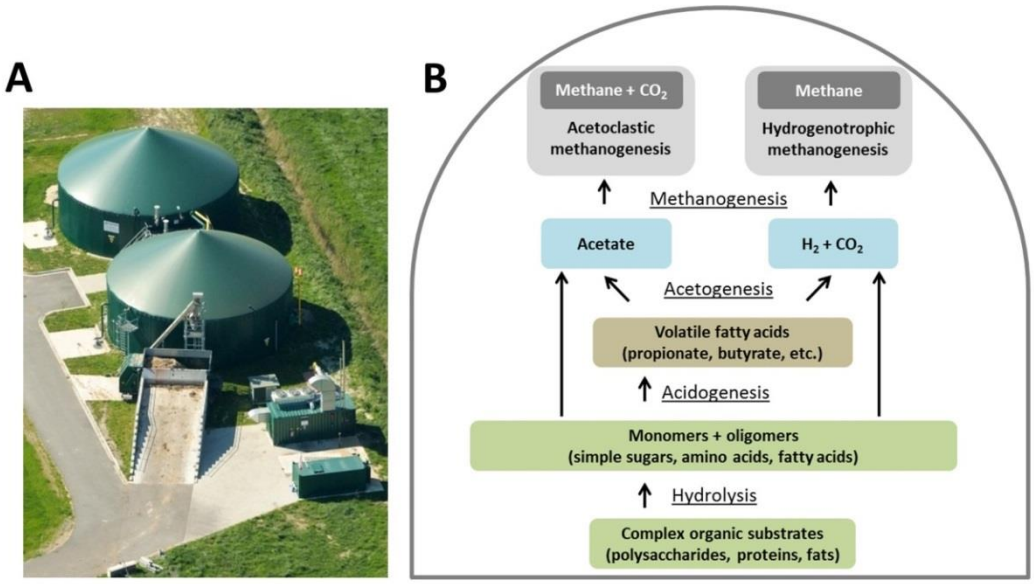

Figure 3. A) Image of a typical agricultural biogas plant ©M. Nolte, CC-by-sa-3.0 DE. **B)** Overview of the anaerobic digestion process in a biogas fermenter. The process can be divided into four steps: hydrolysis, acidogenesis, acetogenesis, and methanogenesis.

In the first hydrolysis step, complex organic substrates are broken down to monomers and oligomers. These simple sugars, amino acids, and fatty acids are then metabolized by fermentative bacteria to mainly volatile (short chain) fatty acids. During this acidogenesis, also acetate, hydrogen, and carbon dioxide are produced. In the acetogenesis step, the fatty acids are oxidized to acetate, hydrogen, and carbon dioxide by acetogenic bacteria via β-oxidation. In the last step, methanogenic archaea produce methane by either acetoclastic methanogenesis or hydrogenotrophic methanogenesis. Today, it is assumed that the methanogenesis in agricultural biogas plants is mostly carried out by hydrogenotrophic archaea (Nettmann, Bergmann et al. 2010, Solli, Havelsrud et al. 2014).

Within this framework, it is important to state that these four steps are interconnected and occur simultaneously in the fermenter. In active biogas fermenters, between several hundred to several thousand microbial species have been identified (Zakrzewski, Goesmann et al. 2012, Li, Chu et al. 2013, Solli, Havelsrud et al. 2014, Stolze, Zakrzewski et al. 2015, Güllert, Fischer et al. 2016). The entirety of the complex metabolic interactions, as well as many of the involved microorganisms are not well understood (Wirth, Kovacs et al. 2012, Zakrzewski, Goesmann et al. 2012). With respect to the initial hydrolysis, it is known that bacteria of the class *Clostridia* are the dominant biomass degrading organisms in agricultural biogas fermenters. The breakdown of the organic substrates is a prerequisite for all subsequent steps. In fact, it is assumed that the initial hydrolysis of the plant biomass is a rate-limiting factor for the overall process (Zverlov, Köck et al. 2015). Therefore, the microbial hydrolysis represents a suitable target for the optimization of biogas plants, as an increase in the hydrolytic performance likely results in higher methane yields. A better understanding of the microorganisms which are essential for an efficient hydrolysis of plant biomass in natural cellulolytic systems might reveal approaches to improve a technical system like a biogas plant.

2.4 Cultivation independent analyses of microbial communities

Recent advances in sequencing technologies have given rise to deep metagenome- and metatranscriptome-based analyses of complex microbiomes. These cultivation independent techniques allow a previously unprecedented insight into the community structure, the genetic potential, and the metabolic activity of the microorganisms in a specific habitat. With regard to the degradation of plant biomass, the microbial groups which are most relevant for the degradation of lignocellulose can be identified and investigated in different cellulolytic systems. By this, a better understanding of the complex factors which are crucial for an efficient enzymatic hydrolysis of lignocellulosic biomass can be gained and, in the long term, be used to improve the microbial conversion of plant biomass to renewable energy resources.

Furthermore, these methods can be applied to screen for novel biocatalysts involved in the depolymerization of biomass. Today, it is assumed that numerous highly efficient lignocellulolytic enzymes are still concealed in hard to cultivate microorganisms (Rubin 2008). A metagenomic screening for novel hydrolytic enzymes can be performed based on

either function or sequence. In case of a function-based screening, environmental DNA is used to construct a metagenomic library. For this purpose, the metagenomic DNA is ligated into fosmid vectors and the fosmids are then introduced into a bacterial host, usually *E.coli*, for heterologous expression. Enzymatic activity of a fosmid carrying clone can be detected by using an appropriate screening substrate. As an alternative strategy, new biocatalysts can be identified based on sequence similarity to known genes and enzymes in databases (Streit and Schmitz 2004, Simon and Daniel 2011). The sequence-based approach has become increasingly popular with the advent of high-throughput next generation sequencing and is especially powerful when combined with an additional metatranscriptome sequencing (RNA-Seq).

2.5 Aims of this study

The main goal of this study was to obtain a better understanding of the processes involved in the hydrolysis of plant biomass in natural and technical systems. For this, deep metagenome- and metatranscriptome-based analyses were used to investigate the microbial communities associated with a herbivorous hindgut fermentation system (gut of an Asian Elephant), a herbivorous foregut fermentation system (rumen of a Holstein Cow), and a technical biomass fermentation system (agricultural biogas fermenter) (Figure 4).

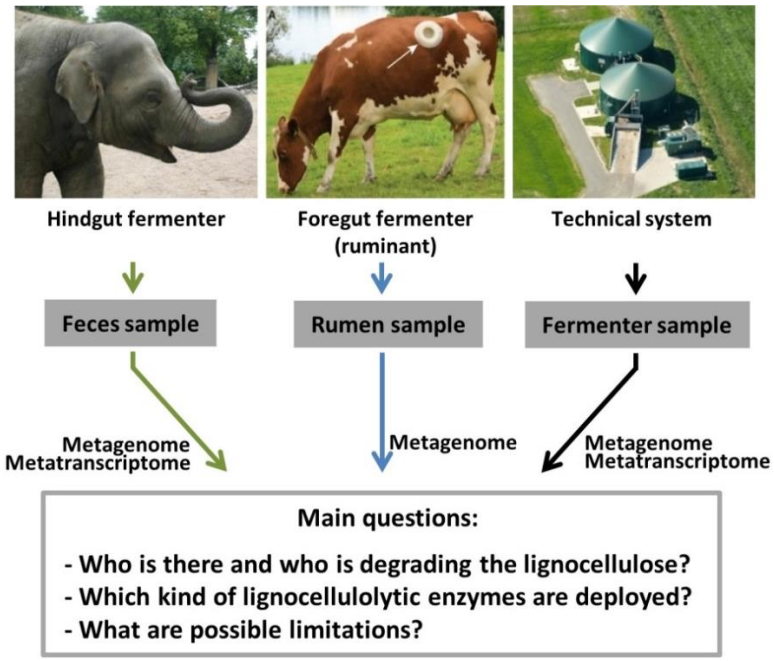

Figure 4. Overview of the three systems analyzed in this study with respect to their potential to degrade lignocellulosic biomass (image copyrights, left to right: Hagenbeck Zoo Hamburg; Hess et al. 2011, Science 331:463-467; M. Nolte, Creative Commons CC-by-sa-3.0 DE).

The three systems were compared with regard to differences in the composition of cellulolytic organisms, the diversity and abundance of genes encoding lignocellulolytic enzymes, the transcription of these genes, and the overall cellulolytic enzyme activity.
Therefore, feces samples were collected from an adult zoo elephant and used to study the microbiome in the hindgut of the largest land-living herbivores which feed on a wide variety of plant-derived biomass. In addition, a published comprehensive metagenomic dataset obtained from a microbiome adherent to switch grass, which was incubated in the rumen of a cow for 72 hours, was included. In contrast to elephants, cows are foregut fermenters which digest their diet in the rumen and mainly feed on grass. Finally, a sample from a typical agricultural biogas plant, which was mainly operated with maize silage, was obtained and the microbial community in this technical biomass converting system was examined.

The comprehensive omics-based analyses of the plant-biomass degradation in the three sampled systems directly resulted in the following publications:

1. **Güllert, Fischer et al. 2016, Biotechnology for Biofuels 9(1):1-20. doi: 10.1186/s13068-016-0534-x**
   Using comparative analyses of the microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces, major differences in carbohydrate hydrolysis strategies were revealed.
2. **Ilmberger, Güllert et al. 2014, PLoS ONE 9(9). doi: 10.1371/journal.pone.0106707**
   The focus of this study was a comparative in-depth analysis of the fecal microbiota of a breast- and a plant-fed Asian elephant. This analysis discovered an unexpectedly high diversity of glycoside hydrolase family enzymes.
3. **Fischer, Güllert et al. 2016, Frontiers in Microbiology 7. doi: 10.3389/fmicb.2016.01297**
   This study evaluated different methods for the monitoring of microbial community structures and showed that there was a high reproducibility within and a low comparability between datasets generated with multiple archaeal and bacterial 16S rRNA gene primer pairs.
4. **Unpublished (manuscript in preparation)**
   Metagenomic DNA isolated from an agricultural biogas fermenter sample was used for a function based screening to identify potentially novel feruloyl esterases.

3 Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies

Simon Güllert[1], Martin A. Fischer[2], Dmitrij Turaev[3], Britta Nöbauer[3], Nele Ilmberger[1], Bernd Wemheuer[4], Malik Alawi[5, 6], Thomas Rattei[3], Rolf Daniel[4], Ruth A. Schmitz[2], Adam Grundhoff[6] and Wolfgang R. Streit[1]

[1]Biocenter Klein Flottbek, Department of Microbiology and Biotechnology, University of Hamburg, Hamburg, Germany; [2]Institute for General Microbiology, Christian Albrecht University Kiel, Kiel, Germany; [3]CUBE - Division for Computational Systems Biology, Dept. of Microbiology and Ecosystem Science, University of Vienna, Vienna, Austria; [4]Institute of Microbiology and Genetics, Georg-August-University Göttingen, Göttingen, Germany; [5]University Medical Center Hamburg-Eppendorf, Bioinformatics Core, Hamburg, Germany; [6]Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Hamburg, Germany

Published in:

Biotechnology for Biofuels 9 (1): 1-20. doi: 10.1186/s13068-016-0534-x

Güllert *et al. Biotechnol Biofuels (2016) 9:121*
DOI 10.1186/s13068-016-0534-x

Biotechnology for Biofuels

RESEARCH Open Access

Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies

Simon Güllert[1], Martin A. Fischer[2], Dmitrij Turaev[3], Britta Noebauer[3], Nele Ilmberger[1], Bernd Wemheuer[4], Malik Alawi[5,6], Thomas Rattei[3], Rolf Daniel[4], Ruth A. Schmitz[2], Adam Grundhoff[6] and Wolfgang R. Streit[1*]

Abstract

Background: The diverse microbial communities in agricultural biogas fermenters are assumed to be well adapted for the anaerobic transformation of plant biomass to methane. Compared to natural systems, biogas reactors are limited in their hydrolytic potential. The reasons for this are not understood.

Results: In this paper, we show that a typical industrial biogas reactor fed with maize silage, cow manure, and chicken manure has relatively lower hydrolysis rates compared to feces samples from herbivores. We provide evidence that on average, 2.5 genes encoding cellulolytic GHs/Mbp were identified in the biogas fermenter compared to 3.8 in the elephant feces and 3.2 in the cow rumen data sets. The ratio of genes coding for cellulolytic GH enzymes affiliated with the *Firmicutes* versus the *Bacteroidetes* was 2.8:1 in the biogas fermenter compared to 1:1 in the elephant feces and 1.4:1 in the cow rumen sample. Furthermore, RNA-Seq data indicated that highly transcribed cellulases in the biogas fermenter were four times more often affiliated with the *Firmicutes* compared to the *Bacteroidetes*, while an equal distribution of these enzymes was observed in the elephant feces sample.

Conclusions: Our data indicate that a relatively lower abundance of bacteria affiliated with the phylum of *Bacteroidetes* and, to some extent, *Fibrobacteres* is associated with a decreased richness of predicted lignocellulolytic enzymes in biogas fermenters. This difference can be attributed to a partial lack of genes coding for cellulolytic GH enzymes derived from bacteria which are affiliated with the *Fibrobacteres* and, especially, the *Bacteroidetes*. The partial deficiency of these genes implies a potentially important limitation in the biogas fermenter with regard to the initial hydrolysis of biomass. Based on these findings, we speculate that increasing the members of *Bacteroidetes* and *Fibrobacteres* in biogas fermenters will most likely result in an increased hydrolytic performance.

Keywords: Anaerobic digestion, Biogas, Biofuels, Biorefinery, Lignocellulosic biomass, Metagenomics, Cellulases, Microbial communities, PULs

*Correspondence: wolfgang.streit@uni-hamburg.de
[1] Department of Microbiology and Biotechnology, Biocenter Klein Flottbek, University of Hamburg, Ohnhorststr. 18, 22609 Hamburg, Germany
Full list of author information is available at the end of the article

Background

In the context of climate change, the production of biogas as a renewable energy form has become increasingly attractive over the last two decades. Biogas is composed of mainly methane and carbon dioxide, which are produced in a complex and anaerobic microbial process [1]. While the main microorganisms and mechanisms involved in the methane production are well known, the overall process of the microbial biogas producing communities, starting with the anaerobic digestion of the biomass to the final end product, is not well understood [2–5]. Previously, it has been reported that the anaerobic degradation of the plant biomass and the subsequent generation of biogas require the close interaction of many different and phylogenetically diverse microorganisms. Published research implies that the diversity ranges from several hundred to several thousand microbial species in active biogas reactors [3, 4, 6, 7]. Interestingly, it was further reported that the overall production of biogas is probably limited due to the relatively slow hydrolysis of the agricultural plant biomass [8]. Thereby, *Clostridia* appear to play a major role during the initial biomass degradation. In fact, they are the dominant class of hydrolytic organisms in the biogas fermenters [8]. Numerous cellulolytic *Clostridia* produce cellulosomes. Cellulosomes are large multi exoenzyme complexes, whose purpose is the efficient degradation of cellulose [9, 10]. These membrane-associated complexes can be visualized using electron microscopy [11]. While the clostridial systems are, perhaps, the most competitive group within biogas fermenters, they are less dominant in natural digestive organs, such as the cow rumen or the gut of other studied herbivorous animals. Within this context, recent research has uncovered that the *Bacteroidetes* are present in virtually all rumen, gut, and fecal samples of herbivores. Here, they usually represent the predominant bacterial group, besides the *Firmicutes* [12–19]. In contrast to clostridial organisms, bacteria of the phylum *Bacteroidetes* do not produce cellulosomes. However, they are associated with the production of very versatile polysaccharide utilization loci (PULs). PULs are prevalent in the phylum of *Bacteroidetes* and have only recently attracted increasing attention. Evidence is mounting that PULs might play an important part in the breakdown of cellulose [20]. Furthermore, it was recently proposed that cellulolytic PULs might be considered as an alternative system for the degradation of cellulose, next to cellulosomes and free-enzymes [21]. PULs, which were originally described as starch degradation operons, have been predicted in up to 67 *Bacteroidetes* genomes until now. They can be described as a set of genes organized around an *SusC* and *SusD* gene pair [22]. Intrigued by the differences between the composition of the microbiomes of natural cellulolytic systems and biogas fermenters, we wanted to investigate how these differences may affect the ability to effectively degrade biomass in biogas plants. In this study, we employed deep metagenome sequencing in combination with RNA-Seq to obtain detailed insights into the glycoside hydrolase enzymes (GHs), mainly employed in carbohydrate hydrolysis in different cellulolytic systems. Within this paper, we provide evidence that in published natural cellulolytic systems of herbivorous animals the ratio of the *Firmicutes* vs. *Bacteroidetes* is almost 1:1 [14–18, 23, 24]. In contrast, in a technical system, such as biogas fermenters the *Firmicutes* outcompete the *Bacteroidetes* by four-to-six-fold [2–4, 25]. In line with this observation, we show that the overall abundance of potential glycoside hydrolase genes is lower in the biogas fermenter compared to two natural systems due to an underrepresentation of typical rumen and gut bacteria. Furthermore, we wanted to know, if these differences are associated with the predominant transcription of certain GH families, possibly allowing a more efficient degradation of the plant biomass.

Methods

Total DNA extraction from an agricultural biogas fermenter sample

Samples were taken from the fermenter of an agricultural biogas plant located near Cologne (Germany) in March and May 2013. At the time of sampling, the biogas plant was running under steady conditions. It produced 536 kW output and was fed with maize silage (69 %), cow manure (19 %), and chicken manure (12 %). Fermentation took place at 40 °C and a pH value of 8 in a 2800 m^3 fermenter. Total DNA was isolated (Isolation 1) using the QIAamp DNA Stool kit from Qiagen (Hilden, Germany) according to the manufacturer's protocol for pathogen detection. For this isolation, 2 g of fermenter material were used and the reaction steps were scaled up accordingly. Heating of the suspension was carried out at 95 °C.

For metagenome sequencing, an additional DNA isolation (Isolation 2) from the May sample was conducted using a CTAB-based method according to Weiland–Bräuer [26]. 1.5 g of sample material was mechanically disrupted using a Dismembrator U instrument (Sartorius AG, Göttingen, Germany). Subsequently, 2.7 ml DNA extraction buffer with 5 % CTAB was added to 1 g of homogenized material. Extracted DNA was highly contaminated by humic acids indicated by brownish to yellow color. Contamination was removed using the FastDNA™ SPIN Kit for Soil (MP Biomedicals, Solon, Ohio, US), excluding the initial lysis steps. Purity of DNA was analyzed using a Nanodrop ND-2000 instrument (PEQLAB Biotechnologie GmbH, Erlangen, Germany).

Amplification and sequencing of 16S rRNA genes

Variable regions of bacterial 16SrRNA genes were amplified as previously published [15] with minor modifications. The V3–V5 region was amplified using the primer set: V3 for 5′-CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCTCGACACCTACGGGNGGCWGCAG-3′ and V5rev 5′-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGCCGTCAATTCMTTTRAGTTT-3′. The primers contained Roche 454 pyrosequencing adaptors, keys, and one unique MID per sample (underlined). To target archaeal 16SrRNA genes, the V4–V6 region was amplified using the primer set: A519F 5′-CCATCTCATCCCTGCGTGTCTCCGACTCAGATATCGCGAGCAGCMGCCGCGGTAA-3′ and A1041R 5′-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGGGCCATGCACCWCCTCTC-3′. The PCR reaction (50 µl) contained 0.5 U of Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Braunschweig, Germany), 10 µl 5× Phusion GC Buffer, 200 µM of each dNTP, 2.5 % DMSO, 1.5 mM MgCl2, 4 µM of each primer, and 20 ng isolated DNA. PCR cycling conditions were: initial denaturation at 98 °C for 3 min, followed by 28 cycles of denaturation at 98 °C for 30 s, annealing at 61 °C for 30 s (archaeal primer set: 66 °C), and extension at 72 °C for 25 s. The final extension was conducted at 72 °C for 5 min. Negative controls were performed with H_2O instead of template DNA. The obtained PCR products were purified via Gel/PCR DNA Fragments Extraction Kit (Geneaid Biotech, Taiwan) as recommended by the manufacturer. Three separate PCR reactions were conducted for each sample. After gel extraction, the reaction products were pooled in equal amounts. The 16S rRNA gene sequencing was performed at the Göttingen Genomics Laboratory using a Roche GS FLX++ 454 pyrosequencer with titanium chemistry (Roche, Branford, USA).

Processing and analysis of 16S rRNA genes

Pyrosequencing derived raw sequences were processed according to Wemheuer et al. [27], with the following modifications: After raw data extraction, reads shorter than 300 bp and those possessing long homopolymer stretches (≥8 bp) or primer mismatches (>3 bp) were removed. The sequences were denoised employing Acacia version 1.53b [28]. Chimeric sequences were, subsequently, removed using UCHIME in de novo and in reference mode using the SILVA SSU database (SSURef 119 NR) as reference data set [29, 30]. All non-bacterial as well as singletons OTUs (OTUs containing only one sequence) were removed according to Schneider et al. [31]. The remaining 16S rRNA gene sequences were uploaded to the SILVA NGS (SILVA Next-Generation Sequencing) server for taxonomic classification [29]. Microbial taxonomy was determined using SILVA version 119 and default settings with one adjustment: The cluster sequence identity threshold was increased to 0.99. Rarefaction curves, diversity indices, and shared OTUs were calculated employing the QIIME 1.8 software package [32].

Metagenome sequencing and de novo assembly

20 ng DNA was sheared with the Bioruptor® (Diagenode; 7 times for 15 s on/90 s off) and libraries were generated using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® as recommended by the manufacturer. Size and quality of the libraries were assessed using a BioAnalyzer High Sensitivity Chip. Diluted libraries (2 nM) were multiplex-sequenced on the Illumina HiSeq 2500 instrument. Initially, one lane was sequenced for each DNA isolation in paired end mode (2 × 101 bases). Subsequently, a second lane was sequenced for DNA isolation 2 using the same conditions. The number of generated reads is indicated in Table 1. Sequencing was carried out at the Heinrich-Pette-Institut in Hamburg, Germany. Sequencing adapters were removed using Trimmomatic 0.33 [33]. Different de novo assemblies were performed using either the IDBA-UD 1.1.1 [34] or Ray Meta v.2.3.1 [35] assembler (Table 1).

Binning of metagenomic contigs

For metagenomic binning, the assembly was performed using the Ray Meta assembler with a k-mer length of 31. Contig coverage was determined by mapping the initial reads to the contigs using the short-read mapper BBMap (BBMap, Bushnell B., sourceforge.net/projects/bbmap/). Subsequently, Samtools [36] was used to convert, sort, and merge the sam files. After that, BEDTools [37] were used to calculate contig-wise average coverage. As low-coverage and short contigs are known to be error-prone [38], contigs with a length <1 kb and average coverage <3 were discarded from the assembly.

Taxonomic profiling of reads was performed by a sequence similarity search using blastx (NCBI-BLAST 2.2.26, e-value <0.1) [39] against a database of universally conserved proteins which occur in 98 % of all eukaryotes, bacteria, and archaea. The database was clustered to a sequence similarity level of 97 % to remove redundancy. Blast results were taxonomically assigned by MEGAN [40] with min. bitscore 60 and min. support percent 0.05, and visualized by Krona 2.5 [41].

Taxonomic profiling of contigs was performed using AMPHORA2 [42] using the universal marker set of 31 genes. NCBI taxonomy ids were mapped to phylogenetic lineages given by AMPHORA2. The comparison between read-based and assembly-based communities was made to verify the consistency between the sample and the assembly. The visualization software Elviz [43]

Table 1 Number of generated reads and statistics of different de novo assemblies

Biogas fermenter May 2013 sample	#Reads used for assembly	Ray assembly					IDBA-UD assembly				
		#Contigs total	#Contigs >1000 bp	N50 for contigs >1000 bp	Mb in contigs >1000 bp	Mb total	#Contigs total	#Contigs >1000 bp	N50 for contigs >1000 bp	Mb in contigs >1000 bp	Mb total
DNA isolation 1	159,458,382	1,201,371	57,009	7183	209.7	486.0	556,160	116,123	5587	404.7	613.8
DNA isolation 2	421,986,642	2,003,618	94,702	11,536	425.9	876.6	947,772	200,495	5976	724.0	1069.9
DNA isolation 1 + 2	581,445,024	2,319,807	112,571	10,871	512	1,035.8	1,142,608	236,489	5823	843.0	1255.4[a]
DNA isolation 2 + additional sequencing of isolation 2	737,631,618	2,593,366	123,435	12,418	581.3	1161.7[b]					
DNA isolation 1 + 2 + additional sequencing of isolation 2	897,090,000	2,826,937	140,535	11,784	653.73	1292.2					

[a] Assembly uploaded to IMG/MER and used for the phylogenetic and comparative analyses

[b] Assembly used for metagenomic binning

was used to visualize contig coverages, length, GC content, taxonomy, and to assess possible binning strategies. For binning based on composition and differential coverage data, the software CONCOCT was used with default parameters [44]. As suggested in the CONCOCT documentation, contigs were cut up into sequences of 10 kb length. Subsequently, mapping of the initial reads was carried out using Bowtie2 [45] to determine the coverage of these contigs. Checkm 1.0.3 [46] was used to assess the completeness and contamination of the bins. Strongly contaminated bins were inspected by VizBin [47], which allowed a further separation of bins, if they formed two or more distinct clusters. Genes encoding presumable carbohydrate-active enzymes were annotated based on sequence similarity to sequences in the CAZY database [48]. The CAZy database (May 2015) was downloaded using a custom Python script. After this, a blastp sequence similarity search [49] of open reading frames, which were extracted from the assembly using getorf [50], was performed against the CAZy database using default parameters and an e-value cutoff of 1e−20.

All bins with completeness >80 %, contamination <10 % and heterogeneity (of the contamination) <50 % were classified as "high quality". Contigs belonging to these bins were removed from the assembly. According to the CONCOCT workflow, the remaining contigs were binned again using CONCOCT. The resulting bins were again evaluated using Checkm. The taxonomy of the bins was obtained from the AMPHORA2 results by determining the consensus lineage of all bin-specific marker genes (cut-off confidence scores >0.8). Annotation of the genome bins was performed using the annotation framework ConsPred V1.21 (http://sourceforge.net/p/conspred/wiki/Home/). In the file "conspred_input_specification.txt", the parameters "taxon exclude", "minimal number rRNA", and "minimal number tRNA" were set to "0".

Identification of carbohydrate-active gene candidates, PULs, and cellulosomal scaffoldin proteins

The assembled metagenomic contigs (biogas and elephant) and scaffolds (cow) were subjected to gene prediction using Prodigal 2.6.1 in meta mode [51]. The number of predicted open reading frames (ORFs) for the respective metagenome is indicated in Table 4. Amino acid sequences of the predicted ORFs were screened for similarity to glycoside hydrolase (GH) families and carbohydrate esterase (CE) families as classified in the CAZy database [48]. For this screening, profile hidden Markov models (HMMs) based on the respective CAZy families were downloaded from the dbCAN database [52] and compared to the protein sequences using hmmscan of the HMMER 3.1b1 software package (hmmer.org). All resulting hits were processed as recommended by the author of the dbCAN database. First, overlapping hits were removed; the hit with the higher e-value was discarded. Hits not covering at least 30 % of the respective HMM were also removed. For the remaining hits, an e-value cutoff of 1e−5 for alignments longer than 80aa and 1e−3 for alignments shorter than 80aa was applied. For the GH109 family, a custom made model was used and the covered fraction of the HMM was increased to 55 %. Duplicate hits in the family GH74 were removed by hand. To identify potential bacteroidetal PULs, the dbCAN database was extended by two additional models: a model for SusD like proteins (PF07980) downloaded from the Pfam database (http://www.pfam.xfam.org/) and a model for TonB-dependent receptor/SusC like proteins (TIGR04056) downloaded from the TIGRfam database (http://www.tigr.org/TIGRFAMs). To identify potential cellulosomal gene clusters in the respective metagenomic data set, we used amino acid sequences of known cellulosomal scaffolding proteins for an iterative protein sequence similarity search via Jackhmmer (hmmer 3.1 package). For this search, a score cut-off value of 700 was applied, and the following scaffoldin query sequences were used (NCBI accession numbers and organism names in brackets): cbpA (AAA23218.1, *Clostridium cellulovorans*), CipC (AAC28899.2, *Clostridium cellulolyticum H10*), cipA (BAA32429.1, *Clostridium josui*), cipA (AAK78886.1, *Clostridium acetobutylicum ATCC 824*), cipA (Q06851, *Clostridium thermocellum*), ScaA (AAG01230.2, *Pseudobacteroides cellulosolvens*), ScaB (AAT79550.1 *Bacteroides cellulosolvens*), and ScaB (CAC34385.1, *Ruminococcus flavefaciens 17*). To allow a comparison between the different sized assembled metagenomic data sets, the number of potential GH and CE gene hits was normalized to 1 Gb of assembled DNA for all comparative analysis.

Taxonomic assignment of carbohydrate-active gene candidates

Amino acid sequences of ORFs, which were previously assigned to GH families associated with cellulases and CE families, were used for a protein blast search against the NCBI non-redundant database. The number of maximal target sequences was decreased to 20 and an e-value cutoff of 1e−2 was employed for this search. Next, all obtained hits were loaded into MEGAN5 [40] and the lowest common ancestor (LCA) algorithm (default settings, unless otherwise specified) was used to classify the sequences taxonomically.

RNA extraction and sequencing from an elephant feces sample

A feces sample from an adult female zoo elephant was taken in April 2014 in the same way and from the same

animal described in the publication by Ilmberger et al. [15] living in the Hagenbeck Zoo in Hamburg, Germany. The sample was transported to the lab on ice and then stored at −70 °C. Isolation of ribonucleic acids for RNA-Seq was carried out using the PowerMicrobiome™ RNA Isolation Kit from Mo Bio Laboratories (Carlsbad, Germany) as recommended by the manufacturer. In a next step, ribosomal RNA was depleted using the Ribo-Zero™ rRNA Removal Kit for Bacteria (Illumina, Madison, USA) according to the manufacturer's instructions. The remaining transcripts were fragmented and cDNA libraries for Illumina sequencing were constructed by Vertis Biotechnology AG, Germany (http://www.vertis-biotech.com/), as described previously for eukaryotic microRNAs [53], but omitting the RNA size-fractionation step prior to cDNA synthesis. Equal amounts of RNA samples were poly(A)-tailed using poly(A) polymerase. Then, the 5′-triphosphates were removed by applying tobacco acid pyrophosphatase (TAP) resulting in 5′-monophosphat. Afterwards, a RNA adapter was ligated to the 5′-phosphate of the RNA. First-strand cDNA was synthesized by an oligo(dT)-adapter primer and the M-MLV reverse transcriptase. In a PCR-based amplification step, using a high-fidelity DNA polymerase, the cDNA concentration was increased to 20–30 ng/µl. A library-specific barcode for multiplex sequencing was part of a 3′-sequencing TruSeq adapter. The resulting cDNA libraries were sequenced using a HiSeq 2500 machine in single-read mode running 100 cycles.

RNA extraction and sequencing from a biogas fermenter sample

A biogas fermenter sample was taken in March 2015, mixed with an equal amount of RNAlater solution and immediately frozen on dry ice for transport. In the lab, the sample was stored at −70 °C.

Isolation of ribonucleic acids for RNA-Seq was carried out using the PowerMicrobiome™ RNA Isolation Kit from Mo Bio Laboratories (Carlsbad, Germany), as recommended by the manufacturer. In a next step, ribosomal RNA was depleted using the Ribo-Zero™ rRNA Removal Kit for Bacteria (Illumina, Madison, USA) according to the manufacturer's instructions. The rRNA-depleted samples were purified via the RNA Clean & Concentrator Columns from Zymo Research (Irvine, USA). During this step, an additional in-column DNase I treatment was included to ensure complete removal of DNA. Subsequently, synthesis of double-stranded cDNA was conducted using the Maxima H Minus Double-Stranded cDNA Synthesis Kit from ThermoScientific (Waltham, USA). In the first-strand cDNA synthesis reaction, 2 µl of random hexamer primer were used. Final purification of the blunt-end double-stranded cDNA was carried out using SureClean Plus solution from Bioline (Luckenwalde, Germany). The cDNA was sequenced in the same way as the total DNA. To achieve the required amount of cDNA for library preparation, multiple RNA isolations from the same sample were pooled.

Processing and analysis of RNA-Seq reads

To identify highly transcribed glycoside hydrolases in the biogas fermenter and elephant feces samples, RNA-Seq reads generated for both samples were checked for read quality and sequencing adapters were removed using Trimmomatic 0.33 [33]. In the next step, poly(A) tails >10 were removed using the trim_tail_left/right function of PRINSEQ lite 0.20.4 [54]. Subsequently, short sequences (≤99 nt) were filtered and rRNA gene-derived sequences were removed employing SortMeRNA 2.0 [55]. All remaining non-rRNA reads were used for mapping to the metagenomic contigs of the respective assembly (indicated in Table 4). For this mapping, Bowtie2 [45] was used in end-to-end mode (preset: very sensitive) and allowing 1 mismatch during seed alignment. After this, the htseq-count script from HTSeq 0.6.1 [56] was applied in non-stranded mode with default settings (alignment quality cutoff <10) to count the reads which map to genes predicted in the respective assembly by Prodigal 2.6.1. Finally, the genes were filtered for previously identified potential glycoside hydrolase encoding genes. The taxonomic origin of 100 GHs (including all CAZy families) and 50 GHs (including only cellulolytic CAZy families) with the highest numbers of mapped cDNA reads was determined via a protein blast and MEGAN5 LCA analysis as described above.

To analyze the expression level of cellulolytic GHs genes in the bacterial metagenomic bins created from the biogas fermenter metagenome, the initial raw reads were again processed using PRINSEQ lite: 10 bases were trimmed from the 5′ end, bases with a quality score <5 were trimmed from the 3′ end, and sequences with mean quality <20 or length <70 bp were discarded. Subsequently, the remaining RNA-Seq reads were mapped to the Ray assembly used for metagenomic binning (indicated in Table 1) via Bowtie2. Next, Bedtools multicov [37] was used to calculate coverage values of potential CAZy glycoside hydrolases which were identified in the metagenomic bins as described in the binning section. The coverage values were converted to rpkm values and plotted against bin taxonomy using the heatmap.2 function of the gplots package in R (R Core Team, 2015).

Transmission electron microscopy (TEM)

Slices were prepared using the microtome Reichert-Jung Ultracut E. Fixation was performed in 2 % glutaraldehyde in 75 mM cacodylate buffer (pH 7.0). Next, the samples

were supplied with 2 % agar in 75 mM cacodylate buffer (pH 7.0) and further fixed with 1 % OsO_4 in 50 mM cacodylate buffer (pH 7.0). After washing with 75 mM cacodylate buffer (pH 7.0), water was removed with acetone and the sample was infiltrated with Spurr resin (Polysciences, Warrington, PA, USA). TEM pictures were taken on the LEO 906 E using the camera Gatan 794 and the software Digital micrograph (Gatan GmbH, Munich, Germany).

DNS Assay for determination of total cellulolytic activities

To determine total cellulolytic activities in the sample materials, 0.2 g of biogas fermenter material as well as various fecal samples of herbivorous animals were diluted in 1 ml phosphate buffer (0.1 M, pH 6.6) containing 2 mM EDTA and 1 mM PMSF. The feces samples were obtained from animals living in the Hagenbeck Zoo in Hamburg, Germany. Subsequently, the samples were sonicated on ice for 15 min and centrifuged for 1 min at full speed. The supernatants were transferred to new tubes and 100 µl aliquots were used for total protein quantification via the Pierce™ BCA Protein Assay Kit as recommended by the manufacturer (Thermo Fischer Scientific, Pinneberg, Germany). In the next step, 100 µl of the remaining supernatants were used for the 3,5-dinitrosalicylic acid (DNS) assay. The assay was conducted in triplicates as described by Juergensen and colleagues [57]. Incubation of the samples with carboxymethylcellulose (CMC) was carried out at 37 °C for 90 min. For each sample, an additional reaction with buffer instead of CMC was conducted and used as a blank. The amount of reducing sugar ends was quantified at 546 nm using a SmartSpec Plus spectrophotometer (Biorad, Munich, Germany). For the calculation of specific enzyme activities, the obtained values were corrected against the measured total protein content. Data are mean values of three independent tests. One unit is defined as the amount of enzyme generating 1 µmol of reduced sugar per minute.

Sequence data deposition

Within the framework of this study, generated raw sequence data have been deposited under the NCBI BioProject number PRJNA301928. In addition, an assembly of the biogas fermenter metagenome can be accessed and downloaded via IMG/ER (https://www.img.jgi.doe.gov) using the IMG ID 3300002898. The 104 metagenomic bins are provided in the compressed Additional files 1, 2, 3 and 4.

Results and discussion

Characteristics of the analyzed biogas plant

Biogas plants harbor complex microbial communities that are essential for the different steps of biogas production. However, the overall biogas production rates are limited and depend on the initial hydrolysis of the plant biomass [2, 8, 58]. To identify possible limitations regarding the overall hydrolytic performance of biogas plants, we sampled and analyzed a typical one-stage agricultural plant with respect to its taxonomic structure and its metagenome content. A detailed overview about the process parameters of this plant is provided in the "Methods" section, and additional parameters are shown in Additional file 5: Table S1. Given the fermentation and process conditions, this plant is representative for several thousand one-stage plants across Europe. Samples were taken at two time points (March and May 2013) for DNA extraction and one time point (March 2015) for RNA extraction as described in the "Methods" section.

Community structure and diversity of the agricultural biogas plant

To analyze the community structure and the main actors in lignocellulose degradation, transmission electron microscopy (TEM), 16S rRNA gene amplicon, and metagenome sequencing were conducted. As expected, TEM microscopy indicated a high microbial diversity and a substantial number of cellulosome-producing bacteria in the studied biogas sample (Fig. 1a). Surprisingly, image analysis of several large decomposing plant cells implied that the cellulosome-producing microorganisms were in general closely attached or in proximity to the decomposing cell walls, while other microbes were only observed in some distance from the degrading cell walls. This was also observed when other samples from the same biogas reactor were analyzed using TEM (Additional file 6). In fact, the cellulose decomposing bacteria formed a loose layer or biofilm that was mostly not penetrated by other microorganisms. This is an intriguing and novel observation, since it implies that the cellulosome-producing bacteria have a competitive advantage and that they can colonize and degrade the plant material in the absence of other bacteria. A similar observation was not made when samples from elephant feces [15], a herbivore that is known for its richness in cellulolytic enzymes, were analyzed. Intrigued by this observation, we assayed total cellulolytic activities in the supernatant of biogas fermenter content and feces samples obtained from various herbivores. In all cases, the biogas sample had the lowest cellulolytic activities (Additional file 7). It was approximately three–five-folds less active than supernatants obtained from mara, elephant, cow, and zebra feces samples. Even though carboxymethylcellulose is a model substrate which cannot reflect the total hydrolytic activity of diverse glycoside hydrolases and sample treatment (e.g. sonication) might have an effect on the initial activity, this finding suggested that there were major differences

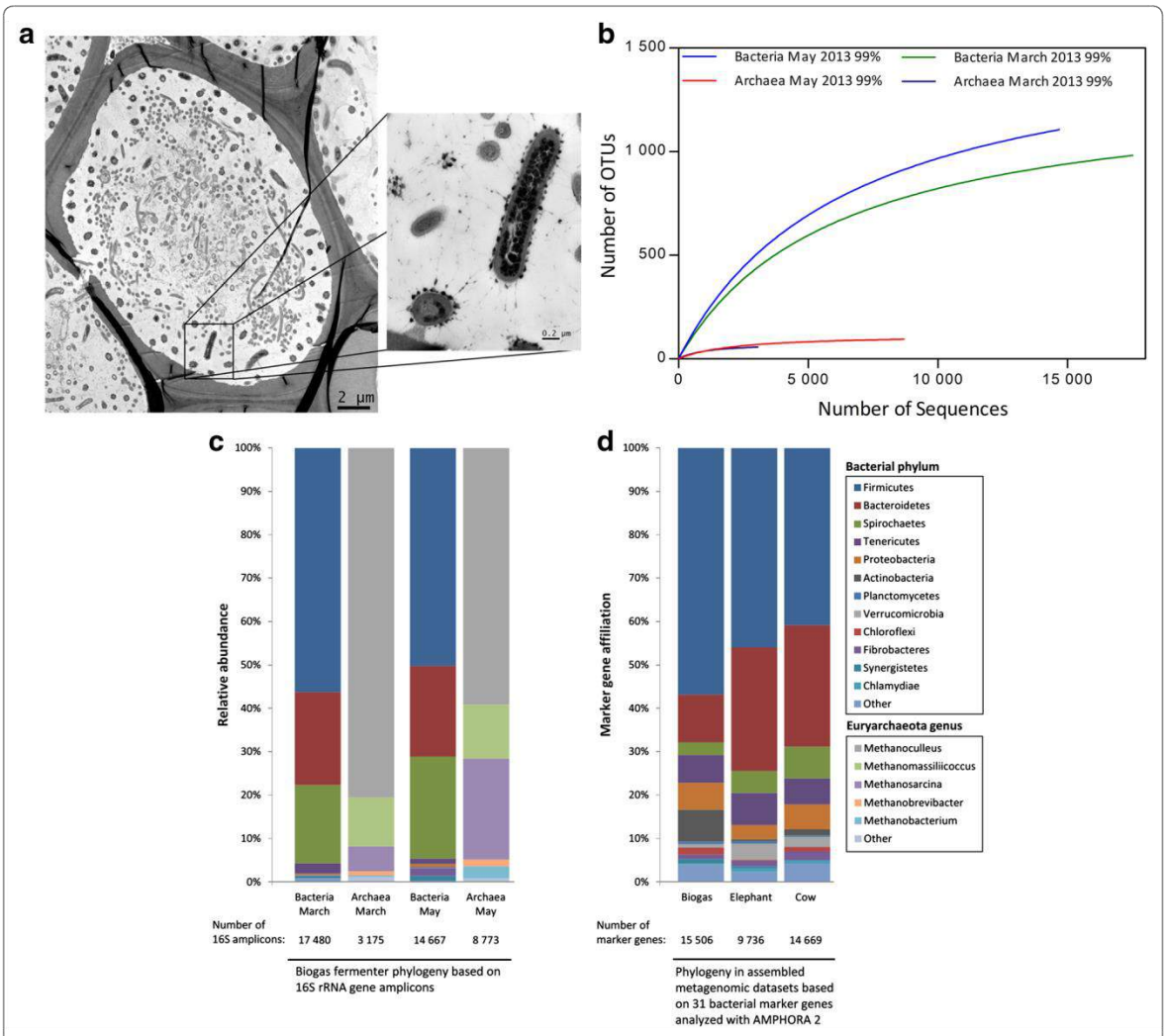

Fig. 1 **a** TEM micrograph of a decomposing plant cell and the associated microorganisms in a biogas fermenter. Cellulosome-producing bacteria are almost exclusively observed in close association with the plant cell wall, where they appear to suppress growth of other microbes. Most other microorganisms are located at the more central part of the decomposing plant cell. Cellulosome-producing bacteria were identified by the large *dark spots* attached to the cells. **b** Rarefaction *curves* calculated for two fermenter samples of the studied agricultural biogas plant. The OTUs were clustered at 99 % genetic similarity of 16S rRNA genes. The sequences were denoised employing Acacia, and chimeric sequences were removed using UCHIME. Singleton OTUs were removed prior to the rarefaction analysis. **c** Phylogenetic analysis of two biogas fermenter samples based on 16S rRNA gene amplicons. The *bars* indicate the relative abundance of bacterial phyla and euryarchaeota genera in two samples taken from the same fermenter at different time points (March and May 2013). **d** Phylogenetic analysis of three assembled metagenomic data sets based on 31 bacterial marker genes. The *bars* show the marker gene affiliation to bacterial phyla in the data sets derived from biogas fermenter May sample and for reasons of comparison from published data sets of elephant feces and cow rumen samples [15, 23]. For this analysis, the AMPHORA 2 software was used

within the bacterial communities which would lead to the observed different activity profiles.

To get a further insight into the diverse community of the studied biogas plant, 16S rRNA gene amplicon sequencing was carried out. Clustering of bacterial 16S rRNA gene amplicons at a 99 % similarity level resulted in 994 OTUs for the March 2013 sample and 1108 OTUs for the May 2013 sample (Fig. 1b). Archaeal diversity was substantially lower in the samples. Archaeal-derived 16S rRNA gene amplicons were clustered to 59 (March) and 95 (May) OTUs. An additional information regarding diversity and richness (Chao1, Shannon-Index) is

provided in Additional file 5: Table S2. A direct comparison between the samples from our plant indicated that 691 bacterial and 52 archaeal OTUs could be observed in both analyzed samples Additional file 5: Table S3. This relatively high number of shared OTUs suggests the presence of a mostly steady microbial core community in the biogas fermenter under the constant process conditions. Within this framework, the taxonomic classification of the bacterial 16S amplicons indicated that most of the obtained sequences (March 56 %, May 50 %) were affiliated with the phylum of the *Firmicutes* (Fig. 1c). Other abundant phyla were the *Bacteroidetes* with a relative abundance of 21 % in both samples and the *Spirochaetes* (March 18 %, May 24 %). The archaeal-derived 16S rRNA amplicons were all classified into the phylum of *Euryarchaeota.* In both samples, the most abundant archaeal genus appeared to be *Methanoculleus* (March 81 %, May 59 %). The next abundant genera were *Methanomasiliicoccus* (March 11 %, May 13 %) and *Methanosarcina* (March 6 %, May 23 %). A more detailed taxonomic breakdown of the 16S rRNA gene analysis results is given in Additional file 8.

Metagenome-based analysis and binning of the May 2013 biogas sample

In addition to amplicon sequencing, a large data set of metagenomic DNA was produced for the May sample. We generated 581 million reads which were assembled to 236,489 contigs (>1000 bp) with a total of 2 million potential open reading frames and 1.25 Gb of assembled DNA (Table 1). This data set comprises the largest currently assembled and published data set for a biogas reactor so far and it is, with respect to cellulolytic communities, the second largest metagenome currently published. Only the data set obtained from a microbiome adherent to switchgrass, which was incubated in the rumen of a cow for 72 h, is larger with respect to the assembly [23].

We utilized this comprehensive assembly to further investigate the community structure in the fermenter and found distinct differences in the phylogenetic make up and the overall diversity compared to the amplicon-based data. For this, we employed the AMPHORA2 software [42], which uses 31 conserved bacterial proteins as phylogenetic markers. A total of 15,506 marker genes were identified in our biogas metagenome data set and classified. Thereby, the analysis revealed that 57 % of the marker genes were affiliated with the *Firmicutes* followed by *Bacteroidetes* (11 %), *Actinobacteria* (7 %), *Tenericutes* (6 %), *Proteobacteria* (6 %), *Spirochaetes* (3 %), and other phyla (Fig. 1d). Almost 50 % of all identified marker genes were assigned to the class *Clostridia* followed by *Bacteroidia* (9 %), *Actinobacteria* (7 %), *Mollicutes* (6 %), and other classes.

In general, the phylogenetic structure of our sampled biogas plant appeared to be similar to the structures described in already published studies of biogas fermenters. These published studies were in part based on 16S rRNA amplicon analyses, but also based on metagenome data sets [2–4, 25]. In these studies, it was repeatedly reported that the *Firmicutes* were the prevalent phylum followed by the *Bacteroidetes.* By comparing the ratios of the *Firmicutes* versus the *Bacteroidetes* in already published studies of agricultural biogas fermenters and our own analysis, we found that the mean ratio of the *Firmicutes* versus the *Bacteroidetes* was 5.6–6.0:1 (Table 2 and included references) indicating an, on average, almost six-fold higher relative abundance of the *Firmicutes* compared to the *Bacteroidetes* in the analyzed fermenters.

To further examine the microbial community in the biogas fermenter, a metagenomic binning based on composition and differential coverage data was performed for the May 2013 sample. For this analysis, the Ray Meta assembly with the highest N50 value was used (Table 1) and binning was conducted. Thereby, 104 high-quality bins were observed and assigned to four binning categories (Table 3). The binning results basically reflected the population structure as determined by the marker gene analysis. In total, 57 of the high-quality bins were affiliated with the *Firmicutes* and most of these with the class Clostridia (51). The second most abundant taxonomic bin classification was *Bacteroidetes* with 21 observed bins of which 16 were further attributed to the class *Bacteroidia*. The remaining bins were mainly affiliated with the *Fibrobacteres* (3), the *Spirochaetes* (4), the *Actinobacteria* (2), the *Verrucomicrobia* (2), and the *Euryarchaeota* (3). Nine bins were not assigned to a specific bacterial phylum. A detailed overview about the taxonomic classification of the bins, the estimated bin completeness, the bin contamination, and the size of the bin is provided in Additional file 5: Table S4. For some of the bins, a taxonomic classification with high confidence was possible down to the species level. For example, highly complete genome reconstructions were possible for *Fibrobacter succinogenes* or a closely related species (Bin-IDs pb121 and pb122). In addition, the binning of genome drafts of the cellulolytic bacteria *Clostridium thermocellum* (Bin-ID 96) and *Clostridium phytofermentans* (Bin-IDs pb35-2, pb186-2, and pb35-1) was possible. In the class *Bacteroidia*, comprehensive genome binnings were obtained for species closely related to *Paludibacter propionicigenes* (Bin-IDs 145 and 201) and a not further classified organism of the family *Porphyromonadaceae* (Bin-ID pb69). Finally, a genome bin of the methanogenic archaeon *Methanosarcina barkeri* was reconstructed with only minor contamination (Bin-ID pb85). Besides the known presence of methanogenic archaea within biogas

Table 2 Ratio of the phyla *Firmicutes* vs. *Bacteroidetes* in biogas fermenters and herbivorous animals

Microbiome/community	*Firmicutes/Bacteroidetes* phyla ratio		Reference/data source
	16S amplicon-based analysis	Metagenome-based analysis	
Biogas fermenter	2.4:1	5.2:1	This study
Biogas fermenter	–	4.7:1	[25]
Biogas fermenter (dry fermentation)	–	4.1:1	[3]
Biogas fermenter (wet fermentation)	–	3.9:1	[3]
Biogas fermenter	9.6:1	5.9:1	[4]
Biogas fermenter	–	9.6:1	[2]
Mean	*6:1*	*5.6:1*	
Asian elephant feces	0.8:1	1.6:1	[15]
Switchgrass incubated in cow rumen	–	1.5:1	[23]
Rumen of a hay-fed cow	0.8–1.7:1	–	[14]
Svalbard reindeer Rumen	0.5:1	0.4:1	[18]
White rhinoceros feces	1.6–2.7:1	–	[24]
Rex rabbit feces	0.8–1.3:1	–	[16]
Rumen of a pasture-fed sheep	0.3–0.5:1	–	[14]
Giraffe rumen	1.6:1	–	[17]
Mean	*1:1*	*1.2:1*	

The data are based on the data sets published in the indicated references or on data produced in this study

Table 3 Binning summary of biogas fermenter May 2013 sample

Bin category	Quality criteria	# of bins in category
Good bins	>95 % completeness and <5 % contamination or >95 % completeness and <10 % contamination with >90 % heterogeneity	20
Nearly complete genome drafts	>90 % completeness and <5 % contamination	20
Nearly complete pangenome drafts	>90 % completeness and >5 % contamination	37
Incomplete genome drafts	60–90 % completeness and <7 % contamination	27

fermenters, the number of high-quality bins related to this taxonomic group was rather small compared to the number of bins obtained for bacteria. In general, the binning of metagenomic contigs into high-quality genome bins allows the reconstruction of key metabolic features of these organisms or OTUs. This is especially useful for, so far, under-represented groups which have not been studied in much detail. Thus, the bins generated during this study provide the basis for a future in-depth analysis of the metabolism and physiology of these organisms.

A direct comparison of the biogas plant microbial community with fecal and rumen microbiomes of herbivores indicates major differences in the ratio of the *Firmicutes* versus the *Bacteroidetes*

Based on the phylogenetic analysis of the biogas fermenter analyzed in this study and the studies listed in Table 2, we asked if the relative ratio of the phylum of the *Firmicutes* versus the *Bacteroidetes* could be used as an indicator for the fitness of a biogas plant and if this would be indicative for the diversity of plant biomass degrading genes. Since most agricultural biogas fermenters are inoculated and fed with various animal manures, it is likely that the *Bacteroidetes* are present at high levels initially, but are then outcompeted. Reasons for this shift in the microbial community might be the operation conditions in the biogas reactors or the lack of growth factors that are usually present in the natural habitats of the *Bacteroidetes*. Within this framework, it is noteworthy that our fermenter was also fed with cow manure and chicken manure that both contain high levels of the *Bacteroidetes* [59, 60].

In a next step, we calculated the ratios of the *Firmicutes* versus the *Bacteroidetes* in published fecal, rumen, and gut samples of herbivorous animals which nurture from various plant-derived biomasses. The resulting ratios, as well as the methods used in the original publication for the analysis of the microbial community, are indicated

in Table 2. Interestingly, when we analyzed the published data, we observed a mean ratio of almost 1:1 (*Firmicutes* vs. *Bacteroidetes*) for seven out of eight analyzed natural microbiomes. Only the published microbiome of the white rhinoceros revealed a slightly higher ratio of 1.6–2.7:1. Altogether, these analyses imply an almost equal abundance of *Firmicutes* and *Bacteroidetes* in these natural systems. This is in contrast to the ratios observed for the biogas fermenter in this study and many other published examples. In all these studied biogas fermenters, the *Firmicutes* were usually 4–6 times more prominent than the *Bacteroidetes*. Thus, it is likely that bacteria affiliated with *Bacteroidetes* do not compete as well in agricultural biogas plants compared to their natural habitats and compared to the *Firmicutes*. It is tempting to speculate that a decreased abundance of the *Firmicutes* together with an increased abundance of the *Bacteroidetes* might be an indicator for the fitness of biogas plants with respect to cellulolytic activities.

In-depth analysis of predicted glycoside hydrolase and carbohydrate esterase family enzyme abundance and origin largely confirms the phylogenetic analyses and implies a lower enzyme abundance compared to natural microbiomes

Intrigued by the above-made findings, we wanted to investigate how the different ratios of the *Firmicutes* vs. the *Bacteroidetes* might affect the abundance of genes encoding GH family enzymes which are involved in the breakdown of the plant biomass and especially the lignocellulose. To address this question, we analyzed our assembled metagenomic data set obtained from the biogas fermenter with respect to the predicted diversity of hydrolytic enzymes involved in lignocellulose degradation and with respect to the taxonomic origin of these genes and enzymes. For this analysis, we used profile hidden Markov models, which were based on entries in the carbohydrate-active enzyme database (CAZy). The CAZy database encompasses a large set of validated carbohydrate-active enzymes and offers a sequence-based family classification of enzymes that are involved in the modification or breakdown of polysaccharides [48]. Within the up to 2 million-predicted potential genes of the analyzed biogas fermenter sample, we identified a total of 17,305 putative genes for glycoside hydrolases from 109 different families. This equals 13.8 GHs per Mbp of assembled DNA. With respect to the lignocellulose degradation, the most predominant GH families observed were GH3, GH5, GH9, GH51, GH74 and GH94 family enzymes. The total number of potential hits observed in the respective family was 977 for GH3, 599 hits for GH5, 216 for GH9, 265 for GH51, 373 for GH 74 and 269 for GH94. These GH families encompass a variety of hydrolytic enzymes, e.g. cellulases, endo- and exoglucanases, arabinofuranosidases, endoxylanases, cellobiohydrolases, and xyloglucanases.

In a next step, we assigned the taxonomic origin to the identified potential cellulolytic GH encoding genes via a protein blast search against the NCBI non-redundant database in combination with the MEGANs LCA algorithm. Using this approach, we were able to elucidate the phyla which contributed most to the hydrolytic metagenomic potential in the fermenter. We found that most of the predicted enzymes belonging to the cellulolytic GH families 1, 3, 5, 8, 9, 30, 45, 51, 74 and 94 showed the highest coverage for affiliates of the *Firmicutes* in the biogas fermenter (Fig. 2a). Notably, the number of predicted enzymes affiliated with the phylum of the *Bacteroidetes* was much lower in the analyzed cellulolytic GH families. A small fraction of predicted enzymes originated from the phyla *Actinobacteria, Spirochaetes*, and *Tenericutes*.

A similar analysis was done for the CE family genes and enzymes. CE family enzymes are mainly carbohydrate-active esterases and were recently introduced into the CAZy database. Within our data set, predicted genes for all 16 known CE families were covered and a total of 7655 genes possibly encoding for CE family enzymes were identified. This equals 6100 CEs per Gb of assembled DNA. The most predominant families were 1, 4, and 10. Thereby, we observed 1826 hits for family CE1, 1697 hits for family CE4, and 1198 hits for family CE10. Altogether these findings implied a high GH and CE enzyme diversity within the data set and suggested that the majority of the genes coding for these enzymes were derived from the *Firmicutes*.

To relate the above-made findings to other highly cellulolytic microbial communities, we compared these values to published and very comprehensive studies of two natural systems. The first data set was derived from a microbiome adherent to switchgrass, which was incubated in the rumen of a cow for 72 h [23]. The second data set was obtained from the feces of an adult Asian zoo elephant and published in 2014 by our group [15]. Both of these assembled metagenomic data sets were of a similar size as the data set generated in this study from an agricultural biogas plant. These data sets are, to our knowledge, the largest data sets of cellulolytic communities publicly available so far. In the original studies, both samples were described as being highly diverse and rich in hydrolytic enzymes. Within this framework, it is noteworthy that cows and elephants rely on different strategies for the digestion of their food. Elephants are hindgut fermenters and degrade their diet in the caecum, whereas cows are foregut fermenters and digest their food in the rumen [61]. While cows mainly fed on grass, elephants

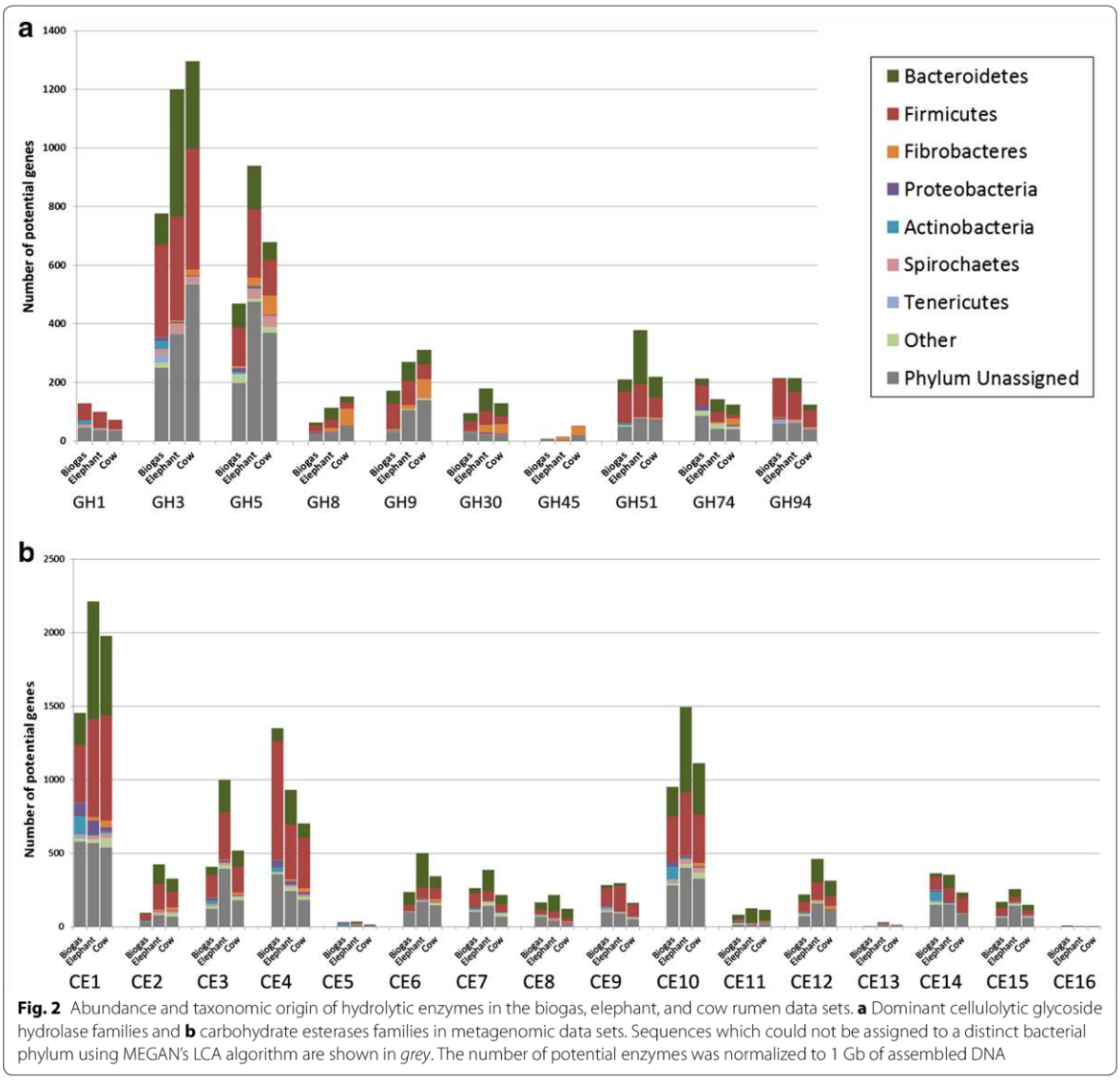

Fig. 2 Abundance and taxonomic origin of hydrolytic enzymes in the biogas, elephant, and cow rumen data sets. **a** Dominant cellulolytic glycoside hydrolase families and **b** carbohydrate esterases families in metagenomic data sets. Sequences which could not be assigned to a distinct bacterial phylum using MEGAN's LCA algorithm are shown in *grey*. The number of potential enzymes was normalized to 1 Gb of assembled DNA

digest a wider spectrum of plant-derived biomass, and as a consequence, their microbiomes are, in part, different.

The two published data sets were downloaded from IMG/MER (https://www.img.jgi.doe.gov/er/) and used for the comparative analyses with respect to the diversity and abundance of GH and CE enzymes (Table 4). We identified 22.5 GHs/Mbp of assembled DNA in the Elephant feces data set and 14.9 in the Cow data set compared to 13.8 in the biogas data set. For this comparison, the quantity of potential GH/CE encoding genes was analyzed independently of contig coverage. The abundance and distribution of the predicted genes coding for GH family enzymes in 1 Gb of assembled DNA are shown in Fig. 3. In a next step, the genes possibly coding for GH family enzymes involved in the breakdown of cellulose were examined in detail. For this, we included the 9 most abundant cellulolytic GH families which were observed in the biogas data set and compared their abundance and taxonomic origin in all three data sets (Fig. 2a). Interestingly, out of the 9 analyzed cellulolytic GH families, the biogas data set revealed the lowest number of potential enzymes in 7 GH families. Our analysis suggests that the lower abundance of GHs in these families can be attributed to a partial lack of

Table 4 General overview on the metagenomic and hydrolytic potential of the analyzed biogas microbial community in relation to two other studies

Metagenome studied	Total DNA in assembly (Gbp)	DNA assembled[a] (Gbp)	No. of predicted ORFs	No. of predicted GHs	No. of predicted CEs	GHs/Mbp	CEs/Mbp	Cellulolytic GHs/Mbp[c]	Ratio cellulolyti GHs *Firmicutes*/ *Bacteroidetes*[c]	Ratio CEs *Firmicutes*/ *Bacteroidetes*	Data source
Agricultural biogas plant fermenter	58.7	1.25	2,011,807	17,305	7655	13.8	6.1	2.5	2.8:1	2.4:1	IMG ID: 3300002898 This study
Elephant feces from zoo animal	54.7	0.92	1,005,402	20,705	8344	22.5	9.1	3.8	1: 1	0.9:1	IMG ID: 3300001598 [15]
Cow rumen, switch grass	111.4	1.55[b]	2,083,556	23,110	9891	14.9	6.4	3.2	1.4:1	1.3:1	IMG ID: 2061766007 [23]

[a] DNA in this study was assembled with the IDBA-UD assembler; DNA in the studies by Ilmberger [15] and Hess [23] was assembled using velvet

[b] Uncalled nucleotides (Ns) from scaffolds not included

[c] GH families included: GH1, GH3, GH5, GH6, GH8, GH9, GH12, GH14, GH30, GH44, GH45, GH48, GH51, GH74 and GH94

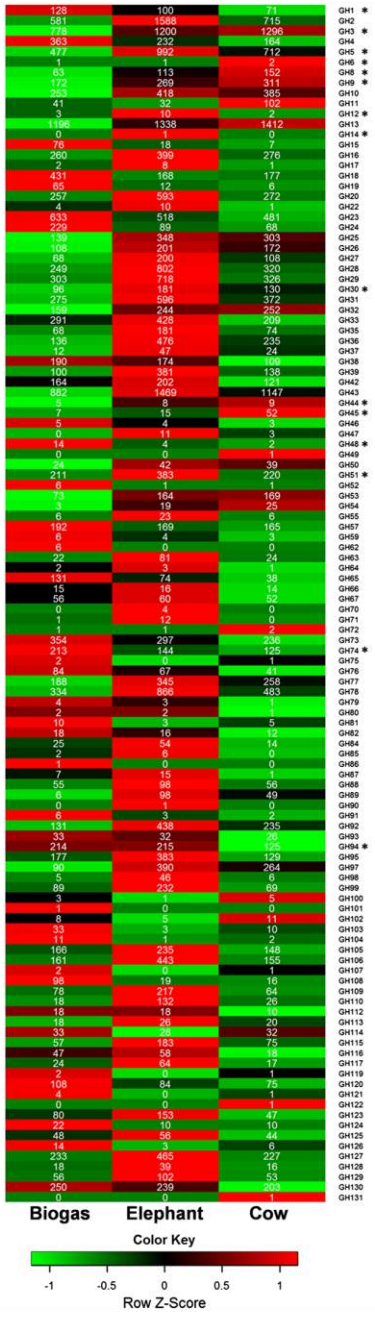

Fig. 3 Heatmap indicating the abundance and distribution of potential GH family enzymes in the assembled metagenomic data sets. Rows are color coded according to *Z*-score. A *Z*-score change of +1 is equal to one standard deviation above the row mean. GH families involved in cellulose degradation are labeled with an *asterisk*. GH families not listed were not observed in the data sets. The amount of potential enzymes was normalized to 1 Gb of assembled DNA

enzymes derived from bacteria affiliated with *Fibrobacteres* and, especially, *Bacteroidetes*. Since both of these phyla comprise important polysaccharide-degrading bacteria in the gut and rumen of animals [12], an underrepresentation of genes coding for GHs derived from these groups implies a potentially important limitation in the biogas fermenter with regard to the hydrolysis of biomass.

With respect to the carbohydrate esterase family enzymes, a similar observation was made. Of the 16 families analyzed, only the CE families 4 and 14 had equal or more hits in the biogas data set compared to the elephant feces and cow rumen data sets (Fig. 2b). Distinct differences in the abundance of potential CEs were observed in the families CE1, CE2, CE3 and CE10. Enzymes assigned to these families share diverse enzymatic activities and substrate specificities, including acetyl xylan esterases and feruloyl esterases. Both of these groups of enzymes have been shown to be important accessory enzymes involved in the degradation of lignocellulosic biomass [62, 63]. A decreased overall diversity of CEs in the biogas fermenter might point to a disadvantage in the ability to efficiently degrade biomass compared to the two natural systems.

In addition to the identification of potential GHs and CEs encoding genes, we wanted to assess the presence of cellulosome encoding gene clusters in the different metagenome data sets. For this analysis, we used amino acid sequences of known cellulosomal scaffolding proteins and an iterative protein sequence similarity search with a cut-off score of 700. We identified a total of 3 hits in the biogas fermenter metagenome data set, 2 hits in the cow rumen data set, and no hits in the elephant feces data set (Additional file 5: Table S6). This analysis demonstrates a reduced diversity of cellulosome-producing bacteria in the cow rumen data set and, possibly, a complete lack of cellulosomes in the elephant feces data set. The potential cellulosomal scaffolding encoding genes, which were identified in the biogas data set, were allocated to the bacterial genome bins 96, pb35-1, and pb235-1. While the bin 96 was assigned to the genus *Clostridium,* the bins pb35-1 and pb235-1 were both assigned to the family *Lachnospiraceae.* Interestingly, a nucleotide blast search of the putative cellulosomal scaffolding genes found in the bin pb35-1 showed a 99 % identity to the recently described thermophilic cellulose-degrading bacterium *Herbinix hemicellulosilytica* which was isolated from a thermophilic biogas reactor [64, 65]. This finding suggests that this organism is also present in the community of our sampled fermenter and indicates that this species produces cellulosomes.

RNA-Seq identifies metabolically highly active bacterial and archaeal groups as well as highly transcribed genes in the biogas fermenter

In the light of the above-made findings, we asked which families of GH enzymes were highly transcribed in the biogas fermenter at the time of sampling. In addition, we wanted to know, whether the highly transcribed genes were affiliated mainly with the *Bacteroidetes*, the *Firmicutes,* or other phyla, and relate our findings to a natural cellulolytic system. For this, we conducted RNA-Seq of a biogas fermenter sample taken from the same biogas plant at a later time point and an elephant feces sample. Because the initial elephant feces data set was published by our group, we had access to samples from the same animal, as described in the original publication by Ilmberger et al. [15]. Due to difficulties obtaining a comparable sample as analyzed by Hess and colleagues [23], we did not include the cow rumen in the RNA-Seq-based analysis. RNA extraction, sample preparation, sequencing, and data processing are described in "Methods" section. An overview about the processing steps and the number of cDNA sequence reads obtained for both samples is provided in Table 5.

In a first step, we wanted to get a general idea of the functional and taxonomic affiliation of highly transcribed genes in the sampled biogas fermenter. To do so, we examined the 100 ORFs with the highest levels of absolute transcripts via a protein blast search against the non-redundant NCBI protein database. It is important to state that this does not necessarily mean that these ORFs, seen individually, are the highest expressed ones. For 94 ORFs, homologs in the non-redundant protein data base were observed, while for the remaining 6 predicted ORFs no homologies were observed at all (Additional file 5: Table S5). The largest fraction of 31 ORFs showed the highest similarity to hypothetical proteins. Thus, this result might suggest that many gene functions of the microbial community in biogas fermenters are not well characterized. Of the ORFs with an assigned function, 20 ORFs were affiliated with bacterial ABC transporter substrate-binding proteins, and 13 ORFs scored the best hit for archaeal enzymes involved in methanogenesis. A large fraction of these methanogenesis-related ORFs encoded for different subunits of the Methyl-coenzyme M reductase.

Furthermore, 85 ORFs could be taxonomically classified using MEGANs LCA method. Of these, 67 ORFs were of bacterial origin and 17 of archaeal. The bacterial ORFs mainly originated from the *Firmicutes* (50) and within this phylum from the genus *Clostridia* (38). Notably, a large fraction of 23 ORFs was assigned to uncultured bacteria of the family *Peptococcaceae* suggesting a very high metabolic activity of these physiologically diverse and partly acetogenic bacteria [66] in our sampled fermenter. In addition, 6 ORFs were assigned to the order *Halanaerobiales*. The majority of species in the order *Halanaerobiales* is known for sugar fermentation or homoacetogenesis [67]. These results might indicate a high relevance of these two bacterial groups for acidogenesis and acetogenesis in this biogas fermenter and perhaps for agricultural biogas reactors in general. Most of the archaeal ORFs were assigned to the class *Methanomicrobia* (9) and within this class to the genus *Methanoculleus* (5). Two highly transcribed ORFs originated from the hydrogenotrophic methanogen *Methanoculleus bourgensis.* This finding is in accordance with recent research suggesting a predominant methanogenesis via the hydrogenotrophic pathway in agricultural biogas plants [68, 69]. Finally, two highly transcribed ORFs were attributed to yet uncultured archaea.

Table 5 RNA-Seq processing steps and number of cDNA sequence reads obtained for elephant feces and biogas fermenter samples

Processing step		Elephant	Biogas
Pre-processing	No. of input reads	141,700,987	282,930,624
	After removal of polyA tails and short sequences <99	103,591,215	282,711,062
SortMeRNA	After rRNA removal	77,904,289	274,661,662
Bowtie2	Input reads	77,904,289	274,661,662
	Reads aligned 0 times	62,293,074 (79.96 %)	49,200,963 (17.91 %)
	Reads aligned exactly 1 time	12,328,533 (15.85 %)	172,293,650 (62.73 %)
	Reads aligned >1 times	3,282,682 (4.21 %)	53,167,049 (19.36 %)
	Overall alignment rate	20.04 %	82.09 %
HTSeq-count	Counted	8,140,194 (52.14 %)	159,743,444 (70.85 %)
	No feature	3,784,252 (24.24 %)	37,124,802 (16.47 %)
	Ambiguous	382,196 (2.45 %)	6,133,988 (2.72 %)
	Too low alignment quality (MAPQ <10)	3,304,573 (21.17 %)	22,458,465 (9.96 %)

RNA-Seq data imply that highly transcribed GH encoding genes in the biogas fermenter mainly originate from the Firmicutes, while bacteroidetal GH encoding genes are most transcribed in the elephant feces

In a next step, we wanted to examine differences in the transcription of GHs enzymes in the biogas and elephant samples in detail. For this purpose, we restricted our RNA-Seq data analysis to ORFs, which were previously identified to potentially encode for GHs. Thereby, we examined the taxonomic origin and CAZy family distribution of 100 GHs (including all CAZy families) and 50 GHs (including only cellulolytic CAZy families) with the highest numbers of mapped cDNA reads. We found that the majority of GHs in both of these groups was affiliated with the *Firmicutes* in the biogas fermenter sample (Fig. 4a). The ratios of GHs derived from the *Firmicutes* versus the *Bacteroidetes* were 2.3:1, including all families and 4.3:1, including only the cellulolytic families. These results further supported the observation that the *Firmicutes* are the predominant group responsible for the hydrolysis of biomass in the biogas fermenter. In contrast, the same analysis conducted for the elephant feces indicated that most of the highly transcribed GHs originated from the *Bacteroidetes*. Here, the ratios of GHs derived from the *Firmicutes* versus the *Bacteroidetes* were 0.68:1, including all families and 1:1, including only the cellulolytic families. Compared to the biogas fermenter, a larger fraction of GHs was also affiliated with the *Fibrobacteres*.

Furthermore, when examining the CAZy family distribution of the 50 cellulolytic GHs with the highest levels of absolute transcript, we observed that in the elephant feces, most of the putative enzymes were assigned to the family GH51 followed by GH5 and GH9 (Fig. 4b). In the biogas fermenter, the GH9 family was most frequently observed followed by GH5 and GH51. While these three GH families are all involved in the hydrolysis of lignocellulose, they differ in their substrate specificities. Interestingly, the GH 51 family contains many hemicellulases while the GH9 family mainly includes cellulose-specific enzymes. Altogether these results supported the notion that there are distinct differences in the cellulolytic bacteria and enzymes involved in the degradation of lignocellulosic biomass between the biogas fermenter and the elephant feces sample. In the biogas fermenter, highly transcribed cellulolytic GHs were four times more often affiliated with the *Firmicutes* compared to the *Bacteroidetes* (ratio 4.3:1), while an almost equal distribution of these enzymes was observed in the elephant feces sample (ratio 1:1).

RNA-Seq identifies transcription of cellulolytic GH family enzymes in the bacterial bins generated from the biogas fermenter metagenome

To further investigate the transcription of cellulolytic GH families in individual organisms in the biogas fermenter, we utilized the binned bacterial contigs which were generated from the biogas fermenter metagenome. For this, RNA-Seq data were mapped onto the binned bacterial contigs (Fig. 5a). Although the binned DNA represented only a part of the complete metagenome, our analysis shows that in individual bacteroidetal genome bins, multiple cellulase-encoding genes were strongly transcribed (e.g. Bin-IDs 36, 138 and 142). An in-depth analysis of these three bins resulted in the identification of multiple PULs, including three putatively cellulolytic PULs (Fig. 5b). In these PULs, cellulase-encoding genes

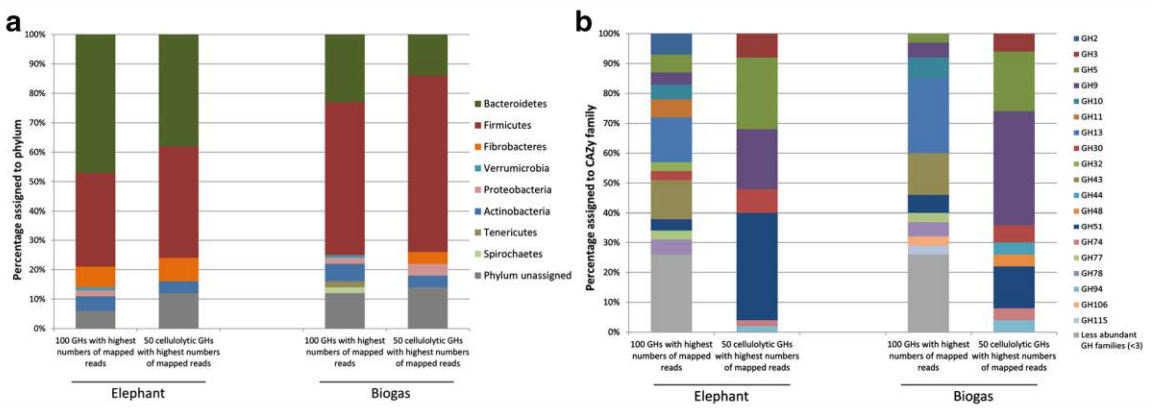

Fig. 4 **a** Taxonomic origin and **b** CAZy family distribution of 100 GHs and 50 cellulolytic GHs with highest numbers of mapped cDNA reads obtained from RNA-Seq of a biogas fermenter sample in relation to an elephant feces sample. Cellulolytic GHs include the families: GH1, GH3, GH5, GH6, GH8, GH9, GH12, GH14, GH30, GH44, GH45, GH48, GH51, GH74 and GH94. For this analysis, Megan LCA parameters were adjusted to "top percent 40" and "LCA percent 50" for assignment of phyla

were identified next to or in close proximity to an *SusC* and *SusD* gene pair. Interestingly, a two-component system histidine kinase gene was also found in all putatively cellulolytic PULs. The presence of this regulatory system might suggest a differential expression of PUL associated genes and enzymes in response to the respective "target" polymer. An induction of PUL gene transcription in response to specific plant polysaccharides was already shown for *B. ovatus* and *B. thetaiotaomicron* [70]. The identification of high transcription levels of cellulase-encoding genes in the bacteroidetal genome bin 36, together with the existence of potential cellulolytic PULs in the same bin, might provide an explanation how the hydrolysis of biomass is carried out by cellulolytic bacteroidetal species. In our opinion, this finding further supports the notion that the importance and potential of bacteroidetal organisms for the degradation of biomass in biogas fermenters were most likely under-estimated in the past.

This analysis also demonstrates and confirms the predominant expression of cellulases by various *Firmicutes*. Not surprisingly, numerous organisms belonging to the class Clostridia showed transcription of cellulolytic enzymes. Particularly, high levels of transcription of cellulolytic enzymes were observed for the clostridial bins 96 and pb215. The genome bin 96 was taxonomically classified as *Clostridium thermocellum* and in agreement with this classification, the cellulosomal-scaffolding protein A of *C. thermocellum* was identified in this bin (Additional file 5 : Table S6). The bin pb215 was classified as a not further specified *Ruminiclostridium*.

Conclusion

In this paper, we provide evidence that the analyzed biogas fermenter contains a relatively lower abundance of glycoside hydrolases and carbohydrate esterases involved in the breakdown of lignocellulosic biomass as compared to two natural plant biomass degrading systems. This difference can be attributed to a partial lack of enzymes derived from bacteria affiliated with the *Fibrobacteres* and, especially, the *Bacteroidetes*. The partial deficiency of these enzymes implies a potentially important limitation in the biogas fermenter with regard to the initial hydrolysis of biomass. In addition, we were able to show that the mean ratio of the phyla *Firmicutes* vs *Bacteroidetes* is close to 1:1 in various fecal or gut microbiomes of herbivorous animals, while the *Bacteroidetes* were usually 5–6 times less prominent in the mainly agricultural biogas fermenters listed in Table 2.

In accordance with this observation, RNA-Seq data showed that highly transcribed cellulolytic GHs in the biogas fermenter were four times more often affiliated with the *Firmicutes* compared to the *Bacteroidetes*, while

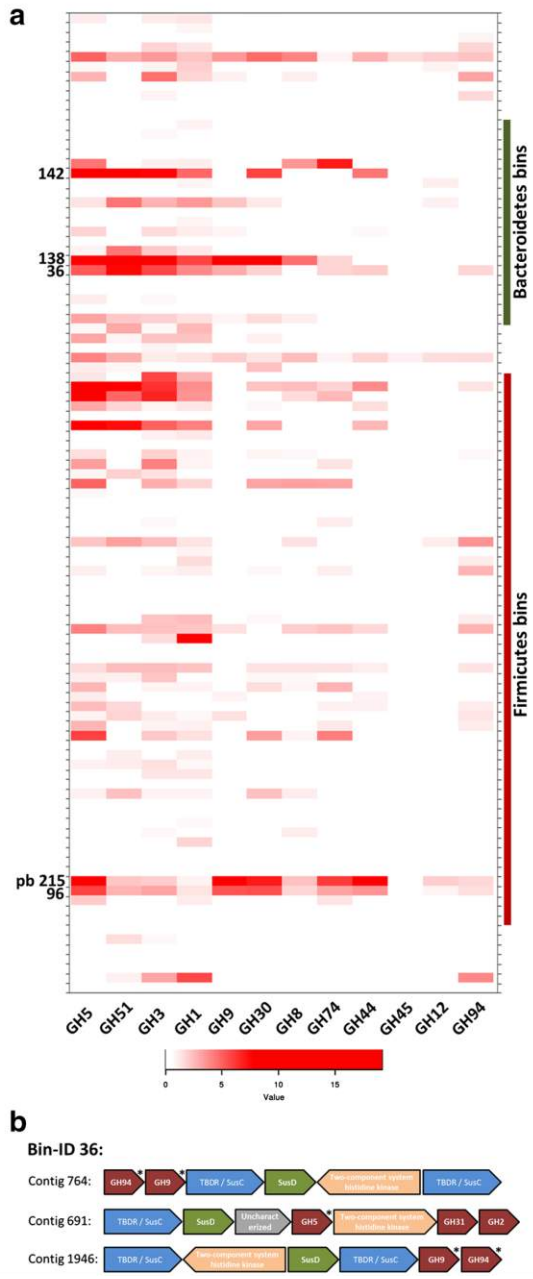

Fig. 5 **a** Heatmap reflecting the expression of cellulolytic GH families in the bacterial bins generated from the biogas fermenter metagenome. Expression strength is shown as square root of the rpkm value. Bin-IDs for selected bins affiliated with the *Bacteroidetes* and the *Firmicutes* are indicated. All other bin-IDs are given in Additional file 9 with a continuous labeling. **b** Physical map of three putative cellulolytic PULs. The PULs were identified on three contigs assigned to the bacteroidetal bin 36. TBDR = TonB-dependent receptor. An *asterisk* indicates GH families with cellulase (GH5, GH9) or cellobiase (GH94) activity

an equal distribution of these enzymes was observed in an elephant feces sample. Finally, we hypothesize that by finding ways to alter the ratio of the *Firmicutes* vs. the *Bacteroidetes* in favor of the *Bacteroidetes*, an increase in the overall hydrolytic performance of biogas plants might be achieved. This can potentially be realized by adding bacteriodetal isolates at high levels. However, it is likely that the added bacteroidetal organisms are quickly outcompeted again due to their better adaptation to natural habitats. To achieve a lasting increase in the abundance of the *Bacteroidetes* in biogas fermenters, the process conditions would have to be altered in a way that favors growth of this bacterial phylum. Consequently, further research is required to identify these conditions and factors, particularly as the microbiomes in natural systems are actively influenced and shaped by the host.

Additional files

Additional file 1. Compressed rar file containing the bins generated from the biogas fermenter metagenome, part 1 of 4.

Additional file 2. Compressed rar file containing the bins generated from the biogas fermenter metagenome, part 2 of 4.

Additional file 3. Compressed rar file containing the bins generated from the biogas fermenter metagenome, part 3 of 4.

Additional file 4. Compressed rar file containing the bins generated from the biogas fermenter metagenome, part 4 of 4.

Additional file 5: Table S1. Process parameters of the biogas fermenter at time of sampling. These data were kindly determined and provided by Bonalytic GmbH. **Table S2.** Diversity and richness of biogas fermenter samples taken in March and May 2013 based on 16S rRNA gene analysis. OTUs were clustered at 99% sequence similarity level and singleton OTUs (i.e. OTUs with only one sequence assigned to them) were removed. **Table S3.** Shared OTUs between both biogas fermenter samples. **Table S4.** Bin overview (tax. Assignment, completeness, contamination, other assembly stats). **Table S5.** Blastp results and MEGAN LCA assignement of 100 ORFs with the highest numbers of mapped reads in the biogas fermenter. **Table S6.** Results of an iterative protein sequence search using Jackhmmer (hmmer 3.1 package) to identify potential cellulosomal scaffoldin proteins in the metagenomic datasets using a score cutoff of 700.

Additional file 6. Additional TEM pictures of biogas fermenter content.

Additional file 7. Comparison of total cellulolytic activities in biogas fermenter content and feces samples of various herbivorous animals determined via 3,5-dinitrosalicylic acid (DNS) assay.

Additional file 8. Compressed rar file containing four krona plots. The krona plots illustrate the results of the 16S rRNA gene amplicon analysis conducted for the March and May biogas fermenter samples.

Additional file 9. High-resolution version of Fig. 5a with continuous labeling of all bacterial bins. Bin-IDs are color coded according to assigned phylum.

Abbreviations

CAZy: carbohydrate-active enzyme; CE: carbohydrate esterase; CMC: carboxymethycellulose; CTAB: cetyltrimethylammoniumbromid; DMSO: dimethylsulfoxid; DNS: 3,5-dinitrosalicylic acid; GH: glycoside hydrolase; HMM: hidden Markov model; LCA: lowest common ancestor; ORF: open reading frame; OTU: operational taxonomic unit; PULs: polysaccharide utilization loci; TEM: transmission electron microscopy.

Authors' contributions

SG conceived the study, extracted DNA and RNA, analyzed the assembled metagenome data sets, carried out the comparative analyses, and wrote the manuscript. MF and RAS provided a biogas fermenter sample and carried out an additional DNA extraction. DT, BN, and TR provided metagenomic assemblies, performed the metagenomic binning, and the analysis of the generated bins. NI provided an elephant feces sample and extracted RNA from this sample. RD and BW performed 16S amplicon sequencing and bioinformatic processing of amplicon raw data. MA and AG performed HiSeq sequencing, raw data processing, and provided metagenomic assemblies. WRS participated in the design of the study, coordinated the experiments, and also wrote and edited the manuscript. All authors read and approved the final manuscript.

Author details

[1] Department of Microbiology and Biotechnology, Biocenter Klein Flottbek, University of Hamburg, Ohnhorststr. 18, 22609 Hamburg, Germany. [2] Institute for General Microbiology, Christian Albrecht University Kiel, Kiel, Germany. [3] CUBE-Division for Computational Systems Biology, Dept. of Microbiology and Ecosystem Science, University of Vienna, Althanstr. 14, Vienna, Austria. [4] Institute of Microbiology and Genetics, Georg-August-University Göttingen, Grisebachstr. 8, Göttingen, Germany. [5] Bioinformatics Core, University Medical Center Hamburg-Eppendorf, Martinistr. 52, Hamburg, Germany. [6] Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Martinistr. 52, Hamburg, Germany.

Acknowledgements

We would like to thank Bioreact and Bonalytic GmbH (Troisdorf, Germany) for kindly providing the biogas plant fermenter samples and for the determination of the chemical process parameters provided in Additional file 5: Table S1. We further would like to thank the team of the Hagenbeck Zoo (Hamburg, Germany) for providing various feces samples.

Availability of supporting data

Within the framework of this study, generated raw sequence data have been deposited under the NCBI BioProject number PRJNA301928. In addition, an assembly of the biogas fermenter metagenome can be accessed and downloaded via IMG/ER (https://www.img.jgi.doe.gov) using the IMG ID 3300002898. The 104 metagenomic bins are provided in the compressed Additional files 1, 2, 3 and 4.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethical approval and consent to participate

Not applicable.

Funding

This research was funded by the German Federal Ministry of Education and Research (Grant number: 03SF0421H).

Received: 4 February 2016 Accepted: 26 May 2016
Published online: 07 June 2016

References

1. Weiland P. Biogas production: current state and perspectives. Appl Microbiol Biotechnol. 2010;85(4):849–60.
2. Wirth R, Kovacs E, Maroti G, Bagi Z, Rakhely G, Kovacs KL. Characterization of a biogas-producing microbial community by short-read next generation DNA sequencing. Biotechnol Biofuels. 2012;5:41.
3. Stolze Y, Zakrzewski M, Maus I, Eikmeyer F, Jaenicke S, Rottmann N, Siebner C, Puhler A, Schluter A. Comparative metagenomics of biogas-producing microbial communities from production-scale biogas plants operating under wet or dry fermentation conditions. Biotechnol Biofuels. 2015;8:14.

4. Zakrzewski M, Goesmann A, Jaenicke S, Junemann S, Eikmeyer F, Szczepanowski R, Al-Soud WA, Sorensen S, Puhler A, Schluter A. Profiling of the metabolically active community from a production-scale biogas plant by means of high-throughput metatranscriptome sequencing. J Biotechnol. 2012;158(4):248–58.
5. Schlüter A, Bekel T, Diaz NN, Dondrup M, Eichenlaub R, Gartemann KH, Krahn I, Krause L, Kromeke H, Kruse O, et al. The metagenome of a biogas-producing microbial community of a production-scale biogas plant fermenter analysed by the 454-pyrosequencing technology. J Biotechnol. 2008;136(1–2):77–90.
6. Li A, Chu Y, Wang X, Ren L, Yu J, Liu X, Yan J, Zhang L, Wu S, Li S. A pyrosequencing-based metagenomic study of methane-producing microbial community in solid-state biogas reactor. Biotechnol Biofuels. 2013;6(1):3.
7. Solli L, Havelsrud OE, Horn SJ, Rike AG. A metagenomic study of the microbial communities in four parallel biogas reactors. Biotechnol Biofuels. 2014;7(1):146.
8. Zverlov V, Köck D, Schwarz W. The role of cellulose-hydrolyzing bacteria in the production of biogas from plant biomass. In: Kamm B, editor. Microorganisms in biorefineries, vol. 26. Berlin: Springer; 2015. p. 335–61.
9. Fontes CM, Gilbert HJ. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu Rev Biochem. 2010;79:655–81.
10. Bayer EA, Lamed R, White BA, Flint HJ. From cellulosomes to cellulosomics. Chem Rec. 2008;8(6):364–77.
11. Bayer EA, Shimon LJ, Shoham Y, Lamed R. Cellulosomes-structure and ultrastructure. J Struct Biol. 1998;124(2–3):221–34.
12. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat Rev Microbiol. 2008;6(2):121–31.
13. He S, Ivanova N, Kirton E, Allgaier M, Bergin C, Scheffrahn RH, Kyrpides NC, Warnecke F, Tringe SG, Hugenholtz P. Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. PLoS One. 2013;8(4):e61126.
14. Henderson G, Cox F, Kittelmann S, Miri VH, Zethof M, Noel SJ, Waghorn GC, Janssen PH. Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. PLoS One. 2013;8(9):e74787.
15. Ilmberger N, Güllert S, Dannenberg J, Rabausch U, Torres J, Wemheuer B, Alawi M, Poehlein A, Chow J, Turaev D, et al. A comparative metagenome survey of the fecal microbiota of a breast- and a plant-fed Asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes. PLoS One. 2014;9(9):e106707.
16. Zeng B, Han S, Wang P, Wen B, Jian W, Guo W, Yu Z, Du D, Fu X, Kong F, et al. The bacterial communities associated with fecal types and body weight of rex rabbits. Sci Rep. 2015;5:9342.
17. Roggenbuck M, Sauer C, Poulsen M, Bertelsen MF, Sorensen SJ. The giraffe (*Giraffa camelopardalis*) rumen microbiome. FEMS Microbiol Ecol. 2014;90(1):237–46.
18. Pope PB, Mackenzie AK, Gregor I, Smith W, Sundset MA, McHardy AC, Morrison M, Eijsink VG. Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. PLoS One. 2012;7(6):e38571.
19. Morrison M, Pope PB, Denman SE, McSweeney CS. Plant biomass degradation by gut microbiomes: more of the same or something new? Curr Opin Biotechnol. 2009;20(3):358–63.
20. Mackenzie AK, Pope PB, Pedersen HL, Gupta R, Morrison M, Willats WG, Eijsink VG. Two SusD-like proteins encoded within a polysaccharide utilization locus of an uncultured ruminant Bacteroidetes phylotype bind strongly to cellulose. Appl Environ Microbiol. 2012;78(16):5935–7.
21. Naas AE, Mackenzie AK, Mravec J, Schuckel J, Willats WG, Eijsink VG, Pope PB. Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation? MBio. 2014;5(4):e01401–14.
22. Terrapon N, Lombard V, Gilbert HJ, Henrissat B. Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. Bioinformatics. 2015;31(5):647–55.
23. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science. 2011;331(6016):463–7.
24. Bian G, Ma L, Su Y, Zhu W. The microbial community in the feces of the white rhinoceros (*Ceratotherium simum*) as determined by barcoded pyrosequencing analysis. PLoS One. 2013;8(7):e70103.
25. Jaenicke S, Ander C, Bekel T, Bisdorf R, Droge M, Gartemann KH, Junemann S, Kaiser O, Krause L, Tille F, et al. Comparative and joint analysis of two metagenomic datasets from a biogas fermenter obtained by 454-pyrosequencing. PLoS One. 2011;6(1):e14519.
26. Weiland N, Loscher C, Metzger R, Schmitz R. Construction and screening of marine metagenomic libraries. Methods Mol Biol. 2010;668:51–65.
27. Wemheuer B, Taube R, Akyol P, Wemheuer F, Daniel R. Microbial diversity and biochemical potential encoded by thermal spring metagenomes derived from the Kamchatka Peninsula. Archaea. 2013;2013:136714.
28. Bragg L, Stone G, Imelfort M, Hugenholtz P, Tyson GW. Fast, accurate error-correction of amplicon pyrosequences using Acacia. Nat Methods. 2012;9(5):425–6.
29. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590–6.
30. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27(16):2194–200.
31. Schneider D, Arp G, Reimer A, Reitner J, Daniel R. Phylogenetic analysis of a microbialite-forming microbial mat from a hypersaline lake of the Kiritimati atoll, Central Pacific. PLoS One. 2013;8(6):e66662.
32. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335–6.
33. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
34. Peng Y, Leung HC, Yiu SM, Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics. 2012;28(11):1420–8.
35. Boisvert S, Raymond F, Godzaridis E, Laviolette F, Corbeil J. Ray meta: scalable de novo metagenome assembly and profiling. Genome Biol. 2012;13(12):R122.
36. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome project data processing S: the sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
37. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.
38. Mende DR, Waller AS, Sunagawa S, Jarvelin AI, Chan MM, Arumugam M, Raes J, Bork P. Assessment of metagenomic assembly using simulated next generation sequencing data. PLoS One. 2012;7(2):e31386.
39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
40. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res. 2007;17(3):377–86.
41. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. BMC Bioinform. 2011;12:385.
42. Wu M, Scott AJ. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics. 2012;28(7):1033–4.
43. Cantor M, Nordberg H, Smirnova T, Hess M, Tringe S, Dubchak I. Elviz—exploration of metagenome assemblies with an interactive visualization tool. BMC Bioinform. 2015;16:130.
44. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11(11):1144–6.
45. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
46. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043–55.
47. Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian HH, Coronado S, der Maaten L, Vlassis N, Wilmes P. VizBin—an application for reference-independent visualization and human-augmented binning of metagenomic data. Microbiome. 2015;3(1):1.
48. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic acids Res. 2014;42(Database issue):D490–5.
49. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinform. 2009;10:421.

50. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet. 2000;16(6):276–7.
51. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. 2010;11:119.
52. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic acids Res. 2012;40(Web Server issue):W445–51.
53. Berezikov E, van Tetering G, Verheul M, van de Belt J, van Laake L, Vos J, Verloop R, van de Wetering M, Guryev V, Takada S, et al. Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis. Genome Res. 2006;16(10):1289–98.
54. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27(6):863–4.
55. Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24):3211–7.
56. Anders S, Pyl PT, Huber W. HTSeq–a python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166–9.
57. Juergensen J, Ilmberger N, Streit WR. Screening for cellulases with industrial value and their use in biomass conversion. Methods Mol Biol. 2012;834:1–16.
58. Keating C, Cysneiros D, Mahony T, O'Flaherty V. The hydrolysis and biogas production of complex cellulosic substrates using three anaerobic biomass sources. Water Sci Technol. 2013;67(2):293–8.
59. Oakley BB, Lillehoj HS, Kogut MH, Kim WK, Maurer JJ, Pedroso A, Lee MD, Collett SR, Johnson TJ, Cox NA. The chicken gastrointestinal microbiome. FEMS Microbiol Lett. 2014;360(2):100–12.
60. Dowd SE, Callaway TR, Wolcott RD, Sun Y, McKeehan T, Hagevoort RG, Edrington TS. Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). BMC Microbiol. 2008;8:125.
61. Clauss M, Frey R, Kiefer B, Lechner-Doll M, Loehlein W, Polster C, Rossner GE, Streich WJ. The maximum attainable body size of herbivorous mammals: morphophysiological constraints on foregut, and adaptations of hindgut fermenters. Oecologia. 2003;136(1):14–27.
62. Wong DW. Feruloyl esterase: a key enzyme in biomass degradation. Appl Biochem Biotechnol. 2006;133(2):87–112.
63. Zhang J, Siika-Aho M, Tenkanen M, Viikari L. The role of acetyl xylan esterase in the solubilization of xylan and enzymatic hydrolysis of wheat straw and giant reed. Biotechnol Biofuels. 2011;4(1):60.
64. Koeck DE, Ludwig W, Wanner G, Zverlov VV, Liebl W, Schwarz WH. Herbinix hemicellulosilytica gen. nov., sp. nov., a thermophilic cellulose-degrading bacterium isolated from a thermophilic biogas reactor. Int J Syst Evol Microbiol. 2015;65(8):2365–71.
65. Koeck DE, Maus I, Wibberg D, Winkler A, Zverlov VV, Liebl W, Puhler A, Schwarz WH, Schluter A. Draft genome sequence of Herbinix hemicellulosilytica T3/55 T, a new thermophilic cellulose degrading bacterium isolated from a thermophilic biogas reactor. J Biotechnol. 2015;214:59–60.
66. Liu F, Conrad R. Chemolithotrophic acetogenic H2/CO2 utilization in Italian rice field soil. ISME J. 2011;5(9):1526–39.
67. Daniel WR, Dwayne AE, Melanie RM. Metabolic capabilities of the members of the order Halanaerobiales and their potential biotechnological applications. Curr Biotechnol. 2014;3(1):3–9.
68. Nettmann E, Bergmann I, Pramschufer S, Mundt K, Plogsties V, Herrmann C, Klocke M. Polyphasic analyses of methanogenic archaeal communities in agricultural biogas plants. Appl Environ Microbiol. 2010;76(8):2540–8.
69. Bergmann I, Nettmann E, Mundt K, Klocke M. Determination of methanogenic Archaea abundance in a mesophilic biogas plant based on 16S rRNA gene sequence analysis. Can J Microbiol. 2010;56(5):440–4.
70. Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, Abbott DW, Henrissat B, Gilbert HJ, Bolam DN, et al. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS Biol. 2011;9(12):e1001221.

4 A Comparative Metagenome Survey of the Fecal Microbiota of a Breast- and a Plant-Fed Asian Elephant Reveals an Unexpectedly High Diversity of Glycoside Hydrolase Family Enzymes

Nele Ilmberger[1], Simon Güllert[1], Joana Dannenberg[1], Ulrich Rabausch[1], Jeremy Torres[1], Bernd Wemheuer[2], Malik Alawi[3], Anja Poehlein[2], Jennifer Chow[1], Dimitrij Turaev[4], Thomas Rattei[4], Christel Schmeisser[1], Jesper Salomon[5], Peter B. Olsen[5], Rolf Daniel[2], Adam Grundhoff[6], Martin S. Borchert[5] and Wolfgang R. Streit[1]

[1]Universität Hamburg, Biozentrum Klein Flottbek, Abteilung für Mikrobiologie & Biotechnologie, Hamburg, Germany; [2]Georg-August-University Göttingen, Institute of Microbiology and Genetics, Gö ttingen, Germany; [3]University Medical Center Hamburg-Eppendorf, Bioinformatics Service Facility, Hamburg, Germany; [4]University of Vienna, CUBE - Division for Computational Systems Biology, Department of Microbiology and Ecosystem Science, Vienna, Austria; [5]Novozymes A/S, Microbial Discovery, Bagsværd, Denmark; [6]Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Hamburg, Germany

Published in:

PLoS ONE 9(9): e106707. doi:10.1371/journal.pone.0106707

OPEN ACCESS Freely available online

PLOS ONE

A Comparative Metagenome Survey of the Fecal Microbiota of a Breast- and a Plant-Fed Asian Elephant Reveals an Unexpectedly High Diversity of Glycoside Hydrolase Family Enzymes

Nele Ilmberger[1], Simon Güllert[1], Joana Dannenberg[1], Ulrich Rabausch[1], Jeremy Torres[1], Bernd Wemheuer[2], Malik Alawi[3], Anja Poehlein[2], Jennifer Chow[1], Dimitrij Turaev[4], Thomas Rattei[4], Christel Schmeisser[1], Jesper Salomon[5], Peter B. Olsen[5], Rolf Daniel[2], Adam Grundhoff[6], Martin S. Borchert[5], Wolfgang R. Streit[1]*

1 Universität Hamburg, Biozentrum Klein Flottbek, Abteilung für Mikrobiologie & Biotechnologie, Hamburg, Germany, **2** Georg-August-University Göttingen, Institute of Microbiology and Genetics, Göttingen, Germany, **3** University Medical Center Hamburg-Eppendorf, Bioinformatics Service Facility, Hamburg, Germany, **4** University of Vienna, CUBE - Division for Computational Systems Biology, Department of Microbiology and Ecosystem Science, Vienna, Austria, **5** Novozymes A/S, Microbial Discovery, Bagsværd, Denmark, **6** Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Hamburg, Germany

Abstract

A phylogenetic and metagenomic study of elephant feces samples (derived from a three-weeks-old and a six-years-old Asian elephant) was conducted in order to describe the microbiota inhabiting this large land-living animal. The microbial diversity was examined via 16S rRNA gene analysis. We generated more than 44,000 GS-FLX+454 reads for each animal. For the baby elephant, 380 operational taxonomic units (OTUs) were identified at 97% sequence identity level; in the six-years-old animal, close to 3,000 OTUs were identified, suggesting high microbial diversity in the older animal. In both animals most OTUs belonged to Bacteroidetes and Firmicutes. Additionally, for the baby elephant a high number of Proteobacteria was detected. A metagenomic sequencing approach using Illumina technology resulted in the generation of 1.1 Gbp assembled DNA in contigs with a maximum size of 0.6 Mbp. A KEGG pathway analysis suggested high metabolic diversity regarding the use of polymers and aromatic and non-aromatic compounds. In line with the high phylogenetic diversity, a surprising and not previously described biodiversity of glycoside hydrolase (GH) genes was found. Enzymes of 84 GH families were detected. Polysaccharide utilization loci (PULs), which are found in Bacteroidetes, were highly abundant in the dataset; some of these comprised cellulase genes. Furthermore the highest coverage for GH5 and GH9 family enzymes was detected for Bacteroidetes, suggesting that bacteria of this phylum are mainly responsible for the degradation of cellulose in the Asian elephant. Altogether, this study delivers insight into the biomass conversion by one of the largest plant-fed and land-living animals.

Citation: Ilmberger N, Güllert S, Dannenberg J, Rabausch U, Torres J, et al. (2014) A Comparative Metagenome Survey of the Fecal Microbiota of a Breast- and a Plant-Fed Asian Elephant Reveals an Unexpectedly High Diversity of Glycoside Hydrolase Family Enzymes. PLoS ONE 9(9): e106707. doi:10.1371/journal.pone.0106707

Editor: Stefan Bereswill, Charité-University Medicine Berlin, Germany

Received April 17, 2014; **Accepted** July 29, 2014; **Published** September 10, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. The data will be available upon publication. This project has been deposited at GenBank (www.ncbi.nlm.nih.gov/genbank) using the BioProject number PRJNA240141. The sequences derived from Illumina and 454 sequencing were deposited in the NCBI Short Read Archive (www.ncbi.nlm.nih.gov/sra) under the study accession number SRP040073. Genome assembly together with predicted gene models and annotation is available from www.jgi.doe.gov. (DOE Joint Genome Institute) under the IMG Project Id 50566.

Funding: This work was in part funded by the ministry of education and research "Bundesministerium für Bildung und Forschung" (BMBF; http://www.bmbf.de/) within the network program ExpresSys and the grant FKZ 0315586F to WRS. JT was funded by the German Academic Exchange program (DAAD; https://www.daad.de/de/index.html). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Some of the authors are affiliated with Novozymes A/S. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

* Email: wolfgang.streit@uni-hamburg.de

Introduction

The microbiota of mammalian intestines is the main driver of plant cell wall degradation as mammalian genomes do not encode for a significant number of genes linked to structural polysaccharide, i.e. cellulose, degradation [1]. Thereby herbivores can gain 70% of their energy from microbial polysaccharide breakdown [1–3]. Within the herbivores two groups are distinguished corresponding to the location of the main fermentation: the foregut fermenters (mainly ruminants) and the hindgut fermenters. In foregut fermenters the fermentation takes place in pregastric chambers, the rumen. In hindgut fermenters the main fermentation chamber is the colon or the caecum (e.g. elephants, their caecum is up to 1.5 m long) and therein the dry matter content is significantly higher than in the rumen [4]. In addition, ruminants grind the lignocellulosic material mechanically by chewing regularly during the fermentation process. Ruminants therefore have a very effective way of digesting lignocellulose. Hindgut

fermenters in contrast are able to digest faster; this is an advantage at high body size as more food can be ingested [5]. Recently, the intestinal or fecal microbiota of different animals has been investigated using next generation sequencing (NGS) technology. Among them were cow, reindeer, wallaby, yak, giant panda, buffalo, swine, Iberian lynx and termite [6–16]. These studies delivered a significant amount of novel sequence data giving insight into the phylogeny, the metabolism and the genetic potential of the intestinal microbiota. Furthermore, these studies have led to a better understanding of fecal and gut microbial communities and they have underlined the importance of these communities for host survival, fitness, physiology and nutrient utilization [17–19]. In addition these studies have suggested a core and a variable gene set, influenced by host traits, environment, type of diet and other not yet identified factors [20–22]. In general and as expected, the herbivorous microbiomes encoded for high numbers of carbohydrate active enzymes (CAZymes). For many of the studied systems Clostridia were most abundant and were identified as main cellulose degraders [11,23]. Interestingly, in termites Spirochaetes and Fibrobacteres are supposed to be the main contributors to cellulose digestion [16]. For these microbes the absence of cellulosomes and the release of at least some of the cellulases in the ruminal fluid were proposed [16]. Furthermore, some Bacteroidetes have been described as being highly abundant in different fecal/intestinal samples [6,8–10]. Recent metagenomic and genomic studies have highlighted the presence of polysaccharide utilization loci (PULs) from Bacteroidetes that include cellulase genes [8,10]. PULs were primarily described as starch degradation operon composed of genes encoding proteins designated SusA to SusG. These proteins coordinately cleave carbohydrates and transport sugar into the cell [1]. PULs have been identified in different gut samples and genomes of *Bacteroides* sp. [3,8,24–26].

Adult Asian elephants can reach a weight of 5,000 kg and consume about 150 kg plant materials per day. The nourishment is composed of highly fibrous plant material, mainly grass, fruits, leaves, twigs, roots and bark. As our understanding of the microbiota inhabiting elephants is poor, we were interested in characterizing the microbial population in the feces of this large herbivorous animal. Within the current paper we deliver evidence that the microbiome of Asian elephants is highly diverse and that microbes of the Bacteroidetes phylum are presumably the main cellulose degraders. Thereby we show that elephants are generalists rather than specialists with respect to the degradation of plant biomass.

Materials and Methods

DNA isolation

No specific permissions were required for these activities. The study did not involve endangered or protected species. Coordinates are: Longitude/Latitude 9.941572/9.941572. Fresh feces samples derived from a six-years-old female Asian elephant ("Kandy", *2003 in Hamburg) and from a three-weeks-old male elephant ("Assam", *2012 in Hamburg), both living in the zoo 'Hagenbecks Tierpark' in Hamburg (Germany), were collected from the zoo staff directly after defecation. The samples were transported, directly and on ice, to the laboratory in Hamburg Klein Flottbek for further analysis. The elephants were not treated with antibiotics. The older Asian elephant was mainly fed with grass, hay, leaves and twigs, with additional fruits and vegetables. In contrast, the three-weeks-old male elephant was breast-fed. DNA isolation was performed with the QIAamp DNA Stool kit from Qiagen (Hilden, Germany) as described previously [27].

Metagenome sequencing

Libraries were prepared with the NEB DNA Ultra Kit following the manufacturer's protocol. Illumina sequencing for the six-years-old elephant was performed using a HiSeq 2000 instrument (1.5 lanes, paired-end run (2×100 bases)). For the three-weeks-old animal a HiSeq 2500 instrument (one lane, paired-end run (2×100 bases)) was used for sequencing at the HPI in Hamburg. *De novo* assembly was performed with the Velvet assembly program version 1.2.08 [28]. For the investigation of the sequences the IMG server (https://img.jgi.doe.gov/cgi-bin/mer/main.cgi) was used. To further analyze the possible biological processes linked to the individual genes and ORFs mainly the KEGG [29], the COG [30] and the Pfam [31] databases were employed using a cut off of 10^{-5}.

Amplification and sequencing of 16S rRNA genes

To assess the microbial diversity, variable regions of the 16S rRNA genes were amplified as previously published [32] but with minor modifications.

The V3–V5 region was amplified using the following primer set: V3for 5′-TCTC ATCCCTGCGTGTCTCCGACTCAGACGCTCGACACCTACGGGNGGCWGCAG-3′ and V5rev 5′-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGCCGTCAATTCMTTTRAGTTT-3′. The primers contained Roche 454 pyrosequencing adaptors, keys and one unique MID per sample (underlined). To assess the archaeal diversity, the V4–V6 region was amplified using the primer set: A519F 5′-CCATCTCATCCCTGCGTGTCTCCGACT CAGATATCGGCGAGCAGCMGCCGCGGAA- 3′ and A1041R 5′-CCTATCCCCTGTGT GCCTTGGCAGTCTCAGGGCCATGCACCWCCTCTC-3′. The PCR reaction (50 µl) contained 0.5 U of Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Germany), 10 µl 5x Phusion GC Buffer, 200 µM of each *d*NTP, 2.5% DMSO, 1.5 mM $MgCl_2$, 4 µM of each primer, and 20 ng isolated DNA. PCR cycling conditions were: initial denaturation at 98°C for 3 min, followed by 28 cycles of denaturation at 98°C for 30 s, annealing at 61°C for 30 s (archaeal primer set: 66°C), and extension at 72°C for 25 s. The final extension was conducted at 72°C for 5 min. Negative controls were performed with H_2O instead of template DNA. The obtained PCR products were purified via Gel/PCR DNA Fragments Extraction Kit (Geneaid Biotech, Taiwan) as recommended by the manufacturer. Three separate PCR reactions were conducted for each sample. After gel extraction, the reaction products were pooled in equal amounts. The 16S rRNA gene sequencing was performed at the Göttingen Genomics Laboratory using a Roche GS-FLX+454 pyrosequencer and titanium chemistry (Roche, Branford, USA).

Processing and analysis of 16S rRNA gene data sets

Via pyrosequencing generated raw sequences were processed according to [33], with the following modifications: After raw data extraction, reads shorter than 300 bp and those possessing long homopolymer stretches (>8 bp) or primer mismatches (>5 bp) were removed. The sequences were denoised employing Acacia [34]. Chimeric sequences were removed using UCHIME in reference mode with the most recent SILVA SSU database as reference dataset (SSURef 115 NR) [35]. The processed 16S rRNA gene sequences were uploaded to the SILVA NGS (SILVA next-generation sequencing) server for taxonomic classification [36]. Microbial taxonomy was determined using default settings with two adjustments: The cluster sequence identity threshold was decreased to 0.97 and the maximal taxonomic depth was increased to 30. Rarefaction curves were calculated employing the QIIME 1.8 software package [37].

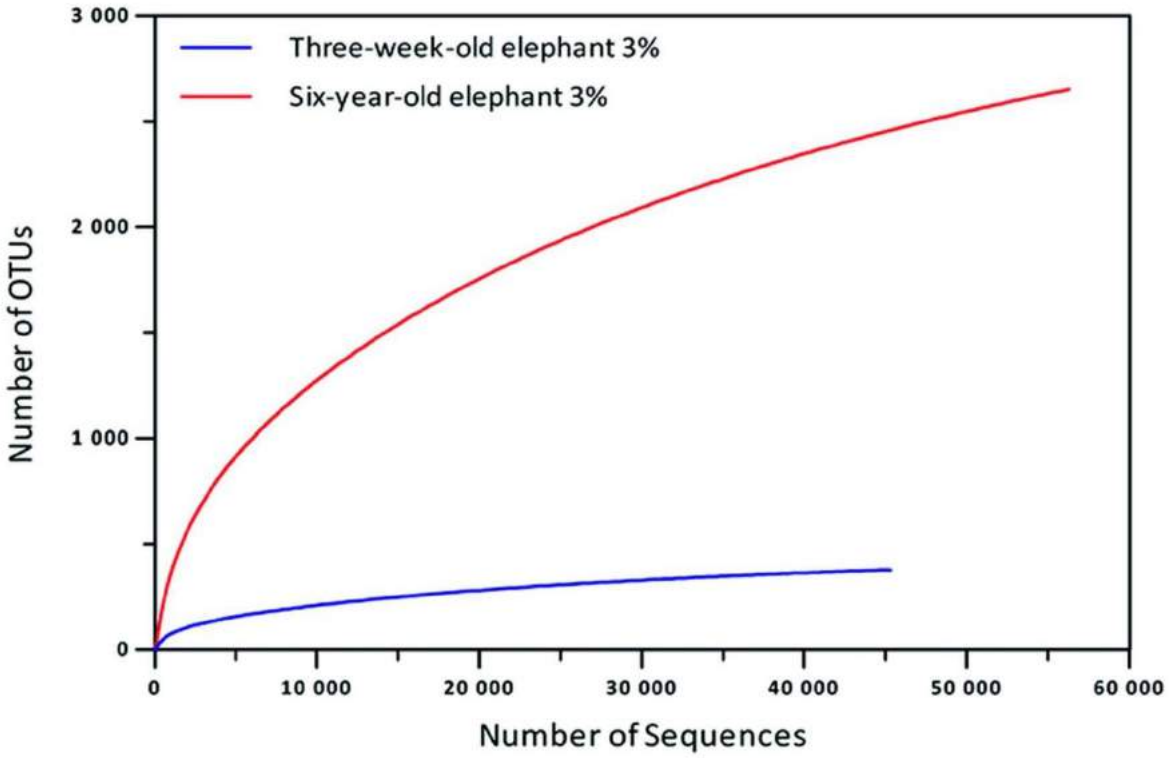

Figure 1. Rarefaction curves calculated for the feces sample of the three-weeks-old and the six-years-old Asian elephant at 3% genetic distance of 16S rRNA genes. The curve for the six-years-old elephant comprises 8,014 archaeal sequences which were clustered to 54 OTUs. The sequences were denoised employing Acacia. Chimeric sequences were removed using UCHIME in reference mode with the most recent SILVA SSU database as reference dataset (SSURef 115 NR).
doi:10.1371/journal.pone.0106707.g001

Table 1. Overall numbers of sequences and contigs generated for the two elephant feces samples.

Parameter	Six-years-old elephant	Three-weeks-old elephant
16S rRNA gene sequences		
No. of GS-FLX+454 sequences	56,124	44,508
Feces Microbiome DNA		
No. of Illumina reads	415,784,264	200,946,703
No. of bases assembled	929,519,943	138,494,152
No. of assembled sequences	260,535	50,765
Coverage	42	73
Mean contig size (bp)	3,566	5,801
N50 (bp)	5,751	22,052
Largest contig (bp)	344,979	597,113
Gene count	1,068,385	173,134
GC mean %	43,96	47,71
Protein coding genes	1,052,245	171,134
with COG	629,485	111,920
with Pfam	818,373	136,417
with KEGG	210,308	42,640

doi:10.1371/journal.pone.0106707.t001

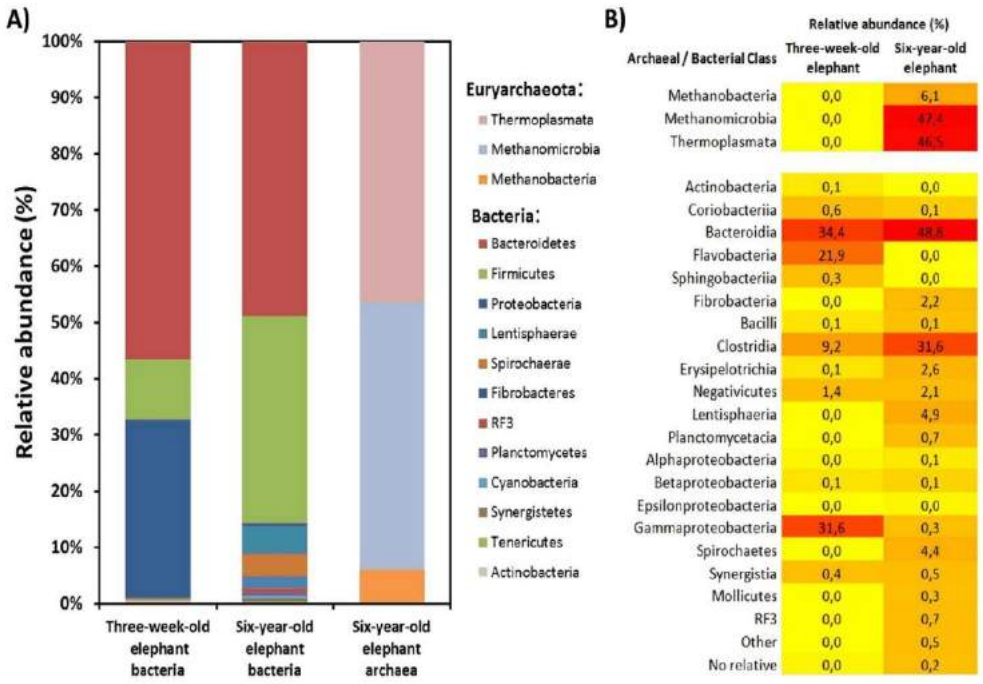

Figure 2. Relative abundances of different phyla and classes in the two elephant feces samples. A: Relative abundance of phyla in the feces of the three-weeks-old and the six-years-old Asian elephant based on 16S rRNA gene sequences. For the three-weeks-old elephant no Archaea were observed. **B:** Phylogenetic comparison on class level between both elephants. Heat map colors indicate the abundances of the respective 16S rRNA genes.
doi:10.1371/journal.pone.0106707.g002

Comparison with other fecal metagenomes and statistical analysis

An additional set of metagenomes of animals' microbiotas was used for comparative analysis by the IMG/M ER webpage of the US Department of Energy Joint Genome Institute (https://img.jgi.doc.gov/cgi-bin/mer/main.cgi) [38]. The phylogenetic distribution of the metagenomes was analyzed. A phylogenetic tree was automatically created with the public Metagenomics RAST server [39].

Binning

The assembly of both metagenomic samples was performed with the Ray Meta assembler [40], version 2.3.1, using a k-mer length of 31 and default parameters. Scaffolds of ≥2 kb length and a mode k-mer coverage depth of >5 were binned based on unsupervised tetranucleotide frequencies using MetaWatt2.0 [41]. To assess bin taxonomy, homogeneity and completeness we identified bacterial and archaeal phylogenetic marker genes using AMPHORA2 [42]. Using project-specific scripts we estimated the most recent taxonomic level supported by most high-confidence AMPHORA2 markers (confidence score ≥0.9) for each bin. At these levels the phylogenetic markers determined the taxonomic affiliation and completeness of each bin. The tRNA genes, rRNA genes and protein-coding genes were predicted in each bin using RNAmmer [43], tRNAscan-SE [44] and GeneMarkS [45], respectively.

Transmission electron microscopy (TEM)

Slices were prepared with the microtome Reichert-Jung Ultracut E. Fixation was performed in 2% glutaraldehyde in 75 mM cacodylate buffer (pH 7.0). After washing it was supplied with 2% agar in 75 mM cacodylate buffer (pH 7.0) and further fixed with 1% OsO_4 in 50 mM cacodylate buffer (pH 7.0). After washing with 75 mM cacodylate buffer (pH 7.0) water was removed with acetone and the sample was infiltrated with Spurr resin (Polysciences, Warrington, PA, USA). TEM pictures were observed with the LEO 906 E, the camera Gatan 794 and the software Digital micrograph.

DNA sequences obtained and GenBank submissions

This project has been deposited at GenBank using the BioProject number PRJNA240141. The sequences derived from Illumina and 454 sequencing were deposited in the NCBI Short Read Archive, the study accession number is SRP040073. Assembled sequence data with predicted gene models and annotation is available from www.jgi.doc.gov. (DOE Joint Genome Institute), the IMG Project Id is 50566.

Results and Discussion

Population structure of the elephant feces samples

Recent metagenomic research has demonstrated a close correlation between host diet and intestinal microbiome and it has highlighted the metabolic diversity within intestinal and fecal samples of mammals and insects [6,8,9,11,15,16,46]. These intriguing findings inspired us to analyze the fecal microbiota of

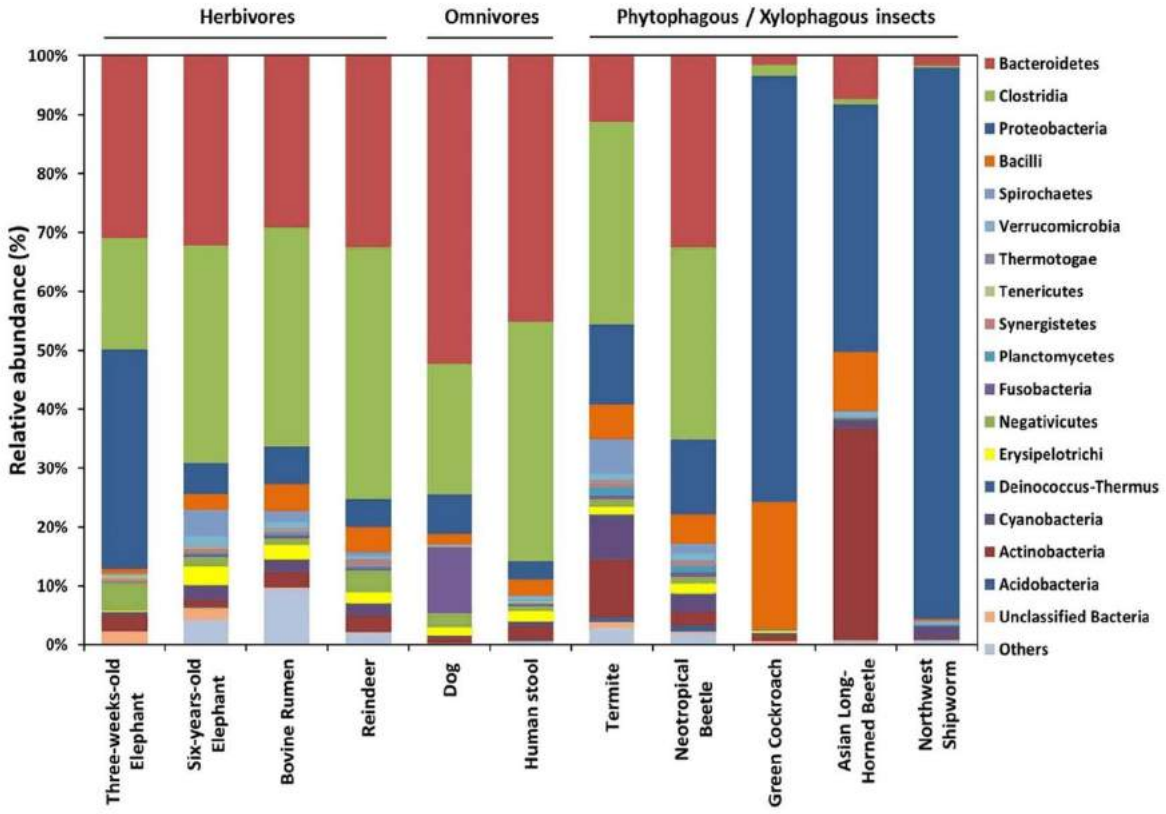

Figure 3. Phylogenetic analysis of the elephant feces in comparison with other fecal and intestinal metagenome data sets. Data indicate the phylogenetic relation based on gene similarities in the metagenome sequences. The percent of sequences assigned to each phylum according to IMG/M ER is shown based on the total number of obtained sequences of each data set. Sequence data for the metagenomes were extracted from the IMG/M ER web page of the US Department of Energy Joint Genome Institute and the respective bioprojects (IMG Genome IDs: Six-years-old Elephant (this study): 3300001598; three-weeks-old Elephant (this study): 3300001919; Green Cockroach: 2228664000; Termite: 3300001544; Dog: 2019105001; Reindeer: 2088090000; Neotropical Beetle: 3300000114; Asian Long-Horned Beetle: 2084038013; Bovine Rumen: 2061766007; Northwest Shipworm: 2189573029; Human stool: 7000000038).
doi:10.1371/journal.pone.0106707.g003

one of the largest land-living herbivores (only the African elephant is larger than the Asian elephant), which is known to be an exception in the world of large foregut fermenters as digesting faster than the high body size would allow [5]. To estimate the diversity of the microbes and their individual tasks that meet the energy need of elephants, a detailed phylogenetic analysis using NGS technologies was performed. Thereby we analyzed the community of a plant-fed female specimen (six-years-old) and of a three-weeks-old breast-fed male animal. Both animals were living in a local zoo and they were not treated with antibiotics. For the baby elephant 44,508 partial 16S rRNA sequences with an average length of 536 bp were generated and 56,124 sequences with an average length of 523 bp for the six-years-old elephant (TABLE 1). Rarefaction curves suggested that both datasets had the required amount of sequence data to assess operational taxonomic unit (OTU) richness and that both datasets had a similar coverage of biodiversity (FIGURE 1).

The feces sample of the six-years-old elephant revealed an almost 10-fold higher diversity than that of the baby animal (FIGURE 1, 2). For the three-weeks-old animal the phylogenetic analysis suggested the presence of approximately 380 OTUs based on a 97% sequence identity cut off for bacterial 16S rRNA genes. However, in the feces of the older animal 2,656 OTUs (calculated max. 3,487 OTUs) were identified (FIGURE 1, TABLE S1). The Shannon index for the baby elephant was 4.3 and for the six-years-old elephant it was 8.4. Furthermore the Chao1 index for species richness was 449 for the baby and 3,281 for the six-years-old elephant (TABLE S1), leading to the overall conclusion that the microbiome of the older elephant was much more diverse than that of the baby.

The baby elephant's feces sample was dominated by bacteria belonging to the Bacteroidetes (FIGURE 2). More than 50% of the microbes were Flavobacteriales (22%) or Bacteroidales (34%). The sample of the baby elephant furthermore comprised a high number of Proteobacteria (32%) and Firmicutes (11%) (FIGURE 2). On genus level most abundant bacteria were *Myroides* (22%), the S24-7 group of Bacteroidetes (14%) and *Pseudomomas* and *Psychrobacter* species (12% each) (supporting information S1). With the archaeal primer set no PCR product was observed.

The bacterial population of the six-years-old animal was dominated by Firmicutes (36%) and Bacteroidetes (47%). Furthermore Spirochaetes (4%), Fibrobacteres (2%) and Lentisphaeria (5%) were detected. Interestingly no bacteria belonging to these phyla were detected in the three-weeks-old elephant's

Table 2. Selected and recent metagenome studies published from insect or mammalian fecal and gut samples.

Microbiome	OTUs detected	Assemble DNA (Mbp)	Sequencing strategy	No. of GH families identified	Total no. of CAZymes	No. of GH per Mbp assembled DNA	[f]Total no. of cellulolytic enzymes	Reference
Iberian lynx	n.d.	18.3	454 GS FLX	42	372	20.3	39	[15]
Tammar wallaby	236	85.2	454 GS FLX	53	557	6.5	175	[9]
Reindeer[a]	1,182	26	454 GS FLX	30	[d]5,000	[e]9.9	[d]400	[8]
Giant panda	85	37	Illumina	44	440	11.9	124	[11]
Bovine	1,000	[b]1,930	Illumina	60	27,755	14.4	5,670	[36]
Bovine	161–259	[c]104	454 GS20	35	3,828	[c]36.8	1,017	[7]
Termite	216	71	454 GS20+ Sanger	45	703	9.9	259	[16]
Elephant gut (six-years-old)	2,656	929.5	Illumina	82	11,038	11.9	2,074	This study
Elephant gut (three-weeks-old)	380	138.5	Illumina	70	1,873	13.5	242	This study

n.d., not determined; a), in 2 samples; b), 286 Gb resulted in approx. 2 Gb of assembled DNAs >1 kb scaffolds; c),unassembled; d), in unassembled reads; e), per 503 Mbp unassembled DNA; f), total gene count with respect to the GH1, GH3, GH 5, GH6, GH8, GH9, GH44, GH45, GH48, GH51, GH74, GH94 family enzymes. Total gene counts, size of assembled DNAs, OTUs and numbers of carbohydrate active enzymes were extracted from the indicated references.

doi:10.1371/journal.pone.0106707.t002

sample. On genus level uncultured *Lachnospiraceae* (10%), *Ruminococcaceae* (10%), *Prevotellaceae* (11%) and the RC9 gut group of *Rickenellaceae* (13%) dominated (supporting information S2). Archaea were found in this sample. About 50% of these belonged to the Methanomicrobiales and the remainder belonged mostly to the Thermoplasmatales, i.e. Candidatus *Methanomethylophilus* (FIGURE 2). The respective organisms most likely contribute to elephant's methane production during feed fermentation [47].

To further verify these data, we analyzed the sequences obtained by Illumina sequencing for the presence of rRNA gene fragments or complete genes. 102 16S rRNA gene fragments were found for the baby elephant and 338 16S rRNA gene fragments for the six-years-old elephant. The results from this analysis largely confirmed the results of the 16S rRNA amplicon sequencing. Additionally, binning was performed with the metagenomic sequence data. For the baby elephant 234 bins were observed, for the six-years-old elephant 1,401. Thereof 131, or 935, respectively, were unassigned (TABLE S2, S3). The results largely confirm the 16S rRNA analysis, the higher microbial richness of the six-years-old elephant and the phylogenetic affiliations. For the older elephant Firmicutes and Bacteroidetes were dominant with 39 and 41% of the bins, respectively. Spirochaetes and Fibrobacteres were present at lower levels and 6% of the bins were assigned to Archaea (TABLE S3). For the baby elephant Firmicutes (35%), Bacteroidetes (32%) and Proteobacteria (28%) were dominant (TABLE S2). This is, in comparison with the 16S rRNA analysis, an overrepresentation of Firmicutes and an underrepresentation of Bacteroidetes. This difference nevertheless can easily be explained by the different methods.

Overall and as expected the microbiota of the six-years-old elephant resembled those of most herbivorous mammals like reindeer, wallaby and cow [6,8,9] with Firmicutes, Bacteroidetes and Proteobacteria being dominant (FIGURE 3). Surprisingly, the number of OTUs was at least two-fold higher than those of other herbivores (FIGURE 3; TABLE 2 and references given herein). Since Asian elephants have a rather diverse diet in the zoo compared to wild living animals this relatively high bacterial diversity may reflect an adaptation to this feeding manner. Furthermore the nourishment is composed of different plants and the microbiome is limited to the small zoo population. Thus, the results altogether do not represent the microbiome of wild living animals. It has been shown e.g. for pandas that the communities of wild and captive animals differ greatly [11].

As the baby elephant was breast-fed, the microbes inhabiting the intestine of this animal were in part determined by the mother's milk and of those bacteria attached to the mother transmitted by direct contact [48]. We speculate that especially the relatively high level of Proteobacteria was a result of the breast-feeding (FIGURE 3). This observation fitted well with the fact that in human breast milk more than 60% of the bacteria were Proteobacteria [49]. While the observed differences in OTU richness of the microbial communities of both elephants can most likely be explained by the differing nourishment, it should be noted that baby elephants also eat feces from older animals. Because of this it is likely that a significant fraction of the microbes in its feces originated from other animals from the herd.

Nevertheless, only 13 OTUs were present in both elephants' microbiomes. These were seven Firmicutes, three Bacteroidetes and three Proteobacteria (TABLE S4). Interestingly, most of these OTUs were considerably higher represented in one of the samples (TABLE S4). Altogether the two communities are distinct with only few overlaps. Though of course analyzing only two

Table 3. Gene count, relative gene count and relative coverage of genes for families of carbohydrate-active enzymes discovered in the sequences of the elephant feces samples according to CAZy.

	Baby elephant			Six-years-old elephant		
[1)]GH family	Gene count	Relative count	Relative coverage	Gene count	Relative count	Relative coverage
1	33	0.01761	0.00979	103	0.00933	0.00667
2	220	0.11745	0.09455	917	0.08311	0.082006
3	148	0.07901	0.08355	804	0.07287	0.07784
4	16	0.00854	0.00430	57	0.00517	0.00405
5	19	0.0101	0.01064	517	0.04686	0.04808
8	4	0.00213	0.00134	85	0.00770	0.01198
9	9	0.00481	0.00406	119	0.01078	0.01199
10	6	0.00320	0.00836	258	0.02338	0.02915
11	–	–	–	20	0.00181	0.00142
13	124	0.06620	0.07574	845	0.07658	0.07423
15	1	0.00053	0.00019	1	0.00009	0.00008
16	22	0.01174	0.01399	198	0.01794	0.01552
17	–	–	–	5	0.00045	0.00022
18	51	0.02723	0.03378	86	0.00779	0.00824
19	1	0.00053	0.00150	3	0.00027	0.00011
20	108	0.05766	0.05088	266	0.02411	0.01881
23	69	0.03684	0.06012	294	0.02664	0.03324
24	12	0.00641	0.00685	26	0.00236	0.00191
25	30	0.01602	0.03030	229	0.02075	0.02255
26	8	0.00427	0.00341	103	0.00933	0.00931
27	12	0.00641	0.00591	158	0.01432	0.01366
28	26	0.01388	0.01502	242	0.02193	0.02112
29	79	0.04218	0.04032	376	0.03408	0.02950
30	17	0.00908	0.00727	95	0.00861	0.00975
31	35	0.01869	0.01000	318	0.02882	0.02962
32	21	0.01121	0.00879	119	0.01078	0.01409
33	31	0.01655	0.01639	149	0.01350	0.01003
35	36	0.01922	0.01745	123	0.01115	0.01155
36	32	0.01708	0.01637	310	0.02809	0.02396
37	3	0.00160	0.00112	17	0.00154	0.00144
38	12	0.00641	0.00653	81	0.00734	0.00645
39	3	0.00160	0.00374	89	0.00807	0.00554
42	9	0.00481	0.00262	37	0.00335	0.00247
43	98	0.05232	0.03978	894	0.08102	0.09836
44	–	–	–	7	0.00063	0.00116
45	–	–	–	7	0.00063	0.00084
46	–	–	–	1	0.00009	0.00004
50	2	0.00107	0.00343	26	0.00236	0.00184
51	24	0.01281	0.00653	239	0.02166	0.02139
53	6	0.00320	0.00632	88	0.00798	0.00976
54	–	–	–	13	0.00118	0.00136
55	2	0.00107	0.00030	8	0.00073	0.00052
57	11	0.00587	0.00893	115	0.01042	0.00941
63	7	0.00374	0.00434	57	0.00517	0.00318
64	–	–	–	2	0.00018	0.00008
65	7	0.00374	0.00574	39	0.00353	0.00285
66	3	0.00160	0.00068	10	0.00091	0.00069
70	–	–	–	1	0.00009	0.00002

Table 3. Cont.

	Baby elephant			Six-years-old elephant		
[1)]GH family	Gene count	Relative count	Relative coverage	Gene count	Relative count	Relative coverage
73	36	0.01922	0.01851	202	0.01831	0.01721
74	–	–	–	11	0.00100	0.00084
76	12	0.00641	0.00510	9	0.00082	0.00081
77	28	0.01495	0.01611	193	0.01749	0.01614
78	41	0.02189	0.03159	413	0.03743	0.02897
79	3	0.00160	0.00264	–	–	–
81	–	–	–	3	0.00027	0.00009
84	9	0.00481	0.00250	19	0.00172	0.00165
88	33	0.01762	0.01190	54	0.00489	0.00342
92	79	0.042178	0.04649	213	0.01930	0.01900
93	–	–	–	2	0.00018	0.00009
94	5	0.00267	0.00140	89	0.00807	0.00906
95	47	0.02509	0.01727	164	0.01486	0.01587
97	44	0.02349	0.02573	197	0.01785	0.02091
98	1	0.00053	0.00021	23	0.00208	0.00330
99	–	–	–	2	0.00018	0.00012
102	3	0.00160	0.00084	4	0.00036	0.00010
103	3	0.00160	0.01026	2	0.00018	0.00005
104	4	0.00214	0.00193	–	–	–
105	25	0.01335	0.00548	126	0.01142	0.01231
106	13	0.00694	0.01614	97	0.00879	0.00832
108	6	0.00320	0.00165	14	0.00127	0.00081
109	22	0.01175	0.01299	125	0.01133	0.00876
110	15	0.00801	0.00812	30	0.00272	0.00237
113	–	–	–	11	0.00100	0.00036
115	7	0.00374	0.00223	74	0.00671	0.00891
116	8	0.00427	0.00147	26	0.00236	0.00139
117	4	0.00214	0.00079	9	0.00082	0.00055
120	6	0.00320	0.00197	45	0.00408	0.00468
123	10	0.00534	0.00288	46	0.00417	0.00276
125	17	0.00908	0.01145	35	0.00317	0.00313
126	–	–	–	2	0.00018	0.00004
127	11	0.00587	0.00802	88	0.00798	0.00639
128	1	0.00053	0.00298	15	0.00136	0.00121
129	1	0.00053	0.00024	7	0.00063	0.00034
130	22	0.01175	0.01020	127	0.01151	0.01197

[1)]GH families according to the CAZy database http://www.cazy.org; Searches for glycoside hydrolases were performed with pfam HMMs, named in accordance with the CAZy nomenclature scheme. - no count observed, GH families not listed were not detected in any of the samples; the relative counts indicate the relative number in comparison to all GHs in the respective sample, and the relative coverage indicates the respective coverage in relation to the coverage of all GHs in the respective sample; GH families with cellulases are in bold.

doi:10.1371/journal.pone.0106707.t003

individuals, these data give hint to the great changes the intestinal microbiota is subjected to during animal's development.

Metagenome survey reveals remarkable metabolic richness

In this study metagenomic sequence data was generated for the intestinal microbiota of a six-years-old Asian elephant and of a three-weeks-old baby elephant. The reads observed from Illumina sequencing were assembled to 929,519,943 bp in 260,535 contigs for the six-years-old animal. The largest contig was 344,979 bp. The average contig length was 3,566 bp (TABLE 1). For the baby elephant 138,494,152 bp were assembled. The average contig length was 2,728 bp. The largest contig was 597,113 bp (TABLE 1).

To broadly characterize the sequence data, general traits were analyzed. Approximately 1,052,245 protein-coding genes could be identified for the-six-years-old elephant and 171,134 for the baby elephant. Of these, a total of 210,308 for the former and 42,640

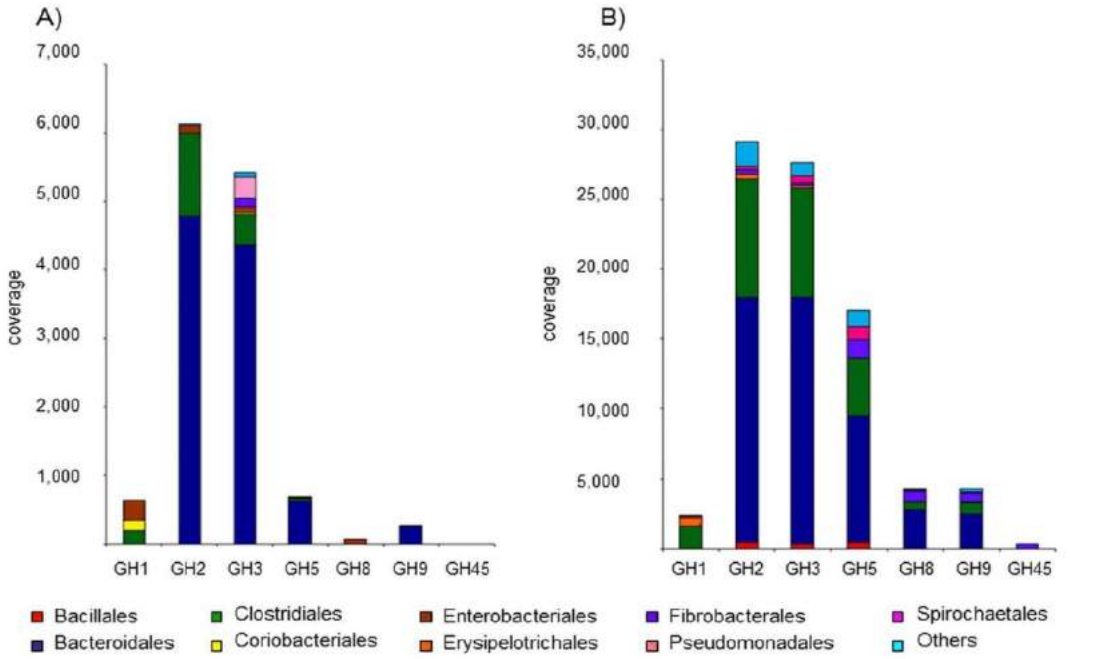

Figure 4. Overall coverage of selected cellulolytic GH family genes in the feces samples in relation to their phylogenetic affiliation. **A**) three-weeks-old elephant, **B**) six-years-old elephant.
doi:10.1371/journal.pone.0106707.g004

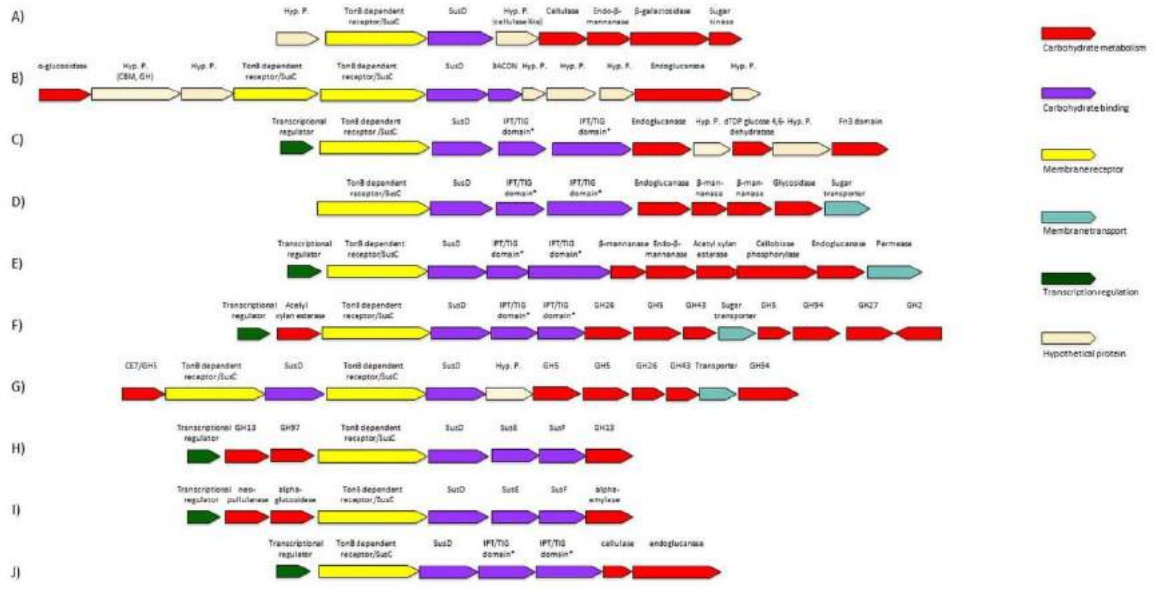

Figure 5. Physical map of selected putative polysaccharide utilization loci (PULs). A)–E) show regions from the sequence data derived from the microbiome of the six-years-old elephant. The different contigs are available from www.jgi.doe.gov (DOE Joint Genome Institute) under the IMG Project Id: 50566 with the scaffold id numbers: EMG_10007792, EMG_10000304, EMG_10002947, EMG_10003848 and EMG_10000174. **F–J)** indicate PULs from the literature or databases. **F**) from Tammar wallaby foregut [9], **G)** from Svalbard reindeer rumen [8,21], **H)** *Bacteroides thetaiotaomicron* [3], **I)** *B. faecis* CAG:32 (GenBank: FR891562.1), **J)** *Prevotella* sp. Sc00026 (GenBank JX424618.1). *indicates IPT/TIG domain containing hypothetical proteins which have also been annotated as SusE and SusF proteins.
doi:10.1371/journal.pone.0106707.g005

for the latter were similar to putative proteins in the KEGG database (TABLE 1). A small fraction (1.0/0.6% respectively) of the putative proteins was derived from eukaryotes. Although the available sequences do not allow a complete analysis, when taking the coverage of the obtained sequences into account, the 1.1 Gbp assembled DNA represented a significant fraction of the metagenome and the sequences therefore gave a good first estimation of the communities' metabolic potential. The amount of assembled sequence data was in general significantly higher than in similar studies (TABLE 2). Only for the cow rumen metagenome a two-fold higher amount of assembled sequence data was reported [6]. Furthermore the N50 of both datasets (5,751 and 22,052 bp, for the baby and the six-years-old elephant, respectively) were average or high in comparison with similar studies. E.g. in a study with Iberian Lynx the N50 was 4,370 bp [15] and for a study with bovine rumen it was 24 kb [6].

The KEGG analysis suggested that the metabolic potential of the microbiomes of the fecal samples is highly diverse and versatile. Genes of many of the classical catabolic pathways linked to the degradation of diverse polysaccharides and proteins, but also for the degradation of aromatic compounds, were identified. In both samples about 25% of the putative proteins were involved in carbohydrate metabolism and glycan biosynthesis (data not shown). A high number of genes coding for proteases (9,998 and 1,396 for the six-years-old and the three-weeks-old animal, respectively) were observed. Further 6,657/843 genes encoding putative esterolytic and/or lipolytic enzymes were detected, suggesting high hydrolytic activity for esters of fatty acids within both microbial communities.

Unique GH diversity in the fecal samples

Since the older elephant was fed on a large variety of different plant-based polysaccharides we focused on the analysis of genes and enzymes linked to carbohydrate catabolism. For a more detailed analysis we compared our set of predicted CAZymes (carbohydrate active enzymes) to entries in the CAZy database. The CAZy database contains a large set of validated CAZymes and it provides a sequence-based family classification of enzymes that are involved in the modification and breakdown of polysaccharides [50]. One of the most striking findings was the observation that the metagenomes of both elephants exhibited extraordinary high diversity of GHs when compared to other fecal samples of herbivores (TABLE 2). Our analyses identified a total of 11,038 putative genes for CAZymes from 82 different GH families in the six-years-old elephant's microbiome (TABLE 2). In the three-weeks-old elephant we detected 1,873 GH genes from 70 GH families. The majority of the identified GH genes were predicted to represent full-length genes. Furthermore, more than 50% of the enzymes showed less than 50% identity to the nearest neighbor (data not shown).

Interestingly, the diversity of GH family enzymes observed in both samples in this study was in general 2-fold higher than reported for any of the previously studied samples (TABLE 2) and with respect to the study on the cow rumen the GH diversity still was 20% higher [6]. This observation is in line with the observed higher phylogenetic diversity in the feces sample of the older elephant. When comparing GH diversity and OTU richness for the baby elephant, a very high GH diversity was observed (TABLE 2). Within the older elephant's fecal metagenome enzymes belonging to GH2, GH3, GH5, GH43 and GH78 families were predominant. Altogether 4,039 genes were linked to these five GH families, being equivalent to one third of all observed CAZymes in this study (TABLE 3). With respect to the overall occurrence of GH2, GH3, GH5 and GH43 family enzymes this observation fitted well with reports on other herbivores such as cow rumen, reindeer and others summarized in TABLE 2. GH2 and GH3 family enzymes encompass ß-galactosidases, ß-glucosidases, exoglucanases but also xylosidases. These enzymes are involved in the breakdown of a large variety of oligosaccharides. GH5 enzymes are endoglucanases involved in cellulose breakdown. Similar to microbiota of other herbivores the six-years-old elephant's data set contained a large number of putative genes matching enzymes of GH families specific for the metabolism of xylo-oligosaccharides (TABLE 3). The most abundant were GH10 mainly acting as ß-1,4-xylanases and the GH43 family enzymes acting mainly as ß-xylosidases (TABLE 3).

Further, we observed a high number of GH78 and GH13 family enzymes. GH78 family enzymes are mainly α-L-rhamnosidases and involved in rhamnose removal from polysaccharides or other molecules including polyphenols. GH78 has only been reported to be predominant in the cow rumen metagenome [6]. Since GH13 family enzymes are mainly involved in starch breakdown the relatively high number of GH13 enzymes suggests that starch depolymerization plays a major role for nutrient uptake in the elephant's intestinal tract. This is likely since the elephants are also fed with fruits and vegetables containing high amounts of starch.

With respect to those enzymes involved in cellulose breakdown our data analysis suggested that within the feces microbiome of the six-years-old animal at least 10 GH families can be identified which are partially involved in cellulose hydrolysis (e.g. GH1, GH3, GH5, GH8, GH9, GH44, GH45, GH51, GH74 and GH94). The predominant ones were GH5 (517) and GH9 (119) cellulases and endoglucanases. Further, 239 GH51 endoglucanases/arabinofuranosidases could be identified (TABLE 3). A total number of 84 GH94 enzymes were found. GH94 enzymes are acting as cellobiose or cellodextrin phosphorylases. Thus altogether 2,074 candidate enzymes were identified to be involved in cellulose degradation in the feces sample of the six-years-old elephant and 242 in the feces of the three-weeks-old elephant, respectively.

The overall occurrence of cellulolytic GH genes per assembled Mbp of metagenome DNA in the elephant feces sample was comparable to those numbers reported for the cow rumen and the reindeer [6,8] (TABLE 2).

The observation that the baby elephant microbiome already encoded a large GH diversity was intriguing. It however may suggest that the mother transfers already very early a part of its microbiome to the baby. This is in line with earlier reports [48]. Baby elephants occasionally eat feces from older elephants what might explain the relatively high content of GH family enzymes in the three-weeks-old animal. Nevertheless, the majority of GH family enzymes in the three-weeks-old elephant were ß-galactosidases (GH2). The rather frequent occurrence of many GH2 family enzymes is most likely linked to the breast-feeding and the high contents of lactose in the mothers' milk.

Bacteroidales constitute the main cellulolytic organisms within the elephant feces microbiome

Interestingly, when considering protein sequence identities, the GH5, GH9 and other GH families associated with cellulolytic enzyme activities belonged in majority to Bacteroidales (FIGURE 4). In both elephants most of these sequences belonged to *Bacteroidaceae*, followed by *Lachnospiraceae* for the older and *Porphyromonadaceae* for the baby elephant (data not shown). This partially differs from other herbivores such as cow rumen in which cellulose degrading enzymes mainly derived from *Clostridium* or *Ruminococcus* species [1]. Nevertheless, a significant role of

Bacteroidetes in carbohydrate degradation was already suggested for the intestines of other animals like wallaby and reindeer [8,9]. Furthermore electron microscopic examinations confirmed the absence of significant numbers of cellulosome-producing bacteria in the elephants' fecal samples (data not shown). Instead small coccoid cells were identified in close proximity to plant cell walls (FIGURE S1). Furthermore searches within the metagenome data set failed to identify high numbers of cellulosome associated proteins. For the older elephant 4 genes encoding a protein with a dockerin and a GH domain were discovered, for the baby elephant none. Because of these observations and the high coverage of GH5 and GH9 genes for Bacteroidales (FIGURE 4) we speculate that these are the main cellulose degraders in the elephant gut. Members of this phylum are known to degrade a variety of carbohydrates but only rarely cellulose [3]. They are well known to degrade starch and other carbohydrates with proteins encoded in an operon designated as PUL (polysaccharide utilization locus) typically containing the genes *susA* to *susG*. We identified 1,383 putative *susD* genes for the six-years-old elephant and 733 for the breast-fed animal. Interestingly, about 25 putative *sus* operons were identified which included a putative cellulase gene suggesting a possible role of PULs during cellulose breakdown. These clusters had a size from 14 kb to 25 kb. Next to cellulases these operons included a variety of different glycoside hydrolases like mannanases, galactosidases and arabinosidases indicating the flexibility of Bacteroidetes. PULs have been identified in a growing number of herbivore microbiomes and in aquatic environments [8,9,51]. The observation here, however, that 1,383 *susD* homologues (733 for the baby elephant) were identified suggests a high importance of these for polysaccharide degradation. Some of the clusters comprising a *susD* gene are exemplarily shown in FIGURE 5. When comparing these to clusters from other feces samples or those deriving from single strains it becomes obvious that most of the operons show high synteny (FIGURE 5). Interestingly, the operon organization in the elephant derived clusters resembles *Prevotella* species rather than *Bacteroides faecis* or the model organism *B. thetaiotaomicron* (FIGURE 5).

Besides the Bacteroidetes, the Fibrobacteres show high abundance of GH5 genes with more than 8% for the six-years-old elephant (FIGURE 4), compared to rather low abundance on phylogenetic level (2%). In contrast to the six-years-old elephant the baby elephant did not show any residing Spirochaetes or Fibrobacteres. Microorganisms of both phyla are known for their ability to digest lignocellulose and to contribute to this task in termites [16]. *Fibrobacter succinogenes* has been described as cellulolytic bacterium able to grow on crystalline cellulose. Nevertheless its mode of cellulose degradation is not elucidated. It does not produce multiprotein complexes like some *Ruminococcus* or *Clostridium* species or secrete cellulases in the surrounding medium [52]. *F. succinogenes* instead produces cellulose-binding proteins that putatively mediate close proximity of the bacterium to the substrate [46]. Therefore and as a relatively high numbers of GH5 genes have high sequence similarity to Fibrobacteres it is likely that they also play a significant role in cellulose degradation in the elephant gut.

Similar to other samples like the termite gut microbiome [16] and the cow rumen [6], exoglucanase genes were absent, suggesting that these plant-cell-wall (GH6) and *Clostridium*-derived (GH48) cellulolytic enzymes play virtually no role during cellulose degradation in the elephant feces microbiome.

Concluding remarks

Elephants are the largest land-living animals, but our knowledge on their microbiome is limited. The data presented here give a first insight into the phylogeny of the elephant gut microbiome (i.e. fecal samples). The phylogenetic analysis of the feces of the six-years-old Asian elephant indicated that these animals host a very diverse community dominated by Bacteroidetes and Firmicutes (approx. 3,000 OTUs at 97% identity level). Furthermore, metagenome sequencing revealed a very high GH diversity of altogether 84 families. Bacteroidetes thereby seem to have a predominant role in biomass degradation. Altogether these findings distinguish the elephant feces microbiota from that of other animals like cow (Clostridiales and *Prevotellaceae* predominant) or termite (Spirochetes and Fibrobacteres predominant). Thus elephants do have a unique and flexible microbiome that meets their high energy need and allows them to digest a wide range of plant-based biopolymers. In this context it is possible to speculate that the elephant is rather a generalist and not a specialist regarding the breakdown of plant biomass. This speculation is in line with the nutrition of the six-years-old animal, which is based on leaves, twigs, hay, grass, vegetables and fruits. While both studied animals were zoo animals, in nature elephants also nurture from a wide variety of plant-derived biomass.

Supporting Information

Figure S1 TEM picture of a typical cell found in the feces of the six-years-old Asian elephant.
(TIF)

Table S1 Description of diversity and richness of the fecal samples of the three-weeks-old and the six-years-old Asian elephant based on 16S rRNA gene analysis. The data for the six-years-old includes Bacteria and Archaea, the data for the three-weeks-old elephant only Bacteria.
(DOCX)

Table S2 Bins observed for the baby elephant.
(DOCX)

Table S3 Bins observed for the six-years-old elephant.
(DOCX)

Table S4 OTUs observed in the 16S rRNA gene datasets of the six-years-old and the three-weeks-old elephant and their frequencies.
(DOCX)

Supporting Information S1 16S rRNA gene analysis of the baby elephant.
(HTML)

Supporting Information S2 16S rRNA gene analysis of the six-years-old elephant.
(HTML)

Author Contributions

Conceived and designed the experiments: NI MSB WRS. Performed the experiments: NI SG JD BW. Analyzed the data: MA AP CS JS PBO TR DT. Contributed reagents/materials/analysis tools: RD AG MSB. Contributed to the writing of the manuscript: NI SG UR JT JC RD AG WRS.

References

1. Björnhag G (1994) The Digestive System in Mammals: Food, Form and Function: Cambridge University Press: 287–309.
2. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA (2008) Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nature Reviews Microbiology 6: 121–131.
3. Flint HJ, Scott KP, Duncan SH, Louis P, Forano E (2012) Microbial degradation of complex carbohydrates in the gut. Gut Microbes 3: 289–306.
4. Bayané A, Guiot S (2011) Animal digestive strategies versus anaerobic digestion bioprocesses for biogas production from lignocellulosic biomass. Reviews in Environmental Science and Bio/Technology 10: 43–62.
5. Clauss M, Frey R, Kiefer B, Lechner-Doll M, Loehlein W, et al. (2003) The maximum attainable body size of herbivorous mammals: morphophysiological constraints on foregut, and adaptations of hindgut fermenters. Oecologia 136: 14–27.
6. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, et al. (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331: 463–467.
7. Brulc JM, Antonopoulos DA, Miller ME, Wilson MK, Yannarell AC, et al. (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. Proceedings of the National Academy of Sciences 106: 1948–1953.
8. Pope PB, Mackenzie AK, Gregor I, Smith W, Sundset MA, et al. (2012) Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. PLoS One 7 (6): e38571.
9. Pope PB, Denman SE, Jones M, Tringe SG, Barry K, et al. (2010) Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. Proceedings of the National Academy of Sciences 107: 14793–14798.
10. Dai X, Zhu Y, Luo Y, Song L, Liu D, et al. (2012) Metagenomic insights into the fibrolytic microbiome in yak rumen. Public Library of Science One 7 (7): e40430.
11. Zhu L, Wu Q, Dai J, Zhang S, Wei F (2011) Evidence of cellulose metabolism by the giant panda gut microbiome. Proceedings of the National Academy of Sciences 108: 17714–17719.
12. Yang S, Ma S, Chen J, Mao H, He Y, et al. (2010) Bacterial diversity in the rumen of Gayals (*Bos frontalis*), Swamp buffaloes (*Bubalus bubalis*) and Holstein cow as revealed by cloned 16S rRNA gene sequences. Molecular Biology Reports 37: 2063–2073.
13. Singh KM, Ahir VB, Tripathi AK, Ramani UV, Sajnani M, et al. (2012) Metagenomic analysis of Surti buffalo (*Bubalus bubalis*) rumen: a preliminary study. Molecular Biology Reports 39: 4841–4848.
14. Lamendella R, Santo Domingo J, Ghosh S, Martinson J, Oerther D (2011) Comparative fecal metagenomics unveils unique functional capacity of the swine gut. BioMed Central Microbiology 11: 103–119.
15. Alcaide M, Messina E, Richter M, Bargiela R, Peplies J, et al. (2012) Gene sets for utilization of primary and secondary nutrition supplies in the distal gut of endangered Iberian lynx. Public Library of Science One 7: e51521.
16. Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature 450: 560–565.
17. Jacobs JP, Braun J (2014) Immune and genetic gardening of the intestinal microbiome. FEBS Letters doi: 10.1016/j.febslet.2014.02.052.
18. Buffie CG, Pamer EG (2013) Microbiota-mediated colonization resistance against intestinal pathogens. Nature Reviews Immunology 13: 790–801.
19. Varel VH (1987) Activity of Fiber-Degrading Microorganisms in the Pig Large Intestine. Journal of Animal Science 65: 488–496.
20. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, et al. (2008) Evolution of Mammals and Their Gut Microbes. Science 320: 1647–1651.
21. Schmitt S, Tsai P, Bell J, Fromont J, Ilan M, et al. (2012) Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. The ISME Journal 6: 564–576.
22. Chassard C, Lacroix C (2013) Carbohydrates and the human gut microbiota. Current Opinion in Clinical Nutrition & Metabolic Care 16: 453–460.
23. Leser TD, Amenuvor JZ, Jensen TK, Lindecrona RH, Boye M, et al. (2002) Culture-Independent Analysis of Gut Bacteria: the Pig Gastrointestinal Tract Microbiota Revisited. Applied and Environmental Microbiology 68: 673–690.
24. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, et al. (2003) A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. Science 299: 2074–2076.
25. Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, et al. (2010) Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. Cell 141: 1241–1252.
26. Mackenzie AK, Pope PB, Pedersen HL, Gupta R, Morrison M, et al. (2012) Two SusD-Like proteins encoded within a polysaccharide utilization locus of an uncultured ruminant *Bacteroidetes* phylotype bind strongly to cellulose. Applied and Environmental Microbiology 78: 5935–5937.
27. Ilmberger N, Meske D, Juergensen J, Schulte M, Barthen P, et al. (2012) Metagenomic cellulases highly tolerant towards the presence of ionic liquids-linking thermostability and halotolerance. Appl Microbiol Biotechnol 95: 135–146.
28. Zerbino DR, Birney E (2008) Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res 18: 821–829.
29. Nakaya A, Katayama T, Itoh M, Hiranuka K, Kawashima S, et al. (2013) KEGG OC: a large-scale automatic construction of taxonomy-based ortholog clusters. Nucleic Acids Research 41: D353–D357.
30. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, et al. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Research 29: 22–28.
31. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, et al. (2008) The Pfam protein families database. Nucleic Acids Research 36: D281–D288.
32. Wemheuer B, Guellert S, Billerbeck S, Giebel HA, Voget S, et al. (2013) Impact of a phytoplankton bloom on the diversity of the active bacterial community in the southern North Sea as revealed by metatranscriptomic approaches. Federation of European Microbiological Societies Microbiology Ecology 87: 378–389.
33. Wemheuer B, Taube R, Akyol P, Wemheuer F, Daniel R (2013) Microbial diversity and biochemical potential encoded by thermal spring metagenomes derived from the Kamchatka Peninsula. Archaea 2013: 136714–136725.
34. Bragg L, Stone G, Imelfort M, Hugenholtz P, Tyson GW (2012) Fast, accurate error-correction of amplicon pyrosequences using Acacia. Nat Methods 9: 425–426.
35. Pruesse E, Peplies J, Glockner FO (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28: 1823–1829.
36. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41: D590–596.
37. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7: 335–336.
38. Markowitz VM, Chen IM, Chu K, Szeto E, Palaniappan K, et al. (2012) IMG/M: the integrated metagenome data management and comparative analysis system. Nucleic Acids Research 40: D123–129.
39. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, et al. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BioMed Central Bioinformatics 9: 386–399.
40. Boisvert S, Raymond F, Godzaridis É, Laviolette F, Corbeil J (2012) Ray Meta: scalable de novo metagenome assembly and profiling. Genome Biology 13: R122.
41. Strous M, Kraft B, Bisdorf R, Tegetmeyer HE (2012) The Binning of Metagenomic Contigs for Microbial Physiology of Mixed Cultures. Frontiers in Microbiology 3: 410.
42. Wu M, Scott AJ (2012) Phylogenomic Analysis of Bacterial and Archaeal Sequences with AMPHORA2. Bioinformatics 28: 1033–1034.
43. Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, et al. (2007) "RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research 35: 3100–3108.
44. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 25: 955–964.
45. Besemer J, Lomsadze A, Borodovsky M (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Research 29: 2607–2618.
46. Engel P, Martinson VG, Moran NA (2012) Functional diversity within the simple gut microbiota of the honey bee. Proceedings of the National Academy of Sciences 109: 11002–11007.
47. Crutzen PJ, Aselmann I, Seiler W (1986) Methane production by domestic animals, wild ruminants, other herbivorous fauna, and humans. Tellus 38B: 271–284.
48. Collado MC, Cernada M, Bauerl C, Vento M, Pérez-Martínez G (2012) Microbial ecology and host-microbiota interactions during early life stages. Gut Microbes 3: 352–365.
49. Ward TL, Hosid S, Ioshikhes I, Altosaar I (2013) Human milk metagenome: a functional capacity analysis. BioMed Central Microbiology 13: 116–127.
50. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research 42: D490–D495.
51. Kabisch A, Otto A, König S, Becher D, Albrecht D, et al. (2014) Functional characterization of polysaccharide utilization loci in the marine Bacteroidetes *Gramella forsetii* KT0803. ISME J. doi: 10.1038/ismej.2014.4.
52. Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, et al. (2011) The Complete Genome Sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. Public Library of Science ONE 6: e18814.

5 Evaluation of 16S rRNA Gene Primer Pairs for Monitoring Microbial Community Structures Showed High Reproducibility within and Low Comparability between Datasets Generated with Multiple Archaeal and Bacterial Primer Pairs

Martin A. Fischer[1], Simon Güllert[2], Sven C. Neulinger[1, 3], Wolfgang R. Streit[2] and Ruth A. Schmitz[1]

[1]Department of Biology, Institute for General Microbiology, Christian-Albrechts-Universität zu Kiel, Kiel, Germany [2]Biocenter Klein Flottbek, Department of Microbiology and Biotechnology, University of Hamburg, Hamburg, Germany; [3]omics2view.consulting GbR, Kiel, Germany

Published in:

Frontiers in Microbiology 7 doi:10.3389/fmicb.2016.01297

frontiers in Microbiology

ORIGINAL RESEARCH
published: 23 August 2016
doi: 10.3389/fmicb.2016.01297

Evaluation of 16S rRNA Gene Primer Pairs for Monitoring Microbial Community Structures Showed High Reproducibility within and Low Comparability between Datasets Generated with Multiple Archaeal and Bacterial Primer Pairs

*Martin A. Fischer[1], Simon Güllert[2], Sven C. Neulinger[1,3], Wolfgang R. Streit[2] and Ruth A. Schmitz[1]**

[1] Department of Biology, Institute for General Microbiology, Christian-Albrechts-Universität zu Kiel, Kiel, Germany,
[2] Biozentrum Klein Flottbek, Institute of Microbiology & Biotechnology, Universität Hamburg, Hamburg, Germany,
[3] omics2view.consulting GbR, Kiel, Germany

OPEN ACCESS

Edited by:
John R. Battista,
Louisiana State University, USA

Reviewed by:
Henning Seedorf,
Temasek Life Sciences Laboratory, Singapore
Jonathan Badger,
National Cancer Institute, USA

***Correspondence:**
Ruth A. Schmitz
rschmitz@ifam.uni-kiel.de

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 26 May 2016
Accepted: 08 August 2016
Published: 23 August 2016

Citation:
Fischer MA, Güllert S, Neulinger SC, Streit WR and Schmitz RA (2016) Evaluation of 16S rRNA Gene Primer Pairs for Monitoring Microbial Community Structures Showed High Reproducibility within and Low Comparability between Datasets Generated with Multiple Archaeal and Bacterial Primer Pairs. Front. Microbiol. 7:1297. doi: 10.3389/fmicb.2016.01297

The application of next-generation sequencing technology in microbial community analysis increased our knowledge and understanding of the complexity and diversity of a variety of ecosystems. In contrast to Bacteria, the archaeal domain was often not particularly addressed in the analysis of microbial communities. Consequently, established primers specifically amplifying the archaeal 16S ribosomal gene region are scarce compared to the variety of primers targeting bacterial sequences. In this study, we aimed to validate archaeal primers suitable for high throughput next generation sequencing. Three archaeal 16S primer pairs as well as two bacterial and one general microbial 16S primer pairs were comprehensively tested by *in-silico* evaluation and performing an experimental analysis of a complex microbial community of a biogas reactor. The results obtained clearly demonstrate that comparability of community profiles established using different primer pairs is difficult. 16S rRNA gene data derived from a shotgun metagenome of the same reactor sample added an additional perspective on the community structure. Furthermore, *in-silico* evaluation of primers, especially those for amplification of archaeal 16S rRNA gene regions, does not necessarily reflect the results obtained in experimental approaches. In the latter, archaeal primer pair ArchV34 showed the highest similarity to the archaeal community structure compared to observed by the metagenomic approach and thus appears to be the appropriate for analyzing archaeal communities in biogas reactors. However, a disadvantage of this primer pair was its low specificity for the archaeal domain in the experimental application leading to high amounts of bacterial sequences within the dataset. Overall our results indicate a rather limited comparability between community structures investigated and determined using different primer pairs as well as between metagenome and 16S rRNA gene amplicon based community structure analysis. This finding, previously shown for Bacteria, was as well observed for the archaeal domain.

Keywords: amplicon sequencing, next-generation sequencing, archaea, metagenome, 16S microbial community, microbial communities

INTRODUCTION

The investigation of the microbial community composition allows a detailed insight in diversity and potential ecosystem function and fosters understanding of complex microbial processes (Vanwonterghem et al., 2014). Recent years have seen a strong increase in sequencing approaches targeting microbial communities via amplicon sequencing or metagenomic and metatranscriptomic approaches (Turnbaugh et al., 2007; Hamady et al., 2008; Raes and Bork, 2008; Caporaso et al., 2012; Grosskopf and Soyer, 2014; Ininbergs et al., 2015). These approaches play an important role in monitoring and comparing large numbers of samples in terms of their microbial composition (Caporaso et al., 2012; Kozich et al., 2013; Sundberg et al., 2013). The by far most often used marker for prokaryotic diversity studies is the 16S rRNA or its corresponding gene. The first to perform extensive research based on the 16S region were Woese and Fox (Woese and Fox, 1977; Woese et al., 1990). Their profound and passionate work led to the discovery of the third domain of life, the Archaea (Woese et al., 1990). Since then, the contribution of the archaeal domain to ecosystem function and diversity was often underestimated in many research fields and studies. While the bacterial fraction of many environments was extensively studied, the Archaea were often not specifically addressed. This underestimation of archaeal contribution to biology can be observed in a variety of studies from Sanger sequencing-based approaches to 454- and MiSeq-based high-throughput sequencing based studies (Frank et al., 2007; Herlemann et al., 2011; Ding and Schloss, 2014; Wang et al., 2015). Many reports focusing on Archaea appear to explore extreme environments like hot springs (Beam et al., 2015), deep sea volcanos (Reysenbach et al., 2006), and black smokers (Takai and Nakamura, 2011) to only mention a few, which further promotes the image of Archaea to represent extremophiles. On the contrary, Archaea are ubiquitously found under rather mesophilic conditions like in fresh and marine waters (DeLong, 1992; DeLong et al., 1994; Karner et al., 2001; Stahl and de la Torre, 2012), biogas reactors (Sundberg et al., 2013), and soil (Leininger et al., 2006), the intestinal tract of termites (Paul et al., 2012), ruminants (Jeyanathan et al., 2011; Kittelmann et al., 2013), but also on the human skin (Probst et al., 2013; Oh et al., 2014), or in the intestine (recently reviewed in Bang and Schmitz, 2015), where they complete the microbiome together with their bacterial, eukaryotic and viral partners.

Regarding biogeochemical cycles, the archaea harbor the unique trait of the methanogenic pathway (Offre et al., 2013). The methane emission by archaeal activity is used in industrial scale as a beneficial source of renewable energy in biogas reactors but is problematic when observed under the perspective of greenhouse gas emission. Two major sources of anthropogenic methane emission is livestock and rice patty fields (Yusuf et al., 2012), two habitats known to harbor methanogenic archaeal communities (Janssen and Kirs, 2008; Kittelmann et al., 2013; Breidenbach and Conrad, 2014) and both contributing notable amounts to the overall anthropogenic greenhouse gas emission (Wuebbles, 2002; Ripple et al., 2014). Additionally, in livestock methane production by the enteric community leads to an energy loss for the host by the emission of the energy rich methane and several studies investigate potential inhibitors of archaeal methane production (Goel and Makkar, 2012; Duin et al., 2016). The archaeal communities of ruminants has therefore been in the focus of several studies in recent years (Skillman et al., 2004; Jeyanathan et al., 2011; Kim et al., 2011; Singh et al., 2012; Tymensen and McAllister, 2012; Kittelmann et al., 2013; Henderson et al., 2015), some of them involving primer evaluation for the archaeal community (Watanabe et al., 2004; Gantner et al., 2011) or extending the microbiome research by adding results for protozoa and fungi (Kittelmann et al., 2013).

In an extensive study, Klindworth et al. (2013) performed a detailed *in-silico* evaluation of a 16S rRNA primer dataset containing 175 primers and 512 primer pairs, with 72 primers targeting archaeal 16S gene sequences. Primers and primer pairs were tested against the SILVA 16S non-redundant reference database to estimate their accuracy and phylogenetic coverage. Inspired by this study, we tested the experimental applicability of several primer combinations—some recommended in the above mentioned study, others supplemented based on literature review. After initial *in-silico* validation, the six most promising primer pairs were chosen; three targeting the archaeal, two the bacterial and one overall prokaryotic 16S rRNA gene sequence. These primer pairs showed high *in-silico* coverage and specificity, and were used to investigate the microbial community of an anaerobic, mesophilic biogas reactor, a habitat known to host a diverse community of Archaea and Bacteria (Eikmeyer et al., 2013; Sundberg et al., 2013). To eliminate disruptive effects and ensure maximum comparability, we used the same template DNA extracted from one sample of the above mentioned biogas reactor for all approaches. Shotgun metagenomic approaches have been introduced into community analysis (Venter et al., 2004) and bear the additional advantage of hinting toward ecosystem potential beside the taxonomic information (Vanwonterghem et al., 2014). Renunciation of 16S rRNA gene amplification is another positive effect of shotgun metagenomics, as it rids the data of primer bias (Shakya et al., 2013; Logares et al., 2014; Tremblay et al., 2015). Thus, as an additional and independent approach, we used 16S rRNA gene data obtained in a very comprehensive metagenome sequencing approach of the same biogas fermenter material (Güllert et al., 2016) as a reference point for comparison.

This study aims to estimate the effect of primer choice on the observed sequence composition of a diverse microbial community. Contrary to other studies focusing on the evaluation of bacterial 16S rRNA primers, we focus here on the evaluation and observation of the archaeal community in more detail. We further critically discuss the reliability of *in-silico* primer evaluation in terms of unspecific amplification and target specificity in application to environmental samples. Additionally, the 16S rRNA gene amplicon based community profiles were compared to the 16S rRNA gene sequences extracted and assembled from shotgun metagenomic data.

MATERIALS AND METHODS

Anaerobic Sludge Sample

For nucleic acid extraction, one sample was taken from a mesophilic (40°C) full-scale biogas plant (power output, 540–580 kWh) located near Cologne (Germany) on May 27th 2013. Main substrate for anaerobic digestion was maize silage (69%), cattle manure (19%), and dry poultry manure (12%). The pH of the reactor was 7.96, volatile fatty acids 3.06 g acetic acid equivalents/L, total inorganic carbon 17.7 g $CaCO_3$/L, free ammonia 2.98 g/L. One liter of sample material was taken under standardized conditions and kept at 4°C during transport to the laboratory, where it was stored at −20°C until DNA extraction.

DNA Extraction

Two milliliter of frozen sample were homogenized prior to extraction using a Dismembrator-U mechanical mortar (Sartorius Stedim Biotech GmbH, Göttingen, Germany) for 5 min at 2.500 rpm. DNA was extracted from the homogenate with the CTAB (Cetrimonium bromide) -chloroform:isoamyl alcohol based protocol as described by Weiland et al. (2010). Due to high concentration of humic acids in the final DNA extracts, DNA was further purified by the FastDNA™ SPIN Kit for Soil (MP Biomedicals, Solon, UH, USA). After extraction, purity was checked spectrophotometrically using the NanoDrop 1000 Fluorospectrometer (Thermo Fisher Scientific, Bremen, Germany) by measuring the absorbance at 260 and 280 nm (260/280 = 1.57).

16S Primer Selection and *In-silico* Evaluation

Three different primer pairs targeting the archaeal 16S rRNA gene region were selected from recent publications (Takai and Horikoshi, 2000; Baker et al., 2003; Yu et al., 2005; Frank et al., 2007; Fierer et al., 2008; Park et al., 2008; Morales and Holben, 2009; Claesson et al., 2010; Herlemann et al., 2011; Klindworth et al., 2013). The main criteria for their selection were (i) *in-silico* specify for archaea, (ii) low bias in amplifying specific groups of prokaryotes and (iii) amplicon length between 250 and 600 base pairs (bp), suited for next-generation sequencing techniques such as 454 Pyrosequencing or Illumina MiSeq. Two primer pairs targeting bacterial 16S rRNA gene (Lane et al., 1985; Hamady and Knight, 2009; Herlemann et al., 2011) and one primer pair for both bacterial and archaeal 16S rRNA gene (Klindworth et al., 2013) were selected applying the same criteria.

In-silico evaluation of the selected primers was performed using the arb-SILVA online tools TestPrime and TestProbe (Klindworth et al., 2013) employing the non-redundant SILVA 16S small subunit reference database (ssu r123, SILVA Ref NR; Pruesse et al., 2007). The results for the tested primers and primer pairs are listed in **Tables 1, 2**.

PCR and Library Preparation

Primers of each tested primer pair included the 454 Adapter set A, an unique identifier barcode sequence and a linker sequence as previously described (Langfeldt et al., 2014). For each primer pair, five separate PCR preparations with 5 individual Identifiers were performed for the investigation of experimental reproducibility. Extracted DNA was adjusted to a concentration of 20 ng/µl and used as template in the amplification reactions. The PCR reaction mixtures consisted of 17.25 µl H_2O (Carl Roth, Karlsruhe, Germany), 250 nM of the forward and reverse Primer (MWG, Ebersberg, Germany), 5 µl 5x Phusion Reaction buffer HF, 250 nM dNTP mix, 0.5 U Phusion HF DNA Polymerase (Thermo Fisher Scientific, Bremen, Germany) and 40 ng template. The annealing temperature for each primer pair was pre-calculated using the Tm Calculator online tool from NEB [version 1.8.1; New England Biolabs, Ipswitch (MA), USA (http://tmcalculator.neb.com)]. Cycling conditions for PCR amplification started with an initial denaturation step for 30 s at 95°C, followed by 30 cycles of 10 s at 95°C, 45 s at the respective annealing temperature and 30 s at 72°C and a final extension for 10 min at 72°C. All reactions were performed in triplicates with one corresponding negative control. Negative control contained no template DNA in the PCR reaction mixture and showed no amplification. Amplicons were checked for correct length and purified via agarose gel electrophoresis using the MinElute Gel Extraction Kit (Qiagen, Hilden, Germany) as previously described in Langfeldt et al. (2014). DNA concentration in eluates ranged from 7.9 to 77.5 ng/µl, as determined with a Nanodrop ND 1000 spectrophotometer (Thermo Fisher Scientific, Bremen, Germany). Samples were combined in the sequencing pool in equal concentrations. Sequencing was performed by MWG in accordance to manufacturer's recommendations on a Roche 454 GS-FLX++ system using the Titanium sequencing chemistry in two sequencing runs.

Amplicon Sequence Processing

Sequences were processed using Mothur v1.35.1 (Schloss et al., 2009) as described in Weiland-Bräuer et al. (2015) with the following modifications. Since some primers contained degenerated positions, sequences containing up to five differences in the primer region and one difference in the barcode region were kept in the dataset. Only those sequences with an average Phred score (Ewing et al., 1998; Ewing and Green, 1998) ≥25 and a maximum of eight homopolymers were retained. Each primer dataset was analyzed separately. Sequence alignment was performed against SILVA-based bacterial and archaeal reference alignments (Release 102; Pruesse et al., 2007). The alignment procedure, the filtering step, the removal of chimeric sequences, and the taxonomic classification were performed as against the mothur formatted SILVA database version 123 (Pruesse et al., 2007). For further processing and comparison between samples, only the domains of interest (Archaea, Bacteria or both) were kept in the dataset used for comparison of the diversity. Unclassified sequences or sequences from a domain not targeted by the respective primer pair were removed. The classification step produced a tax.summary file, listing the abundances of the classified taxa in each sample for each primer pair. For comparison between the different primer pairs, the tax.summary files of the different primer pairs were merged using the command merge.taxsummary.

Each dataset further yielded a shared file listing the sample-by-operational taxonomic unit (OTU) distribution for a given

TABLE 1 | *In-silico* evaluation results for selected primers.

Primer name	Primer sequence 5′ to 3′	Reference	T_M	Coverage of domain		
				Archaea	Bacteria	Eukarya
S-D-Arch-0787-a-S-20	ATTAGATACCCSBGTAGTCC	Yu et al., 2005	56.6	88.3	7.7	0
S-D-Arch-0787-a-A-20	GGACTACVSGGGTATCTAAT					
S-D-Arch-1043-a-A-16	GCCATGCACCWCCTCT	Yu et al., 2005	54.3	82.8	0	0
S-D-Arch-0519-a-S-15	CAGCMGCCGCGGTAA	Park et al., 2008	54.7	95.7	94.4	92.4
S-D-Arch-1041-a-A-18	GGCCATGCACCWCCTCTC	Baker et al., 2003	60.5	81.9	0	0
S-D-Arch-0349-a-S-17	GYGCASCAGKCGMGAAW	Takai and Horikoshi, 2000	56.4	81	0	0
S-D-Bact-0785-a-A-21	TACNVGGGTATCTAATCC	Claesson et al., 2010	51.8	92.1	92.3	0.8
S-D-Bact-0007-a-S-20	AGAGTTTGATCCTGGCTCAG	Frank et al., 2007	57.3	0	65.9	0
S-D-Bact-0338-a-A-19	TGCTGCCTCCCGTAGGAGT	Fierer et al., 2008	61.0	0	90.6	0
S-D-Bact-0341-b-S-17	CCTACGGGNGGCWGCAG	Herlemann et al., 2011	61.2	0.6	94.4	0
S-D-Bact-0907-a-A-20	CCGTCAATTCMTTTRAGTTT	Morales and Holben, 2009	51.2	0.8	90.2	75.4

The T_M temperature was calculated using the NEB online tool. The Coverage was estimated using the SILVA TestProbe tool with the sSU-NR database version 123 and refers to the percentage of potentially covered taxa in the different domains allowing non and one mismatch (MM) during primer annealing.

primer pair. These files were used in the calculation of α-diversity indices on the 97 and 99% similarity level.

Metagenome Sequencing, Assembly, and Annotation

Sequences corresponding to the biogas reactor sample discussed here were extracted from metagenomic data (Güllert et al., 2016) available under the accession PRJNA301928 (http://www.ncbi.nlm.nih.gov/bioproject/PRJNA301928). The extracted reads were trimmed using the Trimmomatic v0.30 software (Bolger et al., 2014). For the extraction of 16S rRNA gene sequences, the software reago v1.1 (Yuan et al., 2015) was used. 336,424 reads were identified as 16S rRNA gene sequences and assembled using the Spades assembler (Bankevich et al., 2012) generating 286 contigs with a length > 200 bp and a coverage above 2. N_{50} of the generated contigs was 336 bp. Before further analysis, the coverage and length information were used to normalize the contig abundances. Taxonomic annotation of metagenomic 16S rRNA gene sequences was performed with Mothur (Schloss et al., 2009). Sequences containing ambiguous bases and/or sequences shorter than 250 bp were removed. The remaining 198 unique contigs were taxonomically classified against the SILVA database (version 123; Pruesse et al., 2007).

Amplicon sequences are accessible via ncbi (http://www.ncbi.nlm.nih.gov/bioproject/PRJNA315559) under PRJNA315559.

All following computational analysis was performed in R v3.2.1 (R Core Team, 2015) using the vegan package version 2.3-5 (Oksanen et al., 2015).

For the calculation of the alpha diversity, OTU's generated by the mothur pipeline at 97 and 99% similarity level after removal of singletons were used. OTU counts were transformed to relative abundances per sample and OTUs with abundance below 0.2% in the corresponding primer dataset were dismissed. Shannon diversity index (H) was calculated using equation 1 from filtered relative OTU counts using vegan's function *diversity()*.Shannon numbers equivalent (^{1}D) was calculated using equation (2) based on the Shannon diversity which was set as exponent to base *e* to yield ^{1}D (Legendre and Legendre, 1998).

$$H = -\sum_{i=1}^{N} p_i \ln(p_i) \quad (1)$$

$${}^1D = e^{\left(-\sum_{i=1}^{N} p_i \ln(p_i)\right)} = e^H \quad (2)$$

where $p_i = n_i/N$ is the relative and n_i is the absolute abundance of species *i* and *N* is the total number of species within the dataset.

Beta-diversity was calculated separately for the bacterial and archaeal domain based on the merged tax.summary files listing abundances of taxonomic bins for each sample. This abundance-per-sample table was transformed and normalized using the Hellinger equation shown in Equation (3) (Legendre and Gallagher, 2001). The Hellinger transformed data was used to calculate a Bray-Curtis dissimilarity matrix by the Equation (4) (Legendre and Legendre, 1998) as well as for the redundancy analysis (RDA).

$$y'_{ij} = \sqrt{\frac{y_{ij}}{y_{j+}}} \quad (3)$$

Where y_{ij} is the abundance of species *i* at site *j* and y_{j+} is the sum of all species at site *j*.

$$D_{BrayCurtis}(i,j) = 1 - \frac{\sum_{k=0}^{n} |y_{ik} - y_{jk}|}{\sum_{k=0}^{n} (y_{ik} + y_{jk})} \quad (4)$$

where the dissimilarity ($D_{BrayCurtis}$) between the community of two sites (*i,j*), *k* as index of the numbers of species and y_{ik} the abundance of species *i* at site *k*.

RESULTS

Primers Show High *In-silico* Coverage

Primers for the amplification of the archaeal and bacterial 16S rRNA genes were selected in a way to avoid amplification

TABLE 2 | Results of the *in-silico* evaluation of the selected primer pairs.

Primer pair	Pair name	$T_{A(exp)}$	Reference	fragment length (nt)	Coverage of domain no MM			Coverage of domain 1 MM		
					Archaea	Bacteria	Eukarya	Archaea	Bacteria	Eukarya
S-D-Arch-0787-a-S-20/S-D-Arch-1043-a-A-16	ArchV56	60	Yu et al., 2005	272	76.50	0.00	0.00	84	0	0
S-D-Arch-0519-a-S-15/S-D-Arch-1041-a-A-18	ArchV46	60	Ogawa et al., 2014	540	79.00	0.00	0.00	93.2	0	0
S-D-Arch-0349-a-S-17/S-D-Arch-0786-a-A-20	ArchV34	56	This study	457	74.00	0.00	0.00	90.7	0	0
S-D-Arch-0519-a-S-15/S-D-Bact-0785-a-A-21	PrkV4	60	van Bleijswijk et al., 2015	287	88.60	88.20	0.70	95.2	95	1.4
S-D-Bact-0007-a-S-20/S-D-Bact-0338-a-A-19	BacV12	56	Langfeldt et al., 2014	350	0.00	60.20	0.00	0	85.8	0
S-D-Bact-0341-b-S-17/S-D Bact-0907-a-A-20	BacV35	60	Güllert et al., 2016	586	0.00	85.60	0.00	60.8	95.3	0.2

The T_A temperature was applied in the amplification reaction. The Coverage of the primer pairs was estimated using the SILVA TestPrime tool with the sSU-NR database version 123 and refers to the percentage of potentially covered taxa in the different domains allowing no and one mismatch (MM) during primer annealing.

of sequences outside the designated domain. The primers designated to the amplification of archaeal and bacterial 16S rRNA gene exhibited high *in-silico* coverage for both domains (between 66 and 96%). All evaluated primers for the archaeal domain covered ≥81% of all phyla in the SILVA database present in January 2016 (ssu r123, SILVA Ref NR). Bacterial primers showed an even higher theoretical coverage of over 90% (see **Table 1**). An exception to this was the primer S-D-Bact-0007-a-S-20 which only matched 66% of the bacterial phyla in the database. The primers designed for the combined amplification of archaeal and bacterial 16S rRNA gene sequences showed the overall highest theoretical coverage targeting 95.7% of all archaeal and 92.3% of all bacterial sequences. When one mismatch in the base pairing of the primers was allowed, the coverage of the primers designed for the amplification of archaeal 16S rRNA gene sequences increased to a coverage of ≥87% of all phyla in the SILVA. A similar increase to ≥95% of all bacterial phyla in the SILVA database was observed for the primers targeting the bacterial 16S rRNA gene sequences. A hint for possible false amplification indicated by theoretical coverage outside the designated domain was observed for the primers S-D-Arch-0787-a-S/A-20, S-D-Bact-0341-b-S-17 and S-D-Bact-0907-a-A-20. The primers S-D-Bact-0907-a-A-20, S-D-Bact-0785-a-A-21, and S-D-Arch-0519-a-S-15 showed potential coverage of the eukaryotic domain (0.8–92.4%). Considering one mismatch during the annealing progress, the potential for false amplification outside the targeted domain increased in the above mentioned primers and was additionally observed in the primers S-D-Arch-0349-a-S-17 and S-D-Bact-0338-a-A-19 in minor amounts (**Table 1**).

Primers were combined into pairs as listed in **Table 2**. *In-silico* evaluation suggested pairs ArchV46 and BacV35 to be the most promising pairs for amplification of archaeal and bacterial 16S rRNA gene sequences, respectively. The primer pairs chosen for amplification of prokaryotic 16S rRNA gene sequences covered 88.6% of the archaeal and 88.2% of the bacterial phyla in the database *in-silico*.

High Read Reduction during Processing in Archaeal 16S rRNA Gene Sequences was Caused by Unspecific Amplification

Selected primers pairs designed for the archaeal domain exhibited low specificity in practical application. **Figure 1** shows amplicon read classification before the taxonomy-based filter step. The percentage of amplicon sequences classified as archaeal after the alignment and annotation step ranged between 64.5 and 10.89% in case of the primers designed for the detection of archaea. In the datasets of the primers designed for the amplification of bacterial 16S rRNA gene over 99% of the amplicon sequences were classified as bacterial (data not shown). The ratio of archaeal to bacterial sequences in the dataset generated with the prokaryotic primers was 3.2–96.8% (see **Figure 1**). Non-archaeal sequences in the archaeal datasets originated mostly from bacterial 16S rRNA gene regions, indicating unspecific amplification by the tested archaeal primer pairs. Read reduction within the mothur pipeline is visualized for each primer pair in **Figure 2** and summarized

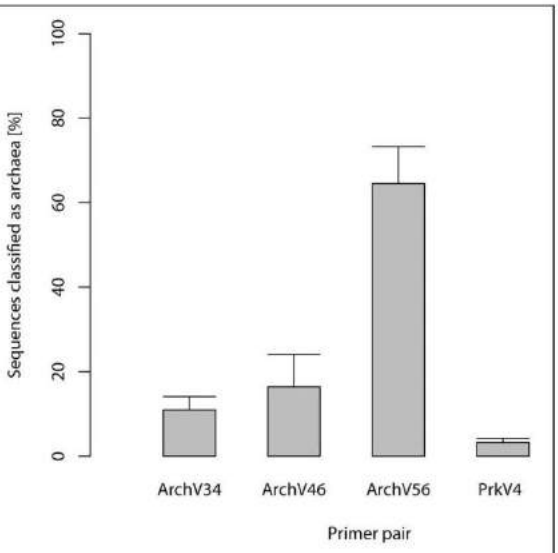

FIGURE 1 | Percentage of sequences classified as archaeal in the datasets generated applying the primers targeting the archaeal 16S ribosomal RNA gene region. 100% refers to all sequences in the separate datasets while the percentage of reads classified as Archaea is indicated by the boxes, standard deviation between the $n = 5$ replicates is indicated by the bars.

in Table S1. In the primer pairs ArchV46 and ArchV34, a strong reduction of the number of sequences in the datasets was observed when filtering for the targeted domain (*get.lineage* command). The primer pairs targeting the bacterial 16S rRNA gene region as well as the PrkV4 and ArchV56 primer pair showed only minor reduction in this filtering step. Reduction in this particular processing step hints toward the deletion of unclassified sequences as well as sequences from untargeted domains.

Taxonomic Composition Differs Depending on Used Primer Pair

The taxonomic annotation of the sequence in the datasets generated by the tested primer pairs and extracted from the metagenome (16S rRNA gene fraction) is shown in the **Figure 3** for the archaeal part on the order level as well as in **Figure 4** for the bacterial part on the class level.

In the archaeal domain, all primer pairs identified sequences annotated as *Methanosarcinales* as the most dominant order (45–66% of the sequences in the datasets) followed by *Methanomicrobiales* reaching between 17 and 43% respectively. *Methanobacteriales* were the third most abundant group ranging between 6 and 15% in relative abundance. 1–5% of the observed sequences were annotated as *Methanomassiliicoccales*, a *Thermoplasmatales* related methanogenic archaeal order (Paul et al., 2012; Sollinger et al., 2016). The overall community composition was consistent with the findings based on the metagenomic 16S rRNA gene sequences. Here, sequences classified as *Methanosarcinales* were the most dominant group with 50% followed by *Methanomicrobiales* with 43%. Sequences classified as *Methanobacteriales* and *Methanomassiliicoccales* each accounted for 4% of the community.

The sequences from the bacterial domain were dominated by the classes *Clostridia* (23–42%), *Bacteroidia* (10–23%), OPB54 belonging to the *Firmicutes* (12–19%), and *Mollicutes* (6–21%), accounting together for over 67% of all sequences in the primer-based datasets. An exception to this was the uncultured "*Candidatus*" *Cloacimonetes* of the SHA-4 candidate division belonging to the *Spirochaetes*. Sequences of these organisms were almost exclusively detected with primer pair BacV35 where it accounted for 17% of all sequences.

Concerning the 16S rRNA gene sequences extracted from the metagenome, the most prominent observation was the higher abundance of the "*Candidatus*" *Cloacimonas* sequences with 16% and the lower abundance of the *Clostridia* with only 18% compared to the primer based method. *Bacteroidia* (16%) and *Mollicutes* (2%) were among the highly abundant taxa as well. Additionally, 16S rRNA gene sequences with no classification below the domain level were of higher abundance compared to the amplicon based approaches. The proportion of sequences originating from Archaea was 2.4% while 97.6% of the sequences were classified as Bacteria.

Alpha Diversity Analysis Differed Highly between Tested Primer Pairs

Alpha diversity was analyzed for the different datasets based on the OTU composition generated at the 99% sequence similarity level (see **Figure 5**) as well as for the 97% sequence similarity level (see Figure S2), both summarized in **Table 3**. The observed diversity depends on (i) the potential of the amplified region to discriminate between different taxonomic units, (ii) the similarity cutoff value for OTU separation and (iii) the range of species amplified by the primer pair. On the 99% similarity level, the datasets of the primer pairs ArchV34 showed lower median alpha diversity of 5.7 for the archaeal domain, diversity was comparably high based on primer pair ArchV46, ArchV56 and PrkV4 with 7.2, 7.6, and 8.8. In contrast, alpha diversity was overall higher within the bacterial datasets. Bacterial diversity observed with primer pair PrkV4 was particularly high with 40.5, while the primer pairs BacV12 and BacV35 resulted in Shannon numbers equivalent of 21.8 and 34.7, respectively.

Beta Diversity Measurement

For comparison of variation in the observed archaeal and bacterial sequence composition, beta diversity was analyzed applying redundancy analysis (RDA) based on abundances of taxonomic bins at the genus level. Therefore, the merged tax.summary files holding information on the abundance of different genera in the communities observed by the primer pairs were transformed to a taxon-per-site table. This taxonomy based approach was chosen since comparison at the usual OTU level was impossible between the different primer pairs. After Hellinger transformation of count data, a Bray-Curtis dissimilarity matrix was calculated. The results are shown as

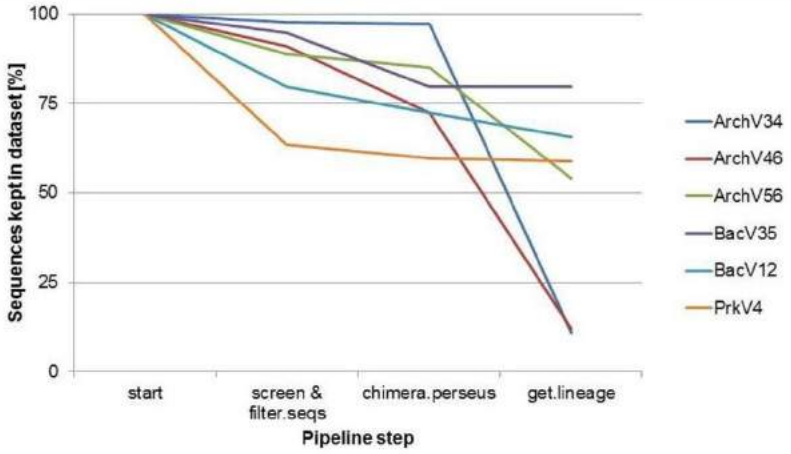

FIGURE 2 | Sequence reduction for the compared 16S rRNA gene primer pairs during the different steps of the mothur analysis pipeline. 100% refers to the number of all sequences in each dataset at the beginning of the Mothur sequence annotation pipeline.

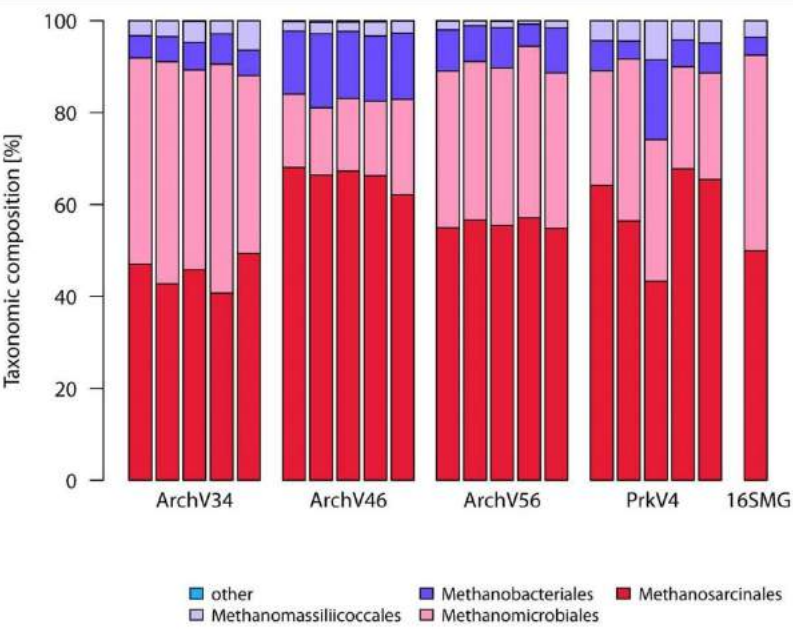

FIGURE 3 | Sequence compositions of the communities observed by the different archaeal and prokaryotic primer pairs on the order level. Shown are the five replicates of each primer pair as well as the 16S rRNA gene sequences extracted from the metagenome (16S-MG). Annotation was done using the SILVA v123 database. Taxa with abundance below 1% were grouped as "other".

a heat map in **Figure 6**. Higher similarity is indicated by red colors whereas low similarity between observed communities is indicated by white. Clustering of the replicates in the archaeal primer pairs (**Figure 6A**) was observed for the ArchV34, ArchV34, and ArchV56 as well as for four of five replicates of the PrkV4 primer pairs. In the bacterial dataset, the clustering of the replicates was stronger compared to the archaeal dataset (**Figure 6B**).

Redundancy analysis based on the Hellinger transformed data was performed using the command *rda* in R with "primer pairs"

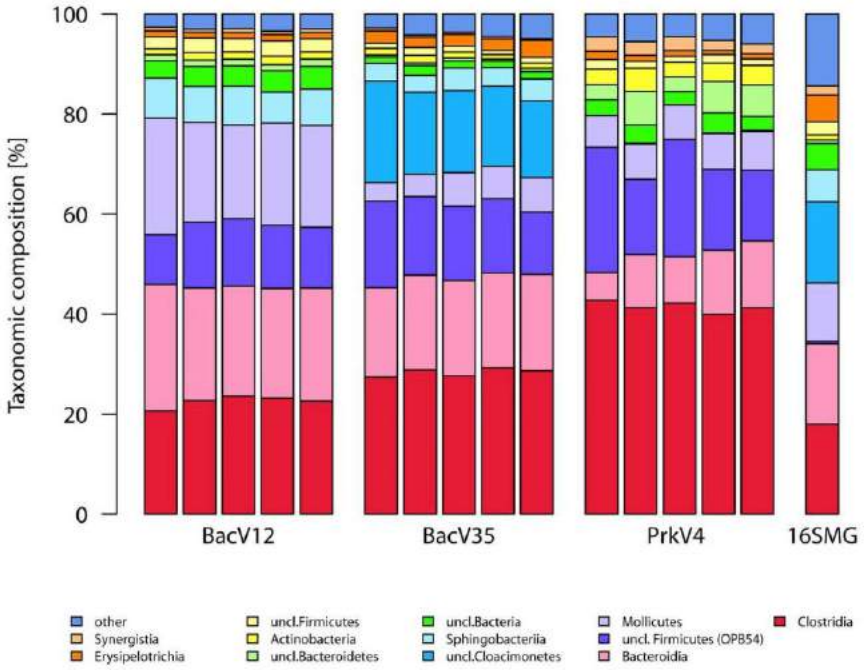

FIGURE 4 | Sequence compositions of the communities observed by the different bacterial and prokaryotic primer pairs on the class level. Shown are the five replicates of each primer pair as well as the 16S rRNA gene sequences extracted from the metagenome data pair (16S-MG). Annotation was done using the SILVA v123 database. Taxa with abundance below 1% were grouped as "other".

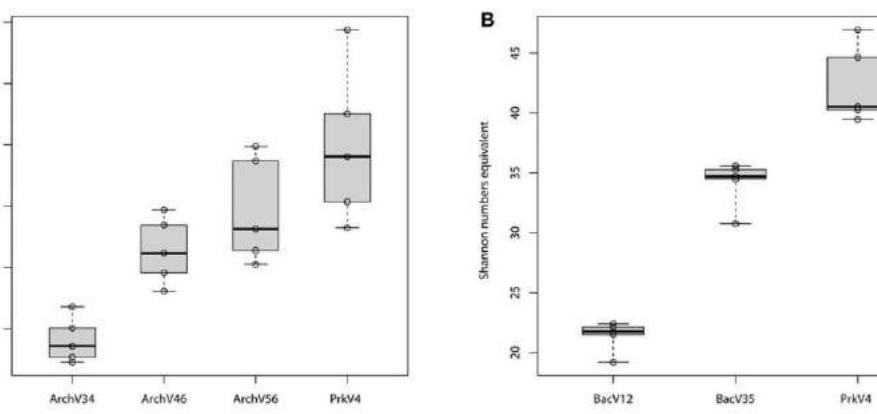

FIGURE 5 | Alpha diversity (Shannon numbers equivalent) observed on the 99% OTU level in the tested archaeal (A) and bacterial (B) datasets amplified with the primer pairs ($n = 5$) after removal singletons and low abundant OTUs (below 0.2%).

as explanatory factor. Overall adjusted variances explained by the first two RDA dimensions as shown in **Figure 7** were 69.35% for the archaeal dataset (**Figure 7A**) and 83.53% for the bacterial dataset (**Figure 7B**). Distinctly separated clusters of the different datasets indicate a specific image of the community for each primer pair. The model was tested for significance by an analysis of variance (ANOVA) using the command *anova* with 1000 permutations. Results showed high significance for the archaeal [$F_{(3,\ 16)} = 15.336$; $p = 0.001$] and bacterial [$F_{(2,\ 12)} = 36.509$; $p = 0.001$] models. Upon significance of the complete model, pairwise

TABLE 3 | Summary of the potential of the tested primer pairs *in-silico* and in application.

Primer set		Specificity [%]	Alpha diversity			Beta diversity
		Experimental	Database coverage observed [%]	Observed α-Diversity in the dataset		Dissimilarity to metagenome
			OTU 97%	OTU 99%		
PrkV4	ArchaeaV34	10.9	74.0	2.4	5.7	0.145
	ArchaeaV46	16.4	79.0	4.1	7.2	0.290
	ArchaeaV56	64.5	76.5	2.7	7.6	0.197
	Archaea	3.2	88.6	4.9	8.8	0.213
	Bacteria	96.8	88.2	41.2	40.5	0.489
	BacteriaV12	100.0	60.2	23.9	21.8	0.453
	BacteriaV35	99.9	85.6	32.5	34.7	0.462

Specificity refers to the percentage of sequences belonging to the targeted domain in the final dataset used for the analysis. Alpha diversity includes the potential coverage as calculated by the use of the SILVA-database version 123 and the observed alpha diversity calculated for the experimental datasets. Beta diversity summarizes the mean Bray-Curtis dissimilarity between the primer pairs and the 16S rRNA gene sequence composition extracted from the shotgun metagenome.

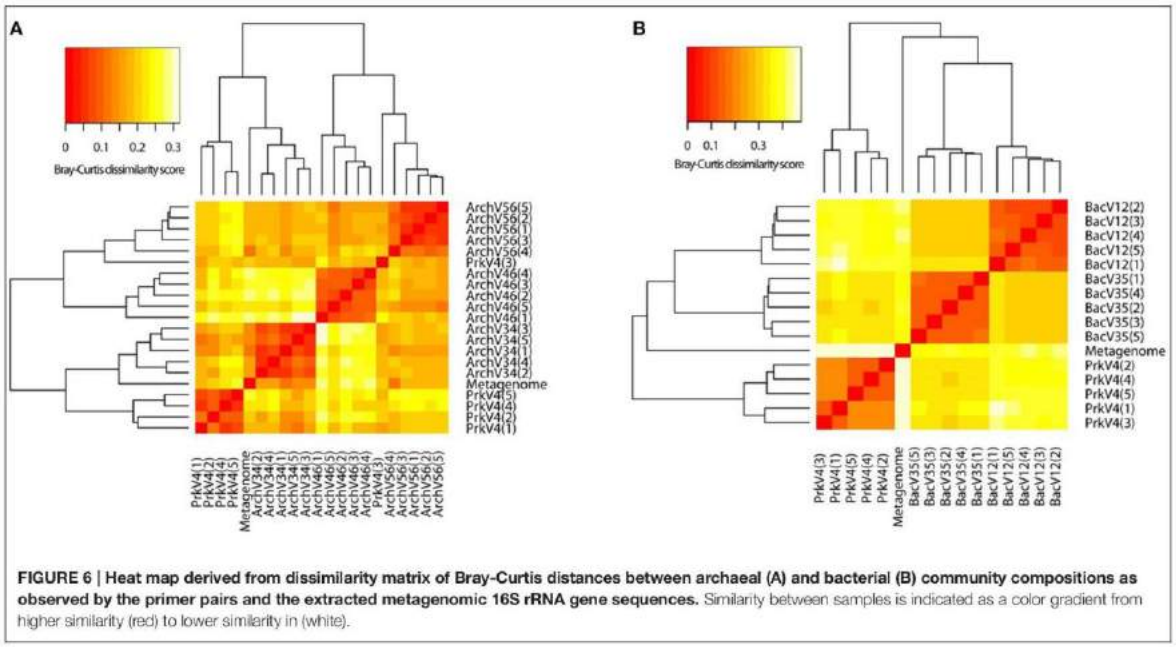

FIGURE 6 | Heat map derived from dissimilarity matrix of Bray-Curtis distances between archaeal (A) and bacterial (B) community compositions as observed by the primer pairs and the extracted metagenomic 16S rRNA gene sequences. Similarity between samples is indicated as a color gradient from higher similarity (red) to lower similarity in (white).

comparisons were conducted between the respective archaeal and bacterial primer pair datasets. Benjamini-Hochberg procedure was applied to correct *p*-values to account for pairwise testing (Benjamini and Hochberg, 1995). *F*-values ranged from 53.791 to 6.4614 with corresponding *q*-values between 0.0132 and 0.0260 for archaeal data (summarized in Table S2). *F*-values for pairwise testing of bacterial primer pairs ranged from 52.012 to 29.381 with corresponding *q*-values from 0.0135 to 0.0140 (summarized in Table S3).

To further investigate differences between observed community structures, indicator taxa were determined as previously described by Weiland-Bräuer et al. (2015) using the command *multipatt* of the R package indicspecies v.1.7.1 (Cáceres and Legendre, 2009) with 10^5 permutations. Indicator taxa are characteristic for a certain environment, or—as in our case—for a primer pair. Therefore, the analysis of indicator species may indicate possible over- or underestimation of certain taxa by the tested primer pairs, which is an important criterion for primer choice and reliability. Six indicator taxa from the archaeal domain were found to be characteristic in the datasets for the different primer pairs (summarized in Table S4). Sequences classified as *Methanosphaera* and *Methanoculleus* were found to be characteristic for the communities observed by primer pair ArchV34, whereas communities amplified by primer pair ArchV46 were more characterized by the presence of sequences annotated as

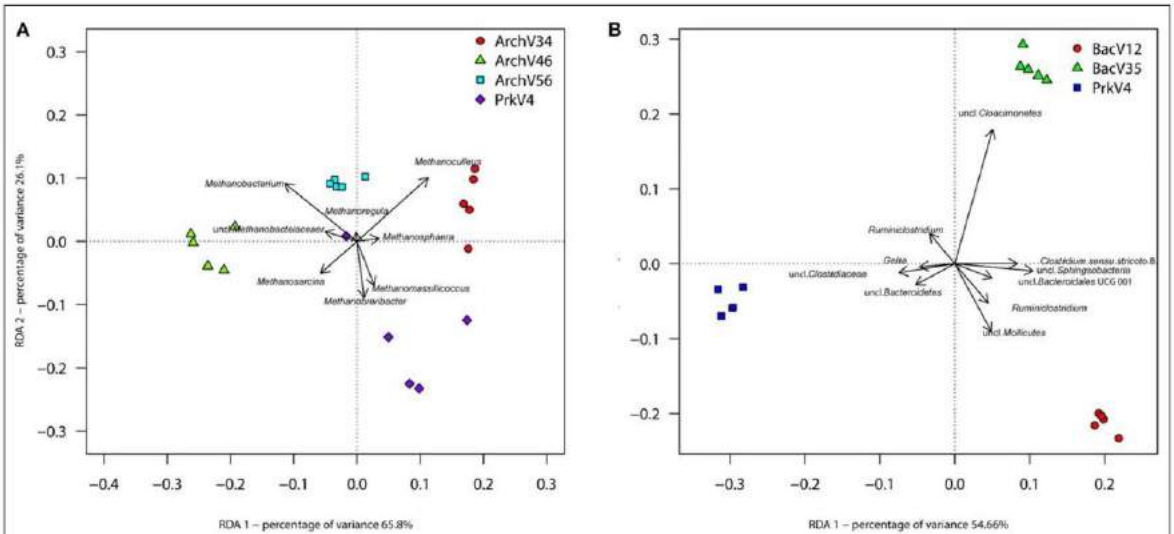

FIGURE 7 | RDA based on Hellinger-transformed taxon count data after Bray-Curtis dissimilarity calculation. Clear separation of the different primer pairs can be observed between the different archaeal **(A)** and bacterial **(B)** communities observed in the sample. The eight (Archaea) and 10 (Bacteria) taxa contributing most of the variance of the dataset are shown as vectors.

Methanobacterium and *Methanosarcina*. The prokaryotic primer pair PrkV4 was characterized by *Methanobrevibacter* and *Methanomassiliicoccus* sequences. No particularly characteristic taxon was found in the dataset generated with the ArchV56 primer pair.

For the bacterial domain, indicator taxon analysis found 10 taxa to be characteristic for the different primer pairs (summarized in Table S5). BacV12 was characterized by sequences annotated as uncultured WCHB1 69 belonging to the *Sphingobacteriales*, *Ruminiclostridium*, as well as uncultured taxa annotated as *Mollicutes*, *Clostridia* and *Bacteroidales*. Sequences annotated as uncultured *Cloacimonetes* were characteristic for primer pair BacV35. Primer pair PrkV4 had *Ruminiclostridium*, *Gelria* as well as two uncharacterized sequences annotated as *Clostridiaceae* and *Bacteroidetes* found to be characteristic for the dataset. Interestingly, the observed indicator taxa did not correlate with higher *in-silico* database coverage of a primer pair for that specific taxon and can therefore not be predicted based on the database analysis. The above mentioned indicator taxa contributed most to the variance explained by RDA and are visualized as vectors in **Figure 7**.

DISCUSSION

The choice of the primer pair for 16S rRNA gene amplification substantially determines quality and perspective on the obtained community data. Numerous recent studies address the topic of comparability between 16S rRNA gene based projects, demonstrating the influences of different effects with the help of mock communities and simulated datasets (Schloss et al., 2011; Brooks et al., 2015; Tremblay et al., 2015). However, complexity of environmental samples cannot be fully mimicked by artificially generated communities and the effects due to the choice of the primer pairs for analyzing complex environmental samples remain in question. Whereas most of the above mentioned studies focus mainly on Bacteria, here we presented comprehensive data generated for the archaeal and bacterial fraction of a complex environment, where we observed similar tendencies of primer effects in both domains. Based on our results five core statements can be formulated:

(A) All primer pairs were able to recover and represent a typical complex microbial community of an anaerobic biogas reactor, yet with a different outcome concerning the details of community structure. Sequences of key organisms for major steps in hydrolysis, acidogenesis and acetogenesis mostly belonging to the *Clostridia*, *Bacteroidia*, and *Actinobacteria* were observed in all bacterial datasets. The archaeal datasets provided sequences of species capable of hydrogenotrophic, acetoclastic, and methylotrophic methanogenesis (Wirth et al., 2012; Sundberg et al., 2013). Especially in the sequences of the archaeal community, a clear abundances ranking of taxonomic orders (*Methanosarcinales* > *Methanomicrobiales* > *Methanobacteriales* ≥ *Methanomassiliicoccales*) was consistently conserved in all tested primer pairs as well as in the sequences obtained from the metagenome. Similarly, ranking order of the two most abundant bacterial classes (*Clostridia*> *Bacteroidia*) was conserved in the sequence

abundance in all bacterial datasets. In the metagenome derived 16S rRNA gene sequences, the classes *Clostridia* and *Bacteroidia* were highly abundant (18 and 16%) but in addition sequences of unclassified *Cloacimonetes* contributed 16% to the dataset (**Figure 4**). This taxon would have been missed using the primer pairs BacV12 or PrkV4 in the analysis of the environment. Organisms of this class were recently found in metagenomic datasets from anaerobic digesters (Solli et al., 2014) and are expected to participate in syntrophic degradation of fatty acids and protein intermediates (Pelletier et al., 2008; Limam et al., 2014). The lower proportion of *Clostridia* in the metagenomic 16S rRNA gene sequences might indicate an overestimation of this class in the amplicon based approaches. In our case, the most promising combination to analyze the community of the sample would have been the combination of the BacV35 and ArchV46 primer pairs, however the metagenomic sequences still show a different overall bacterial and archaeal community compared to that observed by those primer pairs (**Figure 6**).

(B) From a technical point of view, efficiency strongly differed between the evaluated primer pairs, namely due to unspecific amplification by archaeal primer pairs. While 65% of the raw reads from the primer pair ArchV56 could be used for the analysis, the read reduction for the primer pair ArchV34 left only 11% of the raw reads for the final analysis of the archaeal community. Reads filtered from the archaeal datasets mostly belonged to the bacterial domain, with high abundance of the class "*Candidatus*" *Cloacimonas*. While almost all removed sequences from the primer pair ArchV34 belonged to this class, sequences removed from the datasets of the primers ArchV46 and ArchV56 showed a higher diversity but, as expected, were not comparable in composition to the tested bacterial primer pairs (Figure S1). Low specificity of the ArchV34 primer pair toward archaea was not predicted by the Silva TestPrime tool (Klindworth et al., 2013), the reason for this being unclear. One of the primers applied in this primer pair, S-D-Arch-0787-a-A-20, showed potential unspecific amplification within the bacterial domain in the *in-silico* prediction. Inexplicably, the same primer was used as reverse complement in the primer pair ArchV56, which exhibited highest specificity of all archaeal primer pairs tested. This shows that the outcome of sequencing runs is still highly unpredictable and database results cannot be directly transferred to the wet lab application.

(C) Alpha diversity (Shannon numbers equivalent ^{1}D) differed between primer pairs. In general, higher observed alpha diversity for a specific primer pair indicates higher resolution of the present diversity (i.e., better separation of OTUs) in a given sample. The alpha diversity in general as well as the Shannon number equivalents was lower in the archaeal datasets compared to the bacterial. This observation has previously been made for comparable biogas reactors (Francisci et al., 2015). The Shannon numbers equivalent was selected as alpha diversity metric as it is more robust in the application in environmental data and can be seen as the number of equally abundant species needed to form the diversity observed within a given dataset (Jost et al., 2010). For the primers targeting the archaeal 16S rRNA gene regions, the observed alpha diversity correlates well with the prediction of the database coverage (see **Table 3**). High diversities were observed for the primer pairs ArchV46 (^{1}D = 7.23) and PrkV4 (^{1}D = 8.81) covering the highly variable region V4 (Cai et al., 2013). In combination with a high potential coverage of different phyla, ArchV46 thus appears to be quite promising and was recently applied for the analysis of the archaeal domain in a mesophilic anaerobic digesters (Goux et al., 2015). The high diversity observed in the PrkV4 dataset correlates well with the high theoretical coverage as predicted by the *in-silico* evaluation. In our study, the abundances of the domains Archaea and Bacteria were similar in the overall metagenomic (2.4–97.6%) and the amplicon based (3.2–96.8%) 16S rRNA gene sequences generated by primer pair PrkV4, thus showing no strong shift in the proportions of the two domains during amplicon generation. The ArchV56 dataset showed diversity within the same range, a finding which correlated well with the *in-silico* predicted coverage. The observed positive correlation of *in-silico* database coverage and observed diversity was also valid for the tested bacterial primer pairs.

(D) Beta diversity analysis showed good reproducibility within a primer pair, but poor comparability between primer pairs as shown in the heat map (**Figure 6**). Significant differences between the communities amplified by the tested primer pairs resulted from differential amplification of 16S rRNA gene sequences out of the same starting material as well as the ability of the amplified variable region to discriminate between different taxa (Shakya et al., 2013; Tremblay et al., 2015). This bias in amplification and classification is also assumed to be the reason for the observed differences compared to the metagenome derived sequences. The comparison to 16S rRNA gene sequences from metagenomic data has previously shown to conform with 16S rRNA amplicon sequences, generated from environmental samples like sheep rumen (Shi et al., 2014). For the archaeal dataset generated by the PrkV4, we observed lower clustering in the RDA (**Figure 7A**) as well as higher within group variance, which could already be observed in the community composition (**Figure 3**). This resulted from overall lower sequencing depth of the archaeal domain in this dataset. As mentioned before, the overall archaeal abundance within the dataset was 3.2% compared to the also targeted bacterial sequences. As described by Kittelmann et al. (2013), the application of multiple domain specific primer sets can be beneficial when strong differences in the abundance of the different domains can be observed in a habitat. In the before mentioned study, the rumen communities of several ruminants were investigated by the application for several specific primer sets targeting Bacteria, Archaea, Protozoa and Fungi and pooled in alternating proportions for sequencing to account for the different abundances of the rumen microbiota (Kittelmann et al., 2013). In conclusion, to reduce the within group variation of the PrkV4 archaeal

dataset an increased number of sequences would be needed for a saturating analysis of the archaeal domain using the general PrkV4 primer pair. As alternative, a separate analysis of the archaeal and bacterial domain might be beneficial for the investigated habitat.

(E) To give an advice on which primer pair to use is difficult, if not impossible since the choice depends on the habitat and the research question. However we summarized the results obtained in this study in **Table 3** to better compare the dis-advantages of the primer pairs tested in the presented study. Primer pair ArchV34 showed the highest similarity to the archaeal domain of the metagenomic results. Still, it cannot be fully recommended since it was highly unspecific for archaeal sequences. Due to this fact, only 11% of the sequences obtained for this primer pair could be used for analysis. The ArchV46 primer pair showed moderate specificity and high diversity which makes it a reliable candidate for the investigation of new archaeal taxa in diverse environments. It was successfully applied in a recent multi-omics approach investigating the archaeal domain in an anaerobic paddy field. In this complex environment it was able to detect a complex archaeal community, consisting mostly *of Methanomicrobia, Methanomassiliicoccales and Methanobacteria,* as well as some *Thermoprotei* (Ogawa et al., 2014). The primer pair ArchV56 showed the highest specificity for archaeal 16S rRNA gene sequences. Compared to the metagenomic archaeal community, the observed diversity and similarity was average, which makes this primer pair a decent choice for the detection of archaea, especially in environments with a low abundance of archaea. In combination with a fluorescent probe, this primer pair was originally designed for the quantification of archaea (Yu et al., 2005) and was applied in this context in diverse studies (Lee et al., 2008; Nettmann et al., 2008). Beside good theoretical coverage and high diversity, a clear advantage of the PrkV4 primer pair is the simultaneous amplification of archaea and bacterial 16S rRNA gene sequences which can be helpful in the investigation of synergies between the archaeal and bacterial domain in the environment. This advantage was already confirmed in other environmental studies like in the investigation of the coral-associated microbiota containing substantial amounts of *Thaumarchaeota* and minor amounts of *Euryarchaeota,* which were detected in coral mucus for the first time (van Bleijswijk et al., 2015). Unfortunately, compared to the bacterial 16S rRNA gene sequences extracted from the metagenome, the sequences generated with the PrkV4 primer pair showed the lowest similarity. One prominent prokaryotic primer pair is applied in the earth microbiome project (Caporaso et al., 2012). The PrkV4 primer pair showed higher *in-silico* coverage compared to the one applied in the earth microbiome project, still this study cannot provide a direct comparison. The primer pair BacV12 has been used in diverse medical and environmental studies so far (Rausch et al., 2011; Cozen et al., 2013; Langfeldt et al., 2014; Mensch et al., 2016). Even though the primer pair sowed the highest similarity to the bacterial domain observed in the metagenome, the low theoretical coverage and low observed diversity within the samples may hint toward a possible non-observance of present species. Average results in terms of diversity and similarity to the metagenomic results were observed for the primer pair BacV35. The amplification of the highly variable region 4 of the 16S rRNA gene qualifies this primer pair as a good candidate when the focus of the study lies on the bacterial domain only (Güllert et al., 2016).

In summary, it appears most beneficial to use the same primer pair when comparing different sites or environments by amplicon sequencing. This assumption has previously been made for bacterial communities (Baker et al., 2003; Frank et al., 2008; Tremblay et al., 2015) and, as shown here, is also valid for archaeal primer pairs. It should be mentioned that additional effects influencing the observed community structure also occur in form of nucleic acid extraction (LaMontagne et al., 2002; Brooks et al., 2015), the kits applied (Adams et al., 2015), PCR artifacts (Schloss et al., 2011; Brooks et al., 2015), and database bias (Werner et al., 2012), as well as bias introduced by selected hypervariable region (Chakravorty et al., 2007; Yu et al., 2008), the sequencing platform itself (Kim et al., 2011; Luo et al., 2012; Tremblay et al., 2015) or the sequencing center (Hiergeist et al., 2016). In addition, comparability between communities analyzed with different primer pairs is bound to taxonomically assigned sequences and therefore limited and biased by the completeness of the database (Werner et al., 2012). These aspects further emphasize the need for general standards when planning and conducting environmental microbiological research for the sake of improved comparability, like the human (Turnbaugh et al., 2007; Peterson et al., 2009) or earth microbiome project guidelines (Gilbert et al., 2010) with profound and detailed manual and instruction for the sample preparation, which is of great help for between-study comparability. Finally, a comparison to 16S rRNA gene sequences gained from a corresponding metagenome as presented here appears very helpful and can be recommended as an addition to the mock community testing (Brooks et al., 2015; Tremblay et al., 2015) for the evaluation of new archaeal or bacterial primer pairs, especially when the community composition of the investigated environment is yet undetermined.

AUTHOR CONTRIBUTIONS

MF designed the study, carried out experiments, analyzed, and interpreted the data, and wrote the manuscript. SG provided metagenomic data and designed the study. SN supported the statistical analyses on the sequencing data. WS provided metagenomic data and designed the study. RS designed the study, founded, and supervised the work. All authors gave input to the manuscript and participated in the writing process.

FUNDING

Financial support was provided by the BMBF project BioPara funded to RS (grant no. 03SF0421B) and WS (grant no.

03SF0421H). We acknowledge financial support by Land Schleswig-Holstein within the funding program Open Access Publikationsfonds.

ACKNOWLEDGMENTS

We thank Bioreact and Bonalytic GmbH (Troisdorf, Germany), especially Dr. Melanie Hecht and Dipl. BioIng. Thomas Dickhaus, for providing the biogas plant fermenter material and for the determination of the chemical process parameters. Further we would like to thank Anne Kupczok for bioinformatics support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2016.01297

REFERENCES

Adams, R. I., Bateman, A. C., Bik, H. M., and Meadow, J. F. (2015). Microbiota of the indoor environment: a meta-analysis. *Microbiome* 3, 49. doi: 10.1186/s40168-015-0108-3

Baker, G. C., Smith, J. J., and Cowan, D. A. (2003). Review and re-analysis of domain-specific 16S primers. *J. Microbiol. Methods* 55, 541–555. doi: 10.1016/j.mimet.2003.08.009

Bang, C., and Schmitz, R. A. (2015). Archaea associated with human surfaces: not to be underestimated. *FEMS Microbiol. Rev.* 39, 631–648. doi: 10.1093/femsre/fuv010

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Beam, J. P., Jay, Z. J., Schmid, M. C., Rusch, D. B., Romine, M. F., Jennings, R. M., et al. (2015). Ecophysiology of an uncultivated lineage of Aigarchaeota from an oxic, hot spring filamentous 'streamer' community. *ISME J.* 10, 210–224. doi: 10.1038/ismej.2015.83

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Ser. B Methodol.* 57, 289–300. doi: 10.2307/2346101

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Breidenbach, B., and Conrad, R. (2014). Seasonal dynamics of bacterial and archaeal methanogenic communities in flooded rice fields and effect of drainage. *Front. Microbiol.* 5:752. doi: 10.3389/fmicb.2014.00752

Brooks, J. P., Edwards, D. J., Harwich, M. D., Rivera, M. C., Fettweis, J. M., Serrano, M. G., et al. (2015). The truth about metagenomics: quantifying counteracting bias in 16S rRNA studies. *BMC Microbiol.* 15:66. doi: 10.1186/s12866-015-0351-6

Cáceres, M. D., and Legendre, P. (2009). Associations between species and groups of sites: indices and statistical inference. *Ecology* 90, 3566–3574. doi: 10.1890/08-1823.1

Cai, L., Ye, L., Tong, A. H., Lok, S., and Zhang, T. (2013). Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. *PLoS ONE* 8:e53649. doi: 10.1371/journal.pone.0053649

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/ismej.2012.8

Chakravorty, S., Helb, D., Burday, M., Connell, N., and Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J. Microbiol. Methods* 69, 330–339. doi: 10.1016/j.mimet.2007.02.005

Claesson, M. J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J. R., Ross, R. P., et al. (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* 38:e200. doi: 10.1093/nar/gkq873

Cozen, W., Yu, G., Gail, M. H., Ridaura, V. K., Nathwani, B. N., Hwang, A. E., et al. (2013). Fecal microbiota diversity in survivors of adolescent/young adult Hodgkin lymphoma: a study of twins. *Br. J. Cancer* 108, 1163–1167. doi: 10.1038/bjc.2013.60

DeLong, E. F. (1992). Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5685–5689. doi: 10.1073/pnas.89.12.5685

DeLong, E. F., Wu, K. Y., Prézelin, B. B., and Jovine, R. V. (1994). High abundance of Archaea in Antarctic marine picoplankton. *Nature* 371, 695–697. doi: 10.1038/371695a0

Ding, T., Schloss, P. D. (2014). Dynamics and associations of microbial community types across the human body. *Nature* 509, 357–360. doi: 10.1038/nature13178

Duin, E. C., Wagner, T., Shima, S., Prakash, D., Cronin, B., Yanez-Ruiz, D. R., et al. (2016). Mode of action uncovered for the specific reduction of methane emissions from ruminants by the small molecule 3-nitrooxypropanol. *Proc. Natl. Acad. Sci. U.S.A.* 113, 6172–6177. doi: 10.1073/pnas.1600298113

Eikmeyer, F. G., Rademacher, A., Hanreich, A., Hennig, M., Jaenicke, S., Maus, I., et al. (2013). Detailed analysis of metagenome datasets obtained from biogas-producing microbial communities residing in biogas reactors does not indicate the presence of putative pathogenic microorganisms. *Biotechnol. Biofuels* 6:49. doi: 10.1186/1754-6834-6-49

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. *II. error probabilities. Genome Res.* 8, 186–194. doi: 10.1101/gr.8.3.186

Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. *I. accuracy assessment. Genome Res.* 8, 175–185. doi: 10.1101/gr.8.3.175

Fierer, N., Hamady, M., Lauber, C. L., and Knight, R. (2008). The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17994–17999. doi: 10.1073/pnas.0807920105

Francisci, D., de Kougias, P. G., Treu, L., Campanaro, S., and Angelidaki, I. (2015). Microbial diversity and dynamicity of biogas reactors due to radical changes of feedstock composition. *Bioresour. Technol.* 176, 56–64. doi: 10.1016/j.biortech.2014.10.126

Frank, D. N., St Amand, A. L., Feldman, R. A., Boedeker, E. C., Harpaz, N., and Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl. Acad. Sci. U.S.A.* 104, 13780–13785. doi: 10.1073/pnas.0706625104

Frank, J. A., Reich, C. I., Sharma, S., Weisbaum, J. S., Wilson, B. A., and Olsen, G. J. (2008). Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl. Environ. Microbiol.* 74, 2461–2470. doi: 10.1128/AEM.02272-07

Gantner, S., Andersson, A. F., Alonso-Sáez, L., and Bertilsson, S. (2011). Novel primers for 16S rRNA-based archaeal community analyses in environmental samples. *J. Microbiol. Methods* 84, 12–18. doi: 10.1016/j.mimet.2010.10.001

Gilbert, J. A., Meyer, F., Jansson, J., Gordon, J., Pace, N., Tiedje, J., et al. (2010). The earth microbiome project: meeting report of the "1 EMP meeting on sample selection and acquisition" at argonne national laboratory October 6 2010. *Stand. Genomic Sci.* 3, 249–253. doi: 10.4056/aigs.1443528

Goel, G., and Makkar, H. P. S. (2012). Methane mitigation from ruminants using tannins and saponins. *Trop. Anim. Health Prod.* 44, 729–739. doi: 10.1007/s11250-011-9966-2

Goux, X., Calusinska, M., Lemaigre, S., Marynowska, M., Klocke, M., Udelhoven, T., et al. (2015). Microbial community dynamics in replicate anaerobic digesters exposed sequentially to increasing organic loading rate, acidosis, and process recovery. *Biotechnol. Biofuels* 8, 122. doi: 10.1186/s13068-015-0309-9

Grosskopf, T., and Soyer, O. S. (2014). Synthetic microbial communities. *Curr. Opin. Microbiol.* 18, 72–77. doi: 10.1016/j.mib.2014.02.002

Güllert, S., Fischer, M. A., Turaev, D., Noebauer, B., Ilmberger, N., Wemheuer, B., et al. (2016). Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and

elephant feces reveal major differences in carbohydrate hydrolysis strategies. *Biotechnol. Biofuels* 9:121. doi: 10.1186/s13068-016-0534-x

Hamady, M., and Knight, R. (2009). Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Res.* 19, 1141–1152. doi: 10.1101/gr.085464.108

Hamady, M., Walker, J. J., Harris, J. K., Gold, N. J., and Knight, R. (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Methods* 5, 235–237. doi: 10.1038/nmeth.1184

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., and Janssen, P. H. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci. Rep.* 5:14567. doi: 10.1038/srep14567

Herlemann, D. P., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J. J., and Andersson, A. F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* 5, 1571–1579. doi: 10.1038/ismej.2011.41

Hiergeist, A., Reischl, U., and Gessner, A. (2016). Multicenter quality assessment of 16S ribosomal DNA-sequencing for microbiome analyses reveals high inter-center variability. *Int. J. Med. Microbiol.* doi: 10.1016/j.ijmm.2016.03.005. [Epub ahead of print].

Ininbergs, K., Bergman, B., Larsson, J., and Ekman, M. (2015). Microbial metagenomics in the Baltic Sea: recent advancements and prospects for environmental monitoring. *Ambio 44(Suppl.* 3), 439–450. doi: 10.1007/s13280-015-0663-7

Janssen, P. H., and Kirs, M. (2008). Structure of the archaeal community of the rumen. *Appl. Environ. Microbiol.* 74, 3619–3625. doi: 10.1128/AEM.02812-07

Jeyanathan, J., Kirs, M., Ronimus, R. S., Hoskin, S. O., and Janssen, P. H. (2011). Methanogen community structure in the rumens of farmed sheep, cattle and red deer fed different diets. *FEMS Microbiol. Ecol.* 76, 311–326. doi: 10.1111/j.1574-6941.2011.01056.x

Jost, L., DeVries, P., Walla, T., Greeney, H., Chao, A., and Ricotta, C. (2010). Partitioning diversity for conservation analyses. *Divers. Distrib.* 16, 65–76. doi: 10.1111/j.1472-4642.2009.00626.x

Karner, M. B., DeLong, E. F., and Karl, D. M. (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* 409, 507–510. doi: 10.1038/35054051

Kim, M., Morrison, M., and Yu, Z. (2011). Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J. Microbiol. Methods* 84, 81–87. doi: 10.1016/j.mimet.2010.10.020

Kittelmann, S., Seedorf, H., Walters, W. A., Clemente, J. C., Knight, R., Gordon, J. I., et al. (2013). Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. *PLoS ONE* 8:e47879. doi: 10.1371/journal.pone.0047879

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41, e1. doi: 10.1093/nar/gks808

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112–5120. doi: 10.1128/AEM.01043-13

LaMontagne, M. G., Michel, F. C., Holden, P. A., and Reddy, C. A. (2002). Evaluation of extraction and purification methods for obtaining PCR-amplifiable DNA from compost for microbial community analysis. *J. Microbiol. Methods* 49, 255–264. doi: 10.1016/S0167-7012(01)00377-3

Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L., and Pace, N. R (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. U.S.A.* 82, 6955–6959.

Langfeldt, D., Neulinger, S. C., Heuer, W., Staufenbiel, I., Künzel, S., Baines, J. F., et al. (2014). Composition of microbial oral biofilms during maturation in young healthy adults. *PLoS ONE* 9:e87449. doi: 10.1371/journal.pone.0087449

Lee, C., Kim, J., Shin, S. G., and Hwang, S. (2008). Monitoring bacterial and archaeal community shifts in a mesophilic anaerobic batch reactor treating a high-strength organic wastewater. *FEMS Microbiol. Ecol.* 65, 544–554. doi: 10.1111/j.1574-6941.2008.00530.x

Legendre, P., and Gallagher, E. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia* 129, 271–280. doi: 10.1007/s004420100716

Legendre, P., and Legendre, L. (1998). *Numerical Ecology, 2nd English Edn.* Amsterdam, NY: Elsevier.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., et al. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442, 806–809. doi: 10.1038/nature04983

Limam, R. D., Chouari, R., Mazéas, L., Wu, T.-D., Li, T., Grossin-Debattista, J., et al. (2014). Members of the uncultured bacterial candidate division WWE1 are implicated in anaerobic digestion of cellulose. *Microbiologyopen* 3, 157–167. doi: 10.1002/mbo3.144

Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F. M., Ferrera, I., Sarmento, H., et al. (2014). Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ. Microbiol.* 16, 2659–2671. doi: 10.1111/1462-2920.12250

Luo, C., Tsementzi, D., Kyrpides, N., Read, T., and Konstantinidis, K. T. (2012). Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS ONE* 7:e30087. doi: 10.1371/journal.pone.0030087

Mensch, B., Neulinger, S. C., Graiff, A., Pansch, A., Künzel, S., Fischer, M. A., et al. (2016). Restructuring of epibacterial communities on *Fucus vesiculosus* forma mytili in response to elevated pCO2 and increased temperature levels. *Front. Microbiol.* 7:434. doi: 10.3389/fmicb.2016.00434

Morales, S. E., and Holben, W. E. (2009). Empirical testing of 16S rRNA gene PCR primer pairs reveals variance in target specificity and efficacy not suggested by *in silico* analysis. *Appl. Environ. Microbiol.* 75, 2677–2683. doi: 10.1128/AEM.02166-08

Nettmann, E., Bergmann, I., Mundt, K., Linke, B., and Klocke, M. (2008). Archaea diversity within a commercial biogas plant utilizing herbal biomass determined by 16S rDNA and mcrA analysis. *J. Appl. Microbiol.* 105, 1835–1850. doi: 10.1111/j.1365-2672.2008.03949.x

Offre, P., Spang, A., and Schleper, C. (2013). Archaea in biogeochemical cycles. *Annu. Rev. Microbiol.* 67, 437–457. doi: 10.1146/annurev-micro-092412-155614

Ogawa, D. M. O., Moriya, S., Tsuboi, Y., Date, Y., Prieto-da-Silva, Á. R. B., Rádis-Baptista, G., et al. (2014). Biogeochemical typing of paddy field by a data-driven approach revealing sub-systems within a complex environment–a pipeline to filtrate, organize and frame massive dataset from multi-omics analyses. *PLoS ONE* 9:e110723. doi: 10.1371/journal.pone.0110723

Oh, J., Byrd, A. L., Deming, C., Conlan, S., Kong, H. H., and Segre, J. A. (2014). Biogeography and individuality shape function in the human skin metagenome. *Nature* 514, 59–64. doi: 10.1038/nature13786

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., et al. (2015). *Vegan: Community Ecology Package, R Package Version 2.3-0.* Available online at: https://cran.r-project.org/web/packages/vegan/

Park, S.-J., Park, B.-J., and Rhee, S.-K. (2008). Comparative analysis of archaeal 16S rRNA and amoA genes to estimate the abundance and diversity of ammonia-oxidizing archaea in marine sediments. *Extremophiles* 12, 605–615. doi: 10.1007/s00792-008-0165-7

Paul, K., Nonoh, J. O., Mikulski, L., and Brune, A. (2012). "Methanoplasmatales," Thermoplasmatales-related archaea in termite guts and other environments, are the seventh order of methanogens. *Appl. Environ. Microbiol.* 78, 8245–8253. doi: 10.1128/AEM.02193-12

Pelletier, E., Kreimeyer, A., Bocs, S., Rouy, Z., Gyapay, G., Chouari, R., et al. (2008). "Candidatus Cloacamonas acidaminovorans": genome sequence reconstruction provides a first glimpse of a new bacterial division. *J. Bacteriol.* 190, 2572–2579. doi: 10.1128/JB.01248-07

Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J. A., et al. (2009). The NIH human microbiome project. *Genome Res.* 19, 2317–2323. doi: 10.1101/gr.096651.109

Probst, A. J., Auerbach, A. K., and Moissl-Eichinger, C. (2013). Archaea on human skin. *PLoS ONE* 8:e65388. doi: 10.1371/journal.pone.0065388

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196. doi: 10.1093/nar/gkm864

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Austria: Statistical Computing.

Raes, J., and Bork, P. (2008). Molecular eco-systems biology: towards an understanding of community function. *Nat. Rev. Microbiol.* 6, 693–699. doi: 10.1038/nrmicro1935

Rausch, P., Rehman, A., Künzel, S., Häsler, R., Ott, S. J., Schreiber, S., et al. (2011). Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19030–19035. doi: 10.1073/pnas.1106408108

Reysenbach, A.-L., Liu, Y., Banta, A. B., Beveridge, T. J., Kirshtein, J. D., Schouten, S., et al. (2006). A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. *Nature* 442, 444–447. doi: 10.1038/nature04921

Ripple, W. J., Smith, P., Haberl, H., Montzka, S. A., McAlpine, C., and Boucher, D. H. (2014). Ruminants, climate change and climate policy. *Nat. Clim. Change* 4, 2–5. doi: 10.1038/nclimate2081

Schloss, P. D., Gevers, D., Westcott, S. L., and Gilbert, J. A. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE* 6:e27310. doi: 10.1371/journal.pone.0027310

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Shakya, M., Quince, C., Campbell, J. H., Yang, Z. K., Schadt, C. W., and Podar, M. (2013). Comparative metagenomic and rRNA microbial diversity characterization using archaeal and bacterial synthetic communities. *Environ. Microbiol.* 15, 1882–1899. doi: 10.1111/1462-2920.12086

Shi, W., Moon, C. D., Leahy, S. C., Kang, D., Froula, J., Kittelmann, S., et al. (2014). Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res.* 24, 1517–1525. doi: 10.1101/gr.168245.113

Singh, K. M., Tripathi, A. K., Pandya, P. R., Parnerkar, S., Rank, D. N., Kothari, R. K., et al. (2012). Methanogen diversity in the rumen of Indian Surti buffalo (*Bubalus bubalis*), assessed by 16S rDNA analysis. *Res. Vet. Sci.* 92, 451–455. doi: 10.1016/j.rvsc.2011.03.022

Skillman, L. C., Evans, P. N., Naylor, G. E., Morvan, B., Jarvis, G. N., and Joblin, K. N. (2004). 16S ribosomal DNA-directed PCR primers for ruminal methanogens and identification of methanogens colonising young lambs. *Anaerobe* 10, 277–285. doi: 10.1016/j.anaerobe.2004.05.003

Solli, L., Håvelsrud, O. E., Horn, S. J., and Rike, A. G. (2014). A metagenomic study of the microbial communities in four parallel biogas reactors. *Biotechnol. Biofuels* 7, 146. doi: 10.1186/s13068-014-0146-2

Söllinger, A., Schwab, C., Weinmaier, T., Loy, A., Tveit, A. T., Schleper, C., et al. (2016). Phylogenetic and genomic analysis of Methanomassiliicoccales in wetlands and animal intestinal tracts reveals clade-specific habitat preferences. *FEMS Microbiol. Ecol.* 92:fiv149. doi: 10.1093/femsec/fiv149

Stahl, D. A., and de la Torre, J. R. (2012). Physiology and diversity of ammonia-oxidizing archaea. *Annu. Rev. Microbiol.* 66, 83–101. doi: 10.1146/annurev-micro-092611-150128

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

Takai, K., and Horikoshi, K. (2000). Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. *Appl. Environ. Microbiol.* 66, 5066–5072. doi: 10.1128/AEM.66.11.5066-5072.2000

Takai, K., and Nakamura, K. (2011). Archaeal diversity and community development in deep-sea hydrothermal vents. *Curr. Opin. Microbiol.* 14, 282–291. doi: 10.1016/j.mib.2011.04.013

Tremblay, J., Singh, K., Fern, A., Kirton, E. S., He, S., Woyke, T., et al. (2015). Primer and platform effects on 16S rRNA tag sequencing. *Front. Microbiol.* 6:771. doi: 10.3389/fmicb.2015.00771

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The human microbiome project. *Nature* 449, 804–810. doi: 10.1038/nature06244

Tymensen, L. D., and McAllister, T. A. (2012). Community structure analysis of methanogens associated with rumen protozoa reveals bias in universal archaeal primers. *Appl. Environ. Microbiol.* 78, 4051–4056. doi: 10.1128/AEM.07994-11

van Bleijswijk, J. D. L., Whalen, C., Duineveld, G. C. A., Lavaleye, M. S. S., Witte, H. J., and Mienis, F. (2015). Microbial assemblages on a cold-water coral mound at the SE Rockall Bank (NE Atlantic): interactions with hydrography and topography. *Biogeosci. Discuss.* 12, 1509–1542. doi: 10.5194/bgd-12-1509-2015

Vanwonterghem, I., Jensen, P. D., Ho, D. P., Batstone, D. J., and Tyson, G. W. (2014). Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. *Curr. Opin. Biotechnol.* 27, 55–64. doi: 10.1016/j.copbio.2013.11.004

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74. doi: 10.1126/science.1093857

Wang, J., Kalyan, S., Steck, N., Turner, L. M., Harr, B., Kunzel, S., et al. (2015). Analysis of intestinal microbiota in hybrid house mice reveals evolutionary divergence in a vertebrate hologenome. *Nat. Commun.* 6:6440. doi: 10.1038/ncomms7440

Watanabe, T., Asakawa, S., Nakamura, A., Nagaoka, K., and Kimura, M. (2004). DGGE method for analyzing 16S rDNA of methanogenic archaeal community in paddy field soil. *FEMS Microbiol. Lett.* 232, 153–163. doi: 10.1016/S0378-1097(04)00045-X

Weiland, N., Löscher, C., Metzger, R., and Schmitz, R. (2010). Construction and screening of marine metagenomic libraries. *Methods Mol. Biol.* 668, 51–65. doi: 10.1007/978-1-60761-823-2_3

Weiland-Bräuer, N., Neulinger, S. C., Pinnow, N., Künzel, S., Baines, J. F., and Schmitz, R. A. (2015). Composition of bacterial communities associated with *Aurelia aurita* changes with compartment, life stage, and population. *Appl. Environ. Microbiol.* 81, 6038–6052. doi: 10.1128/AEM.01601-15

Werner, J. J., Koren, O., Hugenholtz, P., DeSantis, T. Z., Walters, W. A., Caporaso, J. G., et al. (2012). Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. *ISME J.* 6, 94–103. doi: 10.1038/ismej.2011.82

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

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5088–5090. doi: 10.1073/pnas.74.11.5088

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya. *Proc. Natl. Acad. Sci. U.S.A.* 87, 4576–4579. doi: 10.1073/pnas.87.12.4576

Wuebbles, D. (2002). Atmospheric methane and global change. *Earth Sci. Rev.* 57, 177–210. doi: 10.1016/S0012-8252(01)00062-9

Yu, Y., Lee, C., Kim, J., and Hwang, S. (2005). Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnol. Bioeng.* 89, 670–679. doi: 10.1002/bit.20347

Yu, Z., García-González, R., Schanbacher, F. L., and Morrison, M. (2008). Evaluations of different hypervariable regions of archaeal 16S rRNA genes in profiling of methanogens by Archaea-specific PCR and denaturing gradient gel electrophoresis. *Appl. Environ. Microbiol.* 74, 889–893. doi: 10.1128/AEM.00684-07

Yuan, C., Lei, J., Cole, J., and Sun, Y. (2015). Reconstructing 16S rRNA genes in metagenomic data. *Bioinformatics* 31, 43. doi: 10.1093/bioinformatics/btv231

Yusuf, R. O., Noor, Z. Z., Abba, A. H., Hassan, M. A. A., and Din, M. F. M. (2012). Methane emission by sectors: a comprehensive review of emission sources and mitigation methods. *Renewable Sustain. Energy Rev.* 16, 5059–5070. doi: 10.1016/j.rser.2012.04.008

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

6 Identification and initial characterization of novel feruloyl esterases obtained from an agricultural biogas fermenter metagenome

Simon Güllert[1] and Wolfgang R. Streit*[1]

[1]Biozentrum Klein Flottbek, Abteilung für Mikrobiologie & Biotechnologie, Universität Hamburg, Ohnhorststrasse 18, 22609 Hamburg, Germany

Unpublished

***For Correspondence**
E-Mail: wolfgang.streit@uni-hamburg.de
Abteilung für Mikrobiologie und Biotechnologie,
Biozentrum Klein Flottbek, Universität Hamburg,
Ohnhorststr.18,
D-22609 Hamburg, Germany,
Tel. (+49) 40-42816-463
Fax. (+49) 40-42816-459

INTRODUCTION

The main reason for the high recalcitrance of lignocellulosic biomass against enzymatic hydrolysis is the high degree of cross-linking of the polysaccharides to each other and to lignin in the plant cell wall. In this context, an important molecule is ferulic acid. This hydroxycinnamic acid derivate cross-links the polysaccharides of the hemicellulose fraction via ester bonds and also cross-links the hemicellulose to lignin via ether bonds (Wong *et al.*, 2011). For an efficient hydrolysis of lignocellulosic biomass, the depolymerization of the covalent cross-linkage between the different fractions is an important aspect (Faulds, 2010). Therefore, feruloyl esterases (=ferulic acid esterases, FAEs) have recently attracted increasing attention. FAEs have the ability to hydrolytically cleave the ester linkage of ferulic acid to the polysaccharides (Williamson *et al.*, 1998, Crepin *et al.*, 2004). By this, they make the plant cell wall polymers more accessible to cellulolytic and hemicellulolytic enzymes and thereby improve the biodegradation of biomass to fermentable sugars (Benoit *et al.*, 2008). Due to this synergistic action with other lignocellulolytic enzymes, they are also referred to as accessory enzymes (Wong, 2006).

In addition, FAEs have applications in the food, pharmaceutical, and paper-pulp industries (Koseki *et al.*, 2009). FAEs can be used to release ferulic acid from plant biomass. Ferulic acid has known anti-oxidant and anti-inflammatory functions which can be utilized in the health industry. Ferulic acid can also be converted to vanillic acid and vanillin which are important flavor components in the food industry. Finally, FAEs can be used to improve the efficiency of the bleaching of wood pulp in the paper industry (Benoit *et al.*, 2008).

For these reasons, feruloyl esterases have a high potential for biotechnological applications like the bioprocessing of lignocellulosic biomass to energy resources or value added products. That is why there is a significant interest in extending the spectrum of known FAEs. One approach to screen for novel FAEs is to tap into the genomic potential of hard to cultivate microorganisms using cultivation independent metagenomic techniques. In this study, potentially novel FEAs were identified in a metagenome obtained from an agricultural biogas fermenter sample via a function-based fosmid library screening.

MATERIAL & METHODS

Isolation of total DNA from the biogas fermenter sample

The isolation of total DNA from agricultural biogas fermenter samples is described in Güllert *et al.*, 2016.

Fosmid library construction and functional screening

For the construction of the large insert fosmid library, the CopyControl™ Fosmid Library Production Kit (Epicentre, Madison, USA) was used as recommended in the kit manual. Metagenomic DNA which was isolated from the March 2013 biogas fermenter sample was used for the construction of the fosmid library. Prior to library construction, the isolated DNA was purified using SureClean solution (Bioline, Luckenwalde, Germany) according to the manufacturer's instructions. As a heterologous host for the fosmid library, modified *E. coli* Epi300 cells named "UHH01" which encode the *C. cellulolyticum* σ^{70}-factor (RpoD) in their genome were used (Epi 300 UHH01: rpoD, ampR, Plac, ΔbioF, unpublished).

Functional screening of the fosmid library for feruloyl esterases

Initially, the *E.coli* fosmid clones were screened for their ability to hydrolyze the short-chain fatty acid tributyrin. For this purpose, the clones were grown on LB agar plates supplemented with 1 % (v/v) tributyrin (AppliChem, Darmstadt, Germany). Prior to autoclaving, the medium was warmed up to 60 °C and emulsified using a T18 Ultra Turrax homogenisator (IKA WORKS Inc., Wilmington, NC, USA). After autoclaving, the medium was supplemented with L-arabinose (final concentration 0.001 %) and 12.5 µg/ml chloramphenicol. Cells were grown over night at 37 °C and further incubated for up to 72 hours on the plates. The formation of a clear halo around a colony indicated esterolytic activity.

The clones exhibiting esterolytic activity were subsequently screened for their ability to degrade ethyl ferulate. For this screening, the clones were grown over night on LB agar plates containing 0.001 % L-arabinose. In a next step, the colonies were overlaid with 50 ml 0.5 % agarose solution which contained 100 mg ethyl ferulate (Ethyl 4-hydroxy-3-methoxycinnamate, Sigma Aldrich, Munich, Germany). The ethyl ferulate was previously

solved in 1 ml DMSO. The final ethyl ferulate concentration was 2 mg/ml in the agarose solution. The plates were incubated at 37 °C for approximately 4 hours. The formation of a clear halo around a colony indicated feruloyl esterase activity.

Fosmid subcloning of active colonies

The fosmids of *E.coli* clones which exhibited feruloyl esterase activity were subcloned. For this, the clones were grown in 5ml LB containing 0.001 % L-arabinose and 12.5 µg/ml chloramphenicol. The fosmids were isolated using the high-speed plasmid mini kit (Geneaid Biotech, Taiwan) as instructed by the manufacturer. The isolated fosmid DNA was subsequently sheared into smaller fragments via sonication. For the fragmentation, six impulses (20 % amplitude, 0.5 cycle) from a UP200S ultrasonic processor (Hielscher Ultrasonics GmbH, Teltow, Germany) were used. The fragmentation was verified via agarose gel electrophoresis and the sheared DNA fragments with a size between 2-6 kb were extracted from the gel using the Gel/ PCR DNA fragment extraction kit (Geneaid Biotech, Taiwan) as described by the manufacturer. Subsequently, the ends of the extracted DNA fragments were repaired. For this reaction 14 µl T4 DNA polymerase buffer (5x, Thermo Fischer Scientific), 1 µl dNTPs (10 mM), 1 µl T4 DNA polymerase (5 U/µl Thermo Fischer Scientific), up to 1 µg of sheared fosmid DNA were mixed with water (final volume 70 µl) and incubated at room temperature for 1 hour. Next, the reaction was inactivated at 75 °C for 10 min and purified using SureClean solution (Bioline, Luckenwalde, Germany) according to the manufacturer's instructions. After this purification, the DNA was dephosphorylated and adenine overhangs were generated. For this purpose, 44 µl DNA, 5 µl DreamTaq buffer (10x, Thermo Fischer Scientific), 1 µl FastAP alkaline phosphatase (1 U/µl, Thermo Fischer Scientific) were mixed and incubated for 15min at 37 °C. Subsequently, the reaction was heat inactivated at 75 °C for 5 minutes and 1 µl dATP (100 mM Thermo Fischer Scientific) and 1 µl DreamTaq polymerase (Thermo Fischer Scientific) were added and the reaction was further incubated at 72 °C for 30 minutes. Hereafter, the reaction was again cleaned up using SureClean solution. In the last step, the DNA was ligated into the pDrive cloning vector and introduced into *E.coli* DH5α cells via transformation as described in the manual of the pDrive cloning kit (Qiagen, Hilden, Germany). The obtained subclones were again tested for feruloyl esterase activity.

Sequencing of active fosmid subclones

The fosmid inserts were sequenced from both vector flanks using M13 forward and M13 reverse primers. Subsequently, the obtained insert sequences were used to design new primers. This primer walking was repeated until the insert was completely sequenced. Sanger sequencing was conducted by Eurofins genomics (Eurofins Genomics GmbH, Ebersberg, Germany). Complete insert sequences were subjected to gene prediction using Artemis 16 (Carver *et al.*, 2012).

Cloning into expression vector

Primer pairs corresponding to the coding gene sequences of MTP73F9 and MTP93F12 were designed to introduce *SacI* and *XhoI* restriction sites at the 5' and 3' ends of the respective gene for directional cloning into the pET21a(+) expression vector. The reverse primers did not include the stop codon to allow recombinant protein expression in-frame with a 6x His-tag at the c-terminal end of the proteins. After PCR amplification, the products were digested with *SacI* and *XhoI* DNA restriction enzymes (Thermo Fischer Scientific), followed by ligation into the pET21a(+) vector previously linearized with the same enzymes. The expression constructs were used for transformation into *E.coli* BL21 (DE3) cells via heat shock. The ligation reaction and the transformation were conducted as recommended in the pET system manual (11th edition, Novagen/Merck Millipore, Darmstadt, Germany).

Recombinant protein expression and purification

Overnight cultures of *E.coli* BL21 (DE3) cells containing the respective expression construct were used to inoculate 250 ml LB medium (supplemented with 100 µg/ml ampicillin). The cultures were grown at 37 °C (MTP73F9: 17°C) until an OD_{600} of approximately 1.0 was reached. At this point, IPTG (final concentration 1 mM) was added to the cultures for the induction of protein expression. The cultures were further incubated for 4 hours at 37°C (MTP73F9: incubation at 17 °C over night) and subsequently harvested via centrifugation at 5000 rpm for 10 minutes. The supernatants were discarded and the pellets were frozen at -70 °C for at least 10 minutes. Subsequently, the recombinant his-tagged proteins were purified via immobilized metal affinity chromatography (IMAC). For this purification, the Protino® Ni-TED 1000 kit (Macherey-Nagel, Düren, Germany) was used as recommended by

the manufacturer for the purification of polyhistidine-tagged proteins from *E.coli* under native conditions. For the preparation oy lysates, the cells was resuspended in LEW buffer and treated with Lysozyme (5 mg/ml) at 4 °C for 30 min. Afterwards the cell suspensions were sonicated on ice for 20 min using 20 % amplitude and a cycle of 0.5. Elution of the purified proteins was conducted in 3x 1.5 ml elution buffer. The eluted proteins were concentrated to a final volume of 500 µl using Vivaspin protein concentrator spin columns with a 5 kDa molecular weight cut-off (GE Healthcare, Solingen, Germany). Protein concentrations were determined using the Pierce™ BCA Protein Assay Kit (Thermo Fischer Scientific) as described in the manual. Purification and molecular weight of the recombinant proteins were determined via 10 % SDS-PAGE (Laemmli, 1970).

Phylogenetic analysis

Based on the FAE subclasses reported by Crepin *et al.* (Crepin *et al.*, 2004) and additionally reported termite hindgut derived FAEs by Rashamuse *et al.* (Rashamuse *et al.*, 2014), the phylogenetic position of the newly identifies enzymes in this study was determined. For this, the sequences of known feruloyl esterases and relatives were downloaded from the NBCI database (accession numbers are given in brackets). Nucleotide sequences were translated to amino acid sequences and aligned using T-Coffee v.11.00.8 (Notredame *et al.*, 2000) in accurate mode. In a next step, the aligned sequences were imported into MEGA 6 (Tamura *et al.*, 2013) and the evolutionary history was inferred using the Neighbor-Joining method (Saitou & Nei, 1987). A bootstrap test with one thousand replicates was conducted. The sequences of two acetyl xylan esterases were used an outgroup.

FAE substrate range characterization

Para-nitrophenol conjugated substrates were used for an initial substrate range characterization of the identified FAEs. Upon hydrolyzation by an esterase, the chromogenic *para*-nitrophenol (*p*NP) moiety is cleaved off from the substrate resulting in a quantifiable color change of the enzyme solution. For this assay, the following fatty acid conjugated substrates with increasing acyl chain length were purchased from Sigma-Aldrich (Munich, Germany) and used: *p*NP-acetate (C2), *p*NP-butyrate (C4), *p*NP-octanoate (C8), *p*NP-decanoate (C10), *p*NP-dodecanoate (C12), *p*NP-myristate (C14), *p*NP-palmitate (C16), and *p*NP-stenate (C18). In addition, *p*NP-ferulate (=4-Nitrophenyl trans-ferulate, Carbosynth,

Compton, UK) was used as a substrate. Prior to the activity tests, 10 mM stock solutions of the substrates in 100 % isopropanol were prepared. The activity tests were conducted in triplicates by incubating the respective FAE with 1 mM *p*NP substrate in 0.1 M PB Buffer (pH 7.8) for 20 min at 40 °C in microtiter plates. The increase in yellow color was quantified at 405 nm using a Synergy HT plate spectrophotometer (Biotek, Bad Friedrichshall, Germany). An enzyme free blank with the respective *p*NP-substrate was also measured and the absorption was corrected against the blank. The relative enzymatic activity was calculated separately for each enzyme. The substrate with the highest hydrolyzation rate (=highest absorption) was determined as 100 % relative enzymatic activity.

RESULTS AND DISCUSSION

Metagenomic library construction and functional screening

In order to identify potentially novel feruloyl esterases (FAEs) in the metagenome obtained from an agricultural biogas fermenter, a fosmid library encompassing approximately 273 Mbp of metagenomic DNA was constructed. The library consisted of 9,120 fosmid carrying clones which were stored in 95 microtiter plates and was used for a function based screening for feruloyl esterase activity (Table 1).

Table 1. Characteristics of the generated biogas fermenter fosmid library

Number of clones	9,120
Number of microtiter plates	95
Average insert size	~30 Kbp
Covered environmental DNA	~273 Mbp

In a first step, all clones of this fosmid library were tested for the ability to hydrolyze the short-chain fatty acid tributyrin. In total, 26 clones showed esterolytic activity. These clones were subsequently analyzed for feruloyl esterase activity using ethyl ferulate as a model substrate. Out of the 26 clones, 5 unique clones showed hydrolytic activity on this substrate (Table 2).

Table 2. Fosmid clones with hydrolytic activity on tributyrin plates and ethyl ferulate

No.	Microtiter plate	Clone	Activity on tributyrin	Activity on ethyl ferulate
1	6	G9	+	
2	9	C1	+	
3	12	D12	+	
4	12	E8	+	
5	14	H4	+	
6	17	H1	+	
7	18	F5	+	
8	18	H5	+	
9	22	H12	+	
10	32	C7	+	
11	40	E10	+	
12	46	E1	+	+
13	47	G6	+	+
14	47	H11	+	
15	50	B11	+	+
16	50	H1	+	
17	50	E8	+	
18	52	D9	+	
19	59	H4	+	
20	64	D11	+	
21	71	G2	+	
22	72	A9	+	
23	73	F9	+	+
24	73	H6	+	
25	74	D12	+	
26	79	H9	+	
27	88	H9	+	
28	92	F12	+	+

Characteristics and phylogenetic analysis of identified potential FAEs

The fosmids of these 5 clones were isolated, subcloned, and again tested for activity. The metagenomic inserts of active subclones were sequenced via primer walking. Subsequently, Artemis 16 was used for open reading frame prediction and the putative genes conferring the feruloyl esterase activity were identified by domain similarity to known esterase domains in the Pfam database (http://pfam.xfam.org/). The nucleotide sequences of the potential FAEs were translated to amino acids and their molecular weights, as well as their isoelectric points, were calculated using the Compute pI / Mw Tool (http://web.expasy.org/compute_pi/). In addition, the presence of an n-terminal signal peptide was predicted via SignalIP (http://www.cbs.dtu.dk/services/SignalP/). The results of these analyzes are shown in Table 3 and the corresponding coding sequences of the FAEs are listed in the appendix.

Table 3. Characteristics of the identified potential FAEs

No.	Potential FAE	Amino Acids	Predicted MW (kDa)	Predicted pI	Predicted signal peptide	Domain similarity (Pfam)
1	MTP46E1	381	41.8	4.88	yes	putative Esterase (PF00756)
2	MTP47G6	294	32.3	4.76	no	α/β hydrolase fold (PF07859)
3	MTP50B11	270	30.5	5.54	no	putative Esterase (PF00756)
4	MTP73F9	264	30.2	5.10	no	putative Esterase (PF00756)
5	MTP92F12	320	35.6	4.70	no	α/β hydrolase fold (PF07859)

The predicted sizes of the 5 enzymes range from 264 to 381 amino acids and molecular weights from 30.2 to 41.8 kDa. A potential signal peptide was only present in the potential FAE MTP46E1. In the sequences of MTP46E1, MTP50B11, and MTP73F9 putative esterase domains were identified. In MTP47G6 and MTP92F12 a generic α/β hydrolase fold was present. In a next step, the phylogenetic position of the five newly identifies enzymes in relation to known FAEs was examined. For this, a neighbor joining tree depicting the likely evolutionary history was calculated (Figure 1).

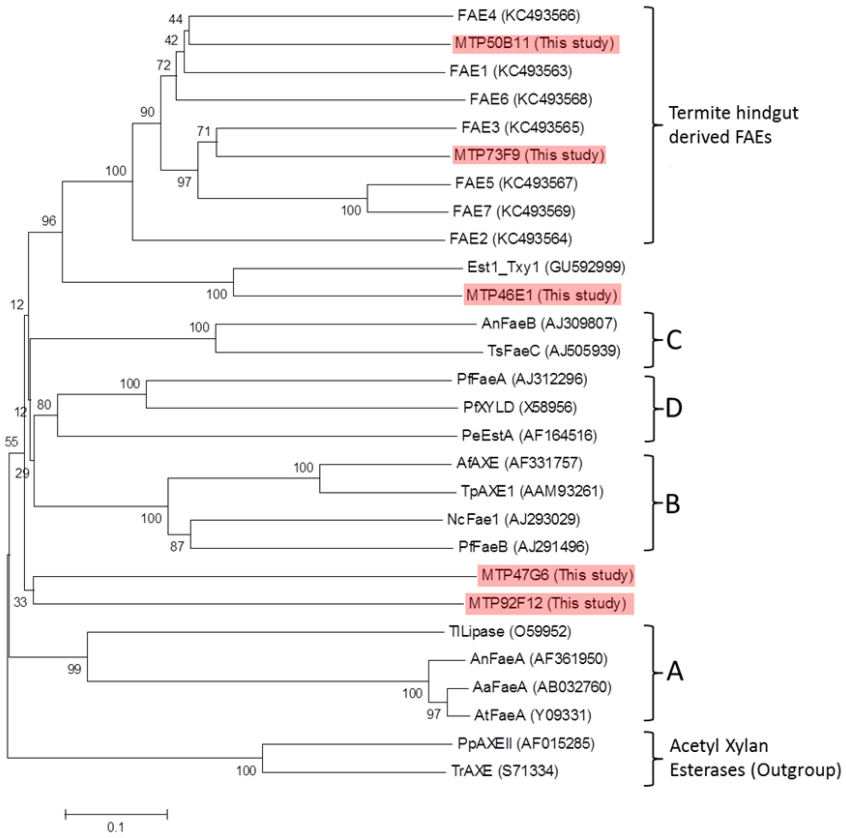

Figure 1. Phylogenetic neighbor joining tree among feruloyl esterases (FAEs) and relatives. Newly identified feruloyl esterases in this study are indicated red. Feruloyl esterase classification (A-D) according to Crepin *et al.*, 2004. Termite hindgut derived FAEs were identified and published by Rashamuse *et al.*, 2014.

Two of the five enzymes (MTP50B11 and MTP73F9) clustered together in a phylogenetic group with recently identified termite hindgut derived FAEs. These enzymes were identified in 2014 by Rashamuse and colleagues (Rashamuse *et al.*, 2014) in a metagenome obtained from a termite hindgut via a similar functional screening approach.

The enzyme MTP46E1 exhibited a close relation to a thermostable feruloyl esterase named Est1_Txy1 which was identified in the hemicellulolytic bacterium *Thermobacillus xylanilyticus*. In the original study, the purified Est1_Txy1 enzyme exhibited an optimal activity at 65 °C and a pH of 8.5 and was able to release ferulic acid from lignocellulosic biomass (Rakotoarivonina *et al.*, 2011).

The enzymes MTP47G6 and MTP92F12 showed a limited similarity to each other but did not demonstrate a substantial similarity to the FAEs included in the phylogenetic tree. Interestingly, none of the five enzymes clustered together with FAEs belonging to the subclasses A-D which were introduced by Crepin *et al.* (Crepin *et al.*, 2004) based on sequence identity and substrate utilization.

Recombinant expression and purification of MTP73F9 and MTP92F12

The enzymes MTP73F9 and MTP92F12 were chosen for recombinant expression in *E.coli* BL21 cells and subsequent purification via IMAC. For this, both enzyme encoding genes were ligated into a pET21a(+) expression vector in-frame with a 6x His-tag at the C-terminal end of the proteins. Both proteins were successfully expressed and were present in the soluble cytoplasmic fraction of the cell extract at the expected sizes (Figure 2).

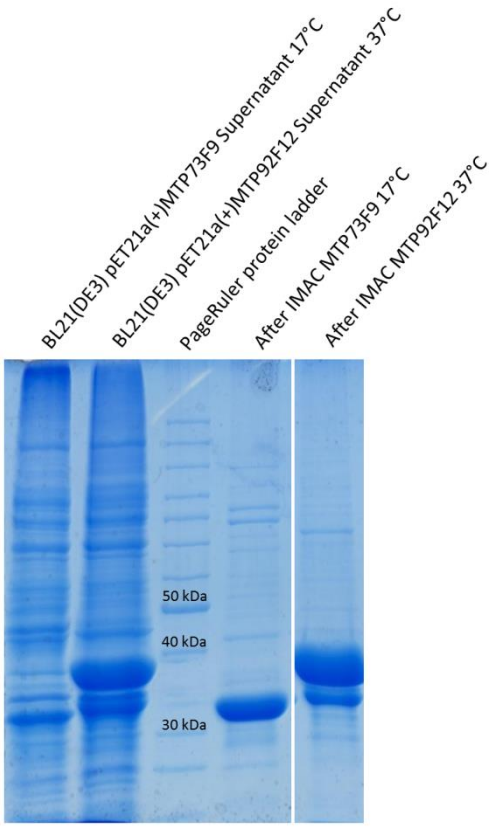

Figure 2. 10 % SDS-PAGE of the supernatant of IPTG induced *E.coli* BL21 expression cultures and purified enzymes after IMAC and coomassie blue staining.

The predicted sizes of the enzymes including the C-terminal histidine-tag were 33.2 kDa for MTP73F9 and 38.6 kDa for MTP92F12. Protein purification via IMAC resulted in a good recovery rate and a reasonably clean up. Interestingly, in addition to the expected signal at approx. 39 kDa for MTP92F12, an additional signal of a slightly smaller protein was also obtained. Even after repeated expression and cleanup this signal was present. It is likely that the smaller protein signal represents an isoform of the MTP92F12 protein or a partially degraded version of MTP92F12.

Initial biochemical characterization and future directions

The purified enzymes MTP73F9 and MTPF12 were used for an initial substrate range characterization. In addition, crude extract from the active subclone encoding MTP46E1 was tested. The enzyme activities against different *p*NP-acyl esters, as well as *p*NP-ferulate, were determined at 40 °C (Figure 3).

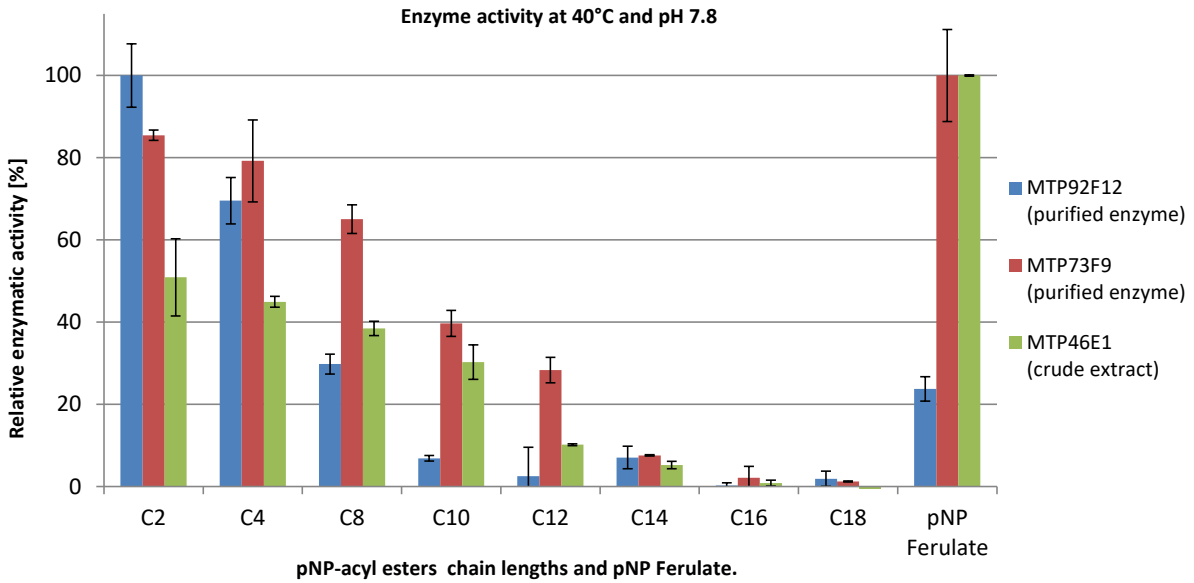

Figure 3. Relative enzyme activity of potential FAEs tested against *p*NP-acyl esters with C-chain lengths between C2 and C18 and *p*NP-ferulate. Activity tests were conducted separately for 20 min at 40°C in 0.1M phosphate buffer with a pH of 7.8.

MTP92F12 exhibited the highest activity against short-chain fatty acid substrates and only minor activity against *p*NP-ferulate. This result suggested that MTP92F12 might be a fatty acid degrading carboxylesterase rather than a true feruloyl esterase. This enzyme was probably identified in the functional screening due to a certain enzyme promiscuity and a resulting side activity on ferulic acid related substrates like ethyl ferulate and *p*NP-ferulate. This assumption might also explain the limited similarity of this enzyme to known FAEs in the phylogenetic analysis.

In contrast, MTP73F9 and MTP46E1 exhibited the highest relative activity using *p*NP-ferulate as substrate. Furthermore, they displayed activity against short-chain fatty acid substrates. The hydrolytic activity of both enzymes decreased with increasing fatty acid acyl chain length of the substrate. Due to the highest observed relative activity on *p*NP-ferulate, it is likely that these enzymes are true feruloyl esterases.

In a next step, the specificity of these enzymes against additional FAE model substrates and their ability to release ferulic acid from plant cell walls has to be evaluated. If the newly identified FAEs are capable of efficiently releasing ferulic acid or related hydroxycinnamic acids from lignocellulose, these enzymes might be of interest for an application in the production of value added compounds from agricultural biomass. Furthermore, the examination of possible synergistic effects between the FAEs and lignocellulolytic enzymes in the degradation of plant biomass might be a direction for further research. If such effects

can be observed, these enzymes might also be suitable candidates to improve the effectiveness of enzyme mixtures designed for the degradation of lignocellulose.

Finally, due to the relationship of MTP46E1 to the recently identified thermostable FAE Est1_Txy1, the optimal reaction temperature of this enzyme and a possible thermostability are further targets for investigation. In case MTP46E1 is indeed thermostable, this enzyme might be especially interesting for industrial applications as many of the relevant processes are conducted at higher temperatures. A possible enzyme structure prediction for MTP46E1 based on *in silico* homology modeling is shown in Figure 4.

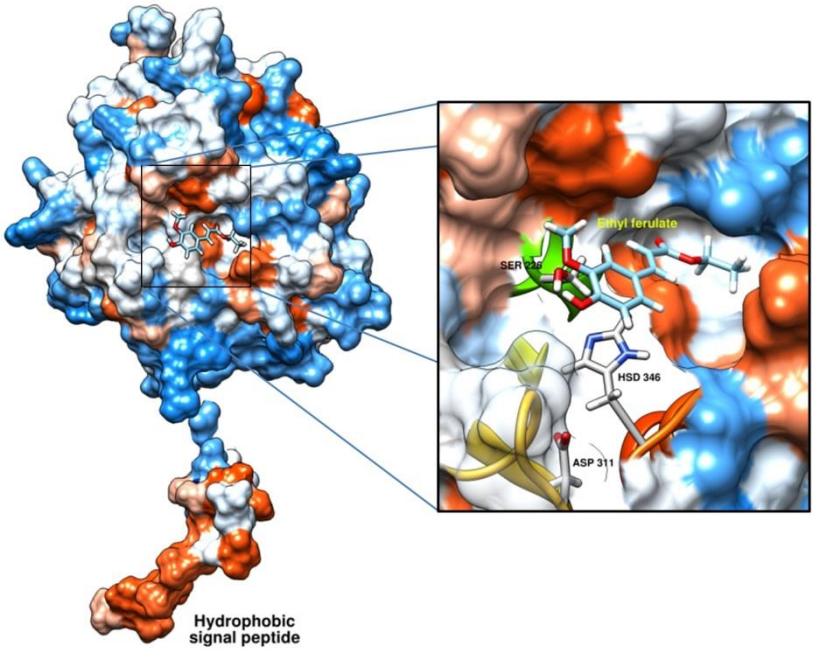

Figure 4. Predicted enzyme structure for MTP46E1. The structure prediction is based on homology modeling and was performed using the Robetta server (Kim, Chivian et al. 2004). The protein surface is colored according to hyprophobicity (hydrophobic regions shown in red, hydrophilic regions in blue). A possible interaction of an ethyl ferulate molecule with the active site of the enzyme is shown. For the prediction of the ligand binding, the SwissDock web service was used (Grosdidier, Zoete et al. 2011). The amino acids forming the catalytic triad Ser(226)-His(346)-Asp(311) are indicated.

The amino acids forming the catalytic triad were identified based on sequence homology to the previously described catalytic residues of Est1_Txy1 (Rakotoarivonina, Hermant et al. 2011). The 17 amino acids comprising hydrophobic N-Terminal signal peptide of this enzyme was included in the modeling. A possible interaction of an ethyl ferulate molecule with the active site of MTP46E1 is shown.

Based on the presence of a signal peptide, a high relative activity on *p*NP-ferulate, and a potential thermostability, the newly identified FAE MTP46E1 might represent the most promising candidate for further examination and a potential industrial application.

APPENDIX

Coding nucleotide sequence of MTP46E1 (1146 nt):

ATGCTTGTCGTGTTGGTACTGGGAGTTTCGTCTATCGCTCTGGCGAATGCAGTGGAAAGTGG
GGGAGAACAGGTGACAATCGAGTTAGGGCACATACCGGAAGACTACCGTAAACCTCTGAAAG
AGGTGGGCCAGATTGGCCGCTTAGTGCAGGTTTACTACGATGTCTACAATTACATTAACGAG
GACCGTCAGCTTGTCTCCAACCAGAACATCAGCCCGGAAGAAGCGGGGCGCGAAGTGGTTAC
AGGTGACCCAATCGAGAAACGGTTCCAGATCTATCTACCTCCAGGCTACCGTGATGAGGATA
AGGAGACTAAATACGACGTACTCTATCTGCTCCATGGGGTAGGCGGTACAAGCTACGAATGG
CCTAACGGCGCCACCTATCTGGGTACCTACGTGATCTGCAACCTTCTGGATCACCTGATCCT
GAATGGCGAAATCGAACCGGTAATCGTGGTGTTCCCCGAGGGCAGAAGCTCACACGATTGGA
AAGACACATCTTTCTCAGCGGATAAGACCAATATCCTTGGCTTTTACTACTTGGACTACGAG
CTGAGATACGATCTGATTCCCTACGTTGAGGCTAATTTCAATACAAAAGCGGATATCAGCGA
CAGGTCACCGGAAGGCATTGCCAAGAACCGTAAGCACCGGGCAATTGGCGGGCTTTCCATGG
GCGGTATGCAGACCTTGAACCTGATTATCGGTGGGTACAGAAGTGACTTCGGTTCCTTCGCA
GGGGCACCGGGGAATCCCGGTAATGGGCTAGCCGGCACTGTTAAAGCACCGGGCATGCTCGA
TCTTTTTGCCCATGTGGCTTCCTTCTCGCCCGCACCCACCTCAAGCGATGGGCGGACATTGG
GTGAGAGCATTGCCGCCTCGCCTTACAAAATCGATCTGCTCTACTTAACCTGTGGCGATGCA
GACGGCACCTCGTTACACGCCTACAACTCTGCCATCGACGGGCTGCGGGAGATGGCGGGGGA
TAAGATCGGCGCCTTTTATCAGGTGACGATGGAACGCCGTGGACATGACTTCAGTGTCTGGA
ACAACGGAGCTTACAACTTCCTTCGGCTCAGCTTCGGTAGGACTCCGGCTGACCCGGCAGAC
ATTGCCATAACTTTGGAATACTCCAGATAG

Amino acid sequence of MTP46E1 (381 aa):

MLVVLVLGVSSIALANAVESGGEQVTIELGHIPEDYRKPLKEVGQIGRLVQVYYDVYNYINE
DRQLVSNQNISPEEAGREVVTGDPIEKRFQIYLPPGYRDEDKETKYDVLYLLHGVGGTSYEW
PNGATYLGTYVICNLLDHLILNGEIEPVIVVFPEGRSSHDWKDTSFSADKTNILGFYYLDYE
LRYDLIPYVEANFNTKADISDRSPEGIAKNRKHRAIGGLSMGGMQTLNLIIGGYRSDFGSFA
GAPGNPGNGLAGTVKAPGMLDLFAHVASFSPAPTSSDGRTLGESIAASPYKIDLLYLTCGDA
DGTSLHAYNSAIDGLREMAGDKIGAFYQVTMERRGHDFSVWNNGAYNFLRLSFGRTPADPAD
IAITLEYSR

Coding nucleotide sequence of MTP47G6 (885 nt):

ATGTTTTGCAGTGAAGCAGACGGCACGGCAATCGCCAGTCGCTGTATGGGCAAGGGGTTCGGTGTAAC
TTGTACATATTTGTATCCTGTAGCCAAAGATTACCGTTTCCCGCAGGTTGTTATTGATCTCATCAAAG
GTATCAAAATCTTGCGTGATCATGCAGAAGAATGGGGAATTAATCCGGACCAGATTGTCATCTCAGGC
AATTCTGCCGGGGCCTTTATCTGCATGACCACCGGTAATATTTGGAACAGGTCGGAGATCATGGAAGC
TGCCGGCTGTACCGGTGAAGAGGGCAAGCCTAATGCCATGATTTTGGGTTTTGGTCCCATGTTTTGCG
GCCAGCAGACCGATGACGGCATTGTTTATGTACCCAATGGTGACCTGGTCGGTCCGCAAACACCGCCT
GCATTTTTCCATCATACGCGTCTTGATATGCTTGTTTCCGTTTATCAGACAATTGCCATGATTGATGC
TATGGAGCGGGCCAAACGCCCTTATGCGGCATATATCTCCAGTACGGGAGGACATGGGGAGACTAATG
CTGTAAACCGCATATTGGCCGAAGACGGTACAGTCGGTCCCTGCATTGATGATTGGTTTGAACAGGCG
TGGCGTTTCCTGGCAAACCAGCTGGGTATTTCACTCATGCCACAGAAGATGCCAATGATGCCGCCTAT
GCCGCCGATGCAGGAAGGGGATGAATTCCCGCAATTCCCCATGCCCATGTTTACGCCGCCGCCTGTTC
CTGAAGGCAGCGTCCCGGTTAAGCCTGAAGAAATGCCATTAGGCCAAGCCGATAACATTCATATGCCT
TTTAACGCCGCTTTCCGTGATAAAGACTTTAAGGTATTTAAAATCACGAATTCTGGATCCGATACGTA
A

Amino acid sequence of MTP47G6 (294 aa):

MFCSEADGTAIASRCMGKGFGVTCTYLYPVAKDYRFPQVVIDLIKGIKILRDHAEEWGINPD
QIVISGNSAGAFICMTTGNIWNRSEIMEAAGCTGEEGKPNAMILGFGPMFCGQQTDDGIVYV
PNGDLVGPQTPPAFFHHTRLDMLVSVYQTIAMIDAMERAKRPYAAYISSTGGHGETNAVNRI
LAEDGTVGPCIDDWFEQAWRFLANQLGISLMPQKMPMMPPMPPMQEGDEFPQFPMPMFTPPP
VPEGSVPVKPEEMPLGQADNIHMPFNAAFRDKDFKVFKITNSGSDT

Coding nucleotide sequence of MTP50B11 (813 nt):

ATGGCGGTTTTTGATATTACCTTCTATTCAAACTCCCTGAACCGGCACACCCAGCTGACGGCAATAAT
CCCTATTGACATGCCGGAATTTCCCGGCTTCCCCAAACCCGATAAAAACAAGCTCTTTAAGTCTCTTT
ACCTTCTGCACGGCTATTCCGGCAACCAGAGCGACTGGCTGCGCGGTACGCAAATTGAGCAGTTCGCA
ATAATGCATCAAATTGCCGTCTTTTGTCCTGCCGGCGAGAACAGCTTCTATCTTGACGACAAGGCGCG
GGGCGCCATGTACGGGGAGATGATCGGCCGGGAGCTGATTGAGTTTACCCGCAGAATATTCCCCTTGT
CTGCCGAAACCAAAGATACGTCTATCGGCGGCCTGTCCATGGGCGGTTACGGCGCCATTCGCAACGGC
TTTAAATACAACGATGTCTTCGGCAGCATAATCGCCCTGTCTTCCGCCCTGATTAATGACAATGTGGC
CAAGGGCTTTGAGCAGGAAAACAACCCGATGATATCTTCCGGCTATTTTATGCATACCTTTGGGGAGC
CCGGCAAAATCAAGGGCAGCGACATTGATCCCGAAGCATTGGCCAGGAAGCTGGTTGACAGCGGCACT
GCAAGGCCCAGGCTCTACATGGCATGCGGCACCGAGGATTTTCTTATTAATGAGAACCGGAAATTCCA
CAAGTGCCTCTTGGAAATGGGATATGAGCATACATACATAGAGGCTCCCGGCATGCACAGCTGGGAAT
TCTGGAATGAATACATTGAAAAGGCAATGGTCTGGCTTACAGGGCAGGACAGGCCGCAAGCCTGA

Amino acid sequence of MTP50B11 (270 aa):

MAVFDITFYSNSLNRHTQLTAIIPIDMPEFPGFPKPDKNKLFKSLYLLHGYSGNQSDWLRGTQIEQFA
IMHQIAVFCPAGENSFYLDDKARGAMYGEMIGRELIEFTRRIFPLSAETKDTSIGGLSMGGYGAIRNG
FKYNDVFGSIIALSSALINDNVAKGFEQENNPMISSGYFMHTFGEPGKIKGSDIDPEALARKLVDSGT
ARPRLYMACGTEDFLINENRKFHKCLLEMGYEHTYIEAPGMHSWEFWNEYIEKAMVWLTGQDRPQA

Coding nucleotide sequence of MTP73F9 (795 nt):

ATGGCGTTATTTCAAGTAGATCTATTCTCAAAATCCCTAGAAAGGAAAACCTGCTTTCATATGGTGAT
ACCTAACGATGTTCCACCTATGATGATTGAAGGAAATGATAATTATAATAGAAAGATGAAGACCCTAT
TTTTACTTCACGGATACTCAGGCTCAAGTAACGATTGGTTATTGGGAAGCCCTGTACAAGACTTAGCT
ACAAAATATAATATGGCTATATTGATGCCATCTGGTGATAATAGCTTTTATCTTAATGGAATAGGTAA
GGGCCGGGCATATTGTAGGCTTGTAGGAGAGGAACTTGTTGACTATATAAGAAGAACATTTAATTTAG
CCATAGATAAAGAGGACACTTTTATAGCTGGGCTTTCTATGGGTGGCTTCGGAGCTATTCACACTGGC
TTATATTATCCTGAGACATTTAGCAAAATAGTTGGTCTTTCGTCTGCCCTTATAATTCATAATATAAA
AAACAAAAAAGAAGGTTTTGAAGATCCTATAGCAGATTACGGATATTACTATTCTTGTTTTGGGGACC
TTGATAAACTTGAAGAGAGTGAGAACAATCCTGAGTATTTAATCAAAGCATTAAAAAGGGATAATAAA
CCCATACCACAAATGTATATAGCCTGTGGAACAGAGGATTTTCTTATAGAAGAAAACAGGGCCTTTCA
TAAATTCCTACAAAATGAGGATGTAGCTGTAGAATATATTGAAAGTCCTGGAACACATGATTGGGCCT
TCTGGAATCATTATCTAGAGCCCTCAATAAAATGGATGCTTAGCTAA

Amino acid sequence of MTP73F9 (264 aa):

MALFQVDLFSKSLERKTCFHMVIPNDVPPMMIEGNDNYNRKMKTLFLLHGYSGSSNDWLLGS
PVQDLATKYNMAILMPSGDNSFYLNGIGKGRAYCRLVGEELVDYIRRTFNLAIDKEDTFIAG
LSMGGFGAIHTGLYYPETFSKIVGLSSALIIHNIKNKKEGFEDPIADYGYYYSCFGDLDKLE
ESENNPEYLIKALKRDNKPIPQMYIACGTEDFLIEENRAFHKFLQNEDVAVEYIESPGTHDW
AFWNHYLEPSIKWMLS

Coding nucleotide sequence of MTP92F12 (963 nt):

ATGCTGCAGACGCGTTACGTATCGGATCCAGAATTCGTGATTCCCGATCCTTCAAAACCTATGGAGTT
GCCCTTGCTGGATGTAAGTGGCATTGAGCGCAAATTTCTTGATATTTCCTATACGCCGGACAATCCCC
ACCCGATGCGCATGCTGGATATCTACCTTCCAGATGAGGGCGAGGGTCCTTTCCCGACGATTATTTAC
CAGCACGGCGGCGCGTTTATTGCCGGAAATAAACGTGATTTCCAGGCTGCCGGGACTCTGGAGGCTAT
TAACAATGGCTTTGCTGTTGTCAGCGTAGAACAGCGACTGGCAACCATGGATATGAACGGAAAGGCCA
ATTCCGAGGGGTTGTTTCCCTACCCGCTGTTTGATTTCAAAGCGGCAATCAGATTTCTGCGGGTAAAC
GCCGCTAAGTATAAGCTTGACCCGGATCGTTTTGCCACATGGGGCGACTCGGCAGGCGGTTACCATGC
GGTAATGGCCGCGCTGACCCAGGATGTGCCCTTCATGTACGATCCCTCTTTGGGCTTTGGGGATATAA
GTGGCAAGGTGCAGGCGGTTGTCAGCTGGTTTGGAGTGGGCGATTTGGTTTTGCAATCAGAATTCACC
GATAACCAGCCGCCCATGGTGGGGCCGGACGGAAAAGAGTATCCTAACTTGAATTACGCGGATGTTTT
TCTCGGCGTGAAGGCAACCGAGCACAAAAACCTGGCCTACTTTGCCAATCCGGAGACCTGGGTCAATC
CTTCCATTCCTCCGGTACTGCTGCAGGCGGGAATTGCCGACGAGGTGGTGCCTTTCGGGTGCTCAGAG
AATCTGGCAAAGCGTATTGAGGAGGTCTGCGGCAAGGACCGTGTGACTCTGGACGCCTTTGAAGGATA
TACCCACGGGGATATGCGTTTTAACGAGCCAGAGAATCTTGCCCGGGTATTTAAGTGGCTCAGGGAAA
AGCTGAAATAG

Amino acid sequence of MTP92F12 (320 aa):

MLQTRYVSDPEFVIPDPSKPMELPLLDVSGIERKFLDISYTPDNPHPMRMLDIYLPDEGEGPFPTIIY
QHGGAFIAGNKRDFQAAGTLEAINNGFAVVSVEQRLATMDMNGKANSEGLFPYPLFDFKAAIRFLRVN
AAKYKLDPDRFATWGDSAGGYHAVMAALTQDVPFMYDPSLGFGDISGKVQAVVSWFGVGDLVLQSEFT
DNQPPMVGPDGKEYPNLNYADVFLGVKATEHKNLAYFANPETWVNPSIPPVLLQAGIADEVVPFGCSE
NLAKRIEEVCGKDRVTLDAFEGYTHGDMRFNEPENLARVFKWLREKLK

REFERENCES

Benoit I, Danchin EGJ, Bleichrodt RJ & de Vries RP (2008) Biotechnological applications and potential of fungal feruloyl esterases based on prevalence, classification and biochemical diversity. *Biotechnol Lett* **30**: 387-396.

Carver T, Harris SR, Berriman M, Parkhill J & McQuillan JA (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**: 464-469.

Crepin VF, Faulds CB & Connerton IF (2004) Functional classification of the microbial feruloyl esterases. *Appl Microbiol Biotechnol* **63**: 647-652.

Faulds CB (2010) What can feruloyl esterases do for us? *Phytochem Rev* **9**: 121-132.

Grosdidier, A., V. Zoete and O. Michielin (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research* **39**: W270-W277.

Güllert S, Fischer MA, Turaev D*, et al.* (2016) Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies. *Biotechnology for Biofuels* **9**: 1-20.

Kim, D. E., D. Chivian and D. Baker (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research* **32**: W526-W531.

Koseki T, Fushinobu S, Ardiansyah, Shirakawa H & Komai M (2009) Occurrence, properties, and applications of feruloyl esterases. *Appl Microbiol Biot* **84**: 803-810.

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Notredame C, Higgins DG & Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* **302**: 205-217.

Rakotoarivonina H, Hermant B, Chabbert B, Touzel JP & Remond C (2011) A thermostable feruloyl-esterase from the hemicellulolytic bacterium Thermobacillus xylanilyticus releases phenolic acids from non-pretreated plant cell walls. *Appl Microbiol Biot* **90**: 541-552.

Rashamuse K, Ronneburg T, Sanyika W, Mathiba K, Mmutlane E & Brady D (2014) Metagenomic mining of feruloyl esterases from termite enteric flora. *Appl Microbiol Biotechnol* **98**: 727-737.

Saitou N & Nei M (1987) The Neighbor-Joining Method - a New Method for Reconstructing Phylogenetic Trees. *Mol Biol Evol* **4**: 406-425.

Tamura K, Stecher G, Peterson D, Filipski A & Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol* **30**: 2725-2729.

Williamson G, Kroon PA & Faulds CB (1998) Hairy plant polysaccharides: a close shave with microbial esterases. *Microbiol-Uk* **144**: 2011-2023.

Wong DW (2006) Feruloyl esterase: a key enzyme in biomass degradation. *Appl Biochem Biotechnol* **133**.

Wong DWS, Chan VJ, Batt SB, Sarath G & Liao H (2011) Engineering Saccharomyces cerevisiae to produce feruloyl esterase for the release of ferulic acid from switchgrass. *J Ind Microbiol Biot* **38**: 1961-1967.

7 Discussion

The combination of high-throughput next generation sequencing with cultivation independent metagenome- and metatranscriptome-based analyses has significantly improved our knowledge of the complex microbial communities present in our environment, in biotechnological systems or in host associated habitats like the gastrointestinal tract of animals.

In this study, the recent progress in sequencing throughput together with improved bioinformatic tools for the analysis of large metagenomic datasets was utilized to study and compare the degradation of lignocellulose by bacteria in two herbivorous animal associated habitats and an agricultural biogas fermenter. Within this framework, a metagenome based analysis of the fecal microbiota of zoo elephants revealed a surprisingly high abundance and diversity of genes encoding glycoside hydrolase family enzymes. Many of the identified GH families were involved in the degradation of lignocellulosic biomass (Ilmberger, Güllert et al. 2014). Due to this finding, the dataset derived from the feces of the adult elephant was included in a detailed comparative analysis of the cellulolytic activity and the metagenomic cellulolytic potential between three different biomass degrading systems. Next to the elephant metagenome, a cow rumen associated metagenome and an agricultural biogas fermenter derived metagenome were analyzed and mayor differences in carbohydrate hydrolysis strategies were identified (Güllert, Fischer et al. 2016).

7.1 Composition of the microbial community in the three analyzed systems and the ratio of the phyla *Firmicutes* vs. *Bacteroidetes* in biogas fermenters and herbivorous animals

The microbial community structure in the agricultural biogas fermenter, the cow rumen, and the adult elephant feces sample was analyzed based on phylogenetic marker genes in the assembled metagenomic datasets and, in part, on 16S rRNA gene amplicons. Most notably, a comparison of the bacterial marker genes in the three datasets showed a higher amount of marker genes affiliated with the phylum *Firmicutes* in the biogas fermenter sample while the

marker genes affiliated with the phylum *Bacteroidetes* were decreased (Figure 1D, Güllert, Fischer et al. 2016). Based on the metagenomic marker genes, the ratios of the phyla *Firmicutes* vs. *Bacteroidetes* in the three samples were calculated. The resulting ratios were 5.2:1 for the biogas fermenter, 1.5:1 for the cow rumen, and 1.6:1 for the elephant feces dataset.

Intrigued by the difference in the ratio of these two bacterial phyla between the biogas fermenter and the two analyzed natural cellulolytic systems, the question arose if comparable ratios would be present in other studied biogas fermenters and herbivores. To answer this question, published studies of mainly agricultural biogas fermenters and various herbivorous animal associated microbiomes were examined and the ratios of the phyla *Firmicutes* vs. *Bacteroidetes* were calculated based on the data in the original studies (Table 2, Güllert, Fischer et al. 2016). By this, it became apparent that similar high ratios were indeed present in other biogas fermenters (Jaenicke, Ander et al. 2011, Wirth, Kovacs et al. 2012, Zakrzewski, Goesmann et al. 2012, Stolze, Zakrzewski et al. 2015) and also substantially lower ratios in other herbivorous animal associated microbiomes (Hess, Sczyrba et al. 2011, Pope, Mackenzie et al. 2012, Bian, Ma et al. 2013, Henderson, Cox et al. 2013, Ilmberger, Güllert et al. 2014, Roggenbuck, Sauer et al. 2014, Zeng, Han et al. 2015). As a result, a distinct difference in the structure of the microbial community in multiple biogas fermenters compared to natural biomass degrading microbiomes was identified.

It can be speculated that bacteria of the *Firmicutes*, especially the *Clostridia*, are very well adapted to the conditions in agricultural biogas fermenters while bacteria of the *Bacteroidetes* are confronted with some pressure limiting their growth. This is in particular interesting as biogas fermenters are commonly inoculated and fed with various animal manures which are rich in bacteroidetal organisms. Therefore, it is likely that the *Bacteroidetes* are initially present in higher abundance in biogas fermenters and are then outcompeted by other groups like e.g. the *Clostridia*.

Given these findings, bacteria of the phylum *Bacteroidetes* seem to be considerably better adapted to natural habitats than to artificial biogas fermenters. The reasons for this remain unknown. However, it can be assumed that certain factors are lacking in the fermenter which are required for an optimal growth of the *Bacteroidetes*. In natural microbiomes, these factors might be provided by the eukaryotic host or other gut microbes. A candidate for such a factor might be cobalamin (vitamin b12) and its analogs. It recently became

apparent that many *Bacteroidetes* genomes are missing the biosynthesis pathway for cobalamin (Magnusdottir, Ravcheev et al. 2015). However, for the majority of gut bacteria vitamin B12 is an essential coenzyme in the anaerobic fermentation process (Martens, Barg et al. 2002). It is possible that the loss of this pathway in various *Bacteroidetes* genomes might resemble an adaption to the highly competitive intestinal environments. There are hints that some members of the *Bacteroidetes* have developed highly effective vitamin B12 uptake systems (Degnan, Barry et al. 2014, Cordonnier, Le Bihan et al. 2016). It is tempting to speculate that these bacteria might conserve energy by omitting a synthesis of this coenzyme, and instead rely on vitamin B12 produced by other gut microorganisms. While this strategy might be beneficial in the gut, it might be a competitive disadvantage in habitats like biogas fermenters.

In the next paragraph, possible consequences for the degradation of lignocellulose that result from the high *Firmicutes vs. Bacteroidetes* ratio in the biogas fermenter will be discussed.

7.2 Decreased cellulolytic activity and lower abundance of predicted glycoside hydrolase and carbohydrate esterase family enzymes in the biogas fermenter compared to the analyzed natural microbiomes

One could argue that an increased abundance of the *Clostridia* in biogas fermenters might be beneficial for the degradation of plant biomass due to the many cellulosome producing cellulolytic bacteria in this class. However, if this was the case, the question arises why a considerably lower ratio has evolved in natural host associated microbiomes. Given the assumption that herbivorous animals rely on a most efficient breakdown of their plant based diet, their associated microbiomes should have evolved to be highly effective in degrading lignocellulose.

In order to address this supposed discrepancy, the cellulolytic potential of the different microbial communities was evaluated using different approaches. At first, the total endoglucanase enzyme activity in various fecal samples and the biogas fermenter sample was determined via a DNS-Assay. This assay showed that the ability to degrade carboxmethlycellulose (CMC) was significantly lower in the biogas fermenter sample as

compared to feces samples obtained from an adult elephant, a cow, a mara, and a zebra (Additional file 7, Güllert, Fischer et al. 2016). In all measurements, the sample obtained from the biogas fermenter had an approximately 3- to 4-fold lower endoglucanase activity. Even though CMC as a model substrate cannot adequately reflect the diversity and complexity of natural lignocellulosic biomass, this finding indicated that there are distinct differences in the cellulolytic activities between the biogas fermenter and the natural habitats.

In addition to the DNS-Assay, the comprehensive assembled metagenomic datasets obtained from the biogas fermenter, the cow rumen, and the adult elephant feces were used for a comparative in-depth analysis of the abundance and diversity of genes encoding potential GHs and CEs family enzymes. This analysis indicated a lower abundance of GH and CE family enzymes per Mbp of assembled DNA in the biogas fermenter compared to the natural microbiomes (Table 4, Güllert, Fischer et al. 2016). When the analysis was restricted to GH families with cellulolytic activity, the richness of potential enzymes was still decreased in the fermenter.

In a next step, the phylogenetic affiliation of the identified putative cellulolytic GH enzymes was examined to identify the phylogenetic groups which contributed most to the cellulolytic potential. This analysis confirmed that especially enzymes derived from the *Bacteroidetes* were underrepresented in the biogas fermenter in relation to the elephant and cow datasets (Figure 2A, Güllert, Fischer et al. 2016). When considering only the cellulolytic GH families, the ratio of enzymes affiliated with the *Firmicutes* vs. *Bacteroidetes* was 2.8:1 in the fermenter. This ratio is almost three-fold higher than the ratio found in the elephant feces metagenome and two-fold higher than the ratio found in the cow rumen associated metagenome. Similar different ratios were obtained for the CE family enzymes. These results demonstrated that a lower abundance of members of the *Bacteroidetes* in the biogas fermenter was associated with a decreased richness of predicted enzymes involved in the degradation of carbohydrates and might provide an explanation for the different activity profiles observed in the DNS-assay.

An important limitation of this analysis was that the comparison of the genes encoding GH and CE family enzymes was performed on a metagenomic level and therefore did not provide any insight into the actual transcription level of the identified genes. To address this limitation, metatranscriptome sequencing (RNA-Seq) was performed for the biogas and the

elephant feces samples. The results obtained from this analysis will be discussed in the next chapter.

7.3 Metatranscriptome sequencing (RNA-Seq) identifies differences between highly transcribed genes encoding GH family enzymes in the biogas fermenter and the elephant feces sample

In order to assess the transcription levels of the previously identified GH encoding genes, the metagenomic datasets generated from the biogas fermenter and the elephant feces sample were complemented with data obtained from metatranscriptome sequencing. By including these data, the GHs which were previously identified in the metagenomic datasets could be sorted according to their transcript levels. It was now possible to examine and compare the phylogenetic affiliation of highly transcribed GHs between the samples (Figure 4a, Güllert, Fischer et al. 2016). When the phylogenetic affiliation of the 50 cellulolytic GHs with the highest numbers of mapped cDNA reads was compared between the biogas fermenter and the elephant feces sample, a distinct difference between the samples was found. The ratio of highly transcribed *Firmicutes* derived enzymes versus *Bacteroidetes* derived enzymes was 1:1 in the elephant feces and 4:1 in the biogas fermenter sample. This result confirmed the divergent phylogenetic affiliation of cellulolytic enzymes in the two metagenomic datasets on a transcriptional level and also showed that the *Bacteroidetes* derived cellulolytic GHs were actively transcribed. Due to the higher amount of highly transcribed *Bacteroidetes* derived enzymes in the elephant gut, it can be speculated that the *Bacteroidetes* (and their enzymes) are equally important as the *Firmicutes* for the degradation of lignocellulose. In contrast, in the biogas fermenter three times more of the highly transcribed cellulolytic GHs were affiliated with the *Firmicutes.*

Furthermore, the largest group of highly transcribed cellulolytic enzymes in the elephant feces sample belonged to the CAZy GH51 family (Figure 4b, Güllert, Fischer et al. 2016). The particularly high abundance of *Bacteroidetes* derived GH51 family enzymes in the elephant feces dataset was already remarkable in the metagenome based analysis. The data obtained from RNA-Seq now confirmed that many of the genes encoding GH51 family enzymes were highly transcribed. Next to cellulases, the GH51 enzyme family comprises many

hemicellulases like xylanases and L-arabinofuranosidases (Eckert and Schneider 2003). The high abundance and expression level of hemicellulases is most likely an adaption to the rather diverse diet of the zoo animal and might be one of the key factors to efficiently degrade a wide variety of lignocellulosic biomass.

Based on these data, in the biogas fermenter most of the highly transcribed cellulolytic GHs were assigned to the CAZy GH9 family which contains mostly endo- and exoglucanases specific for cellulose polymers (Berlemont and Martiny 2013). Many of the GHs in this family are affiliated with the genus *Clostridium* including cellulases found in *Clostridium thermocellum, Clostridium cellulolyticum, Clostridium cellulovorans, Clostridium acetobutylicum,* and others.

In summary, and considering the reduced cellulolytic activity profile observed for the biogas sample in the DNS assay, it can be hypothesized that an overrepresentation of *Firmicutes (Clostridium)* derived cellulases, together with a possible lack of highly transcribed hemicellulases, might in fact actually be disadvantageous for the overall hydrolytic performance of the biogas fermenter.

7.4 The combination of metagenome and metatranscriptome sequencing allows an efficient identification of relevant biocatalysts in a sample and reveals metabolically highly active groups

In addition to comparative analyses between different samples, the combined sequencing of the metagenome and the metatranscriptome of a specific sample provides valuable additional insight for the identification of highly active biocatalysts like e.g. cellulases. Given the very high number of up to several thousand potential cellulase encoding genes commonly identified by sequence based screening methods in comprehensive metagenomic datasets, additional information is required to select promising candidates for heterologous expression and biochemical characterization. Data obtained from RNA-Seq can be used to discover enzymes (e.g. glycoside hydrolases) which are highly transcribed in a sample under the given conditions. If samples are chosen correspondingly to the enzymes of interest, e.g. gut microbiome samples of herbivorous animals for cellulases, plastic biofilm samples for PET esterases, hot springs for thermostable enzymes, etc., it is reasonable to assume that

highly expressed enzymes in the sample represent the most relevant candidates for industrial biocatalysts. This approach can be combined with an additional enrichment step which reduces the diversity in the sample and by this reduces sequencing costs and facilitates the bioinformatic analysis.

For example, in the metagenomic dataset generated from the feces of an adult elephant, a very high abundance and diversity of genes encoding potential GHs was present (Ilmberger, Güllert et al. 2014). In total, up to 20,000 potentially relevant genes were identified; many of these were full-length genes. In order to choose promising cellulolytic enzyme candidates for future examination and heterologous expression from this large pool of genes, RNA-Seq was performed and the resulting cDNA reads were mapped to the previously identified genes encoding putative GH family enzymes (Table 1).

Table 1. Identified potential GH family enzymes in the elephant feces dataset with the highest numbers of overall mapped cDNA reads.

Putative enzyme	CAZy GH family	Full-length?	Signal peptide?	Number of mapped cDNA reads
EFN4HT0G_1	GH5	Y	Y	4,925
EFN4H7X5_24	GH53	Y	Y	2,086
EFN4HTZ3_3	GH11	Y	Y	860
EFN4HCHT_18	GH5	Y	Y	780
EFN4HTTS_2	GH10	Y	Y	710
EFN4HDP9_16	GH74	Y	Y	692
EFN4H8CT_19	GH31	Y	Y	602
EFN4HTZ3_2	GH11	Y	N	569
EFN4HG20_6	GH43	Y	Y	552
EFN4HRM5_7	GH9	Y	N	524
EFN4H85S_46	GH51	Y	Y	485
EFN4HKWL_2	GH43	Y	Y	469
EFN4HP8H_12	GH5	Y	Y	450
EFN4HRRD_3	GH2	Y	Y	438
EFN4HRH8_1	GH11	Y	Y	432
EFN4J3KJ_4	GH9	N	Y	374
EFN4M2C0_1	GH13	Y	N	367
EFN4HDJ7_6	GH3	Y	N	359

Using this approach, the putative cellulase “EFN4HT0G_1” assigned to the CAZy GH5 family was identified as a highly attractive candidate for a detailed future analysis. At the time of sampling, the corresponding gene was the glycoside hydrolase encoding gene with the highest transcript level in the elephant feces metagenome. The encoded enzyme possesses

an N-terminal signal peptide, a GH5 cellulase domain, a carbohydrate binding module, and a por secretion system C-terminal sorting domain. Furthermore, the enzyme displays only a low similarity (<40%) to other cellulases in the NCBI protein database. After the successful identification of this gene, the cellulolytic performance and the biochemical characteristics of the corresponding enzyme will be evaluated in future studies.

In addition to the elephant sample, RNA-Seq was also performed for the studied biogas fermenter. After mapping of the cDNA reads to the metagenomic dataset, additional highly transcribed GH candidates for further characterization were identified.

However, this analysis is not limited to specific groups of genes like GHs. When all predicted genes in a metagenome are included, RNA-Seq can give intriguing insights into metabolically highly active groups of microorganisms and identify the genes with the highest levels of transcript in a sample. In the biogas fermenter sample, the 100 genes with the highest number of total mapped cDNA reads were analyzed with regard to their putative function and phylogenetic affiliation (Table S5, Güllert, Fischer et al. 2016). Among the highest expressed genes, several archaeal genes encoding enzymes involved in methanogenesis were present. These genes were mainly affiliated with hydrogenotrophic species of the genus *Methanoculleus.* This finding supports the assumption of a predominant methanogenesis in agricultural biogas plants via the hydrogenotrophic pathway (Bergmann, Nettmann et al. 2010).

Notably, 23 of the 100 genes originated from bacteria of the family *Peptococcaceae.* This finding demonstrates a substantial metabolic activity of these physiologically diverse and partly acetogenic bacteria (Liu and Conrad 2011) in the sampled fermenter and might further hint to a possible high relevance of this bacterial family in agricultural biogas fermenters in general.

Moreover, a large fraction of the 100 highly transcribed genes encoded hypothetical proteins of unknown function and for six genes, no homologs were found at all in the database. In future studies, it would be highly interesting to examine some of the genes with no known homology in more detail. It is intriguing that some of these highly transcribed genes appear to be completely unknown until today.

In conclusion, an integrative approach combining metagenome, metatranscriptome, and possibly metaproteome analyses is needed to rationally interpret the constantly increasing metagenomic datasets and to identify the metabolically most relevant groups. It further

holds significant potential for the bioprospecting for highly active and novel enzymes (Ferrer, Martinez-Martinez et al. 2016).

7.5 Identification of cellulosomal scaffolding proteins and PULs in the biogas, cow rumen, and elephant feces metagenomes

After elucidating differences in the abundance and phylogenetic affiliation of genes encoding GHs between the different metagenomic datasets, it was also highly interesting to examine the cellulolytic systems in which these enzymes might be embedded in. In the biogas fermenter, it was likely that many of the *Clostridia* derived GHs were part of cellulosomal systems. In contrast, the presence and diversity of cellulosomes in the gut and rumen derived metagenomes would probably be lower. By using the sequences of known cellulosomal scaffolding proteins, an iterative protein search was conducted to identify potential cellulosomal gene clusters in the different metagenomic datasets. This analysis identified one *Clostridium thermocellum* derived scaffoldin and two *Clostridium cellulolyticum* derived scaffoldin proteins in the biogas fermenter dataset (Table S6, Güllert, Fischer et al. 2016). Both of these species are commonly found in biogas fermenters and are presumably among the main lignocellulose degrading organisms in the biogas fermenter (Wirth, Kovacs et al. 2012, Stolze, Zakrzewski et al. 2015, Zverlov, Köck et al. 2015).

In the cow rumen metagenome, two cellulosomal scaffolding proteins, both affiliated with *Ruminococcus flavefaciens,* were identified. In contrast to the biogas fermenter, no cellulosomal scaffoldin protein affiliated with the genus *Clostridium* was identified. *R. flavefaciens* belongs to the genus *Ruminococcus* of the class *Clostridia* and is a cellulolytic bacterium commonly found in the rumen and intestine of animals and humans. The cellulosome of *R. flavefaciens* is highly complex (Venditto, Luis et al. 2016) and the only known example of a cellulosomal system found in a typical gut bacterium (Flint, Bayer et al. 2008).

Interestingly, no potential cellulosomal scaffoldin proteins were found in the elephant feces derived dataset. This is especially noteworthy given the high cellulolytic activity measured in the DNS assay and the high amount of GHs identified in this metagenome. While it is possible that the very comprehensive metagenomic dataset generated from the elephant feces microbiome might still be incomplete and therefor the scaffoldin proteins were

missing, the more likely assumption is that there are no (known) cellulosomal systems present in the gut of the sampled elephant.

With respect to the almost complete absence of cellulosomes in the elephant dataset, the question arises how lignocellulosic biomass is efficiently degraded in the elephant gut despite the lack of cellulosomes and a reduced abundance of the *Firmicutes* compared to the biogas fermenter. For several years, it has been speculated that bacteria of the phylum *Bacteroidetes* might play an important part in the degradation of lignocellulose in the gut and rumen of animals (Mackenzie, Pope et al. 2012, Despres, Forano et al. 2016) The findings presented in this study support this view and demonstrate that the observed higher abundance of cellulolytic GHs in the two studied natural systems is, to a large extend, based on enzymes derived from the *Bacteroidetes.* Due to the high prevalence of PULs in *Bacteroidetes* genomes, until today multiple PULs in 71 genomes have been predicted, it is highly likely that a significant fraction of the *Bacteroidetes* derived GHs are part of PULs (Terrapon, Lombard et al. 2015). A remaining central question is how many of these PULs are involved in the degradation of cellulose.

In a recent study, a putatively cellulolytic PUL was identified in the same published metagenome affiliated with the cow rumen that was analyzed in this study (Naas, Mackenzie et al. 2014). In accordance with this finding, multiple contigs encoding putatively cellulolytic PUL gene clusters were identified in the metagenome obtained from the adult elephant feces associated microbiome (Figure 5, Ilmberger, Güllert et al. 2014). Furthermore, in the biogas fermenter derived metagenome, several potentially cellulolytic PULs were present despite the reduced abundance of bacteroidetal organisms in this sample (Figure 5b, Güllert, Fischer et al. 2016).

In addition to the suspected direct degradation of cellulose by members of the *Bacteroidetes*, bacteria of this phylum might also indirectly contribute to an efficient hydrolysis of lignocellulose by degrading polysaccharides associated with the hemicellulose fraction of the plant cell wall. Recently, PULs involved in the degradation of a several xylans were described (Martens, Koropatkin et al. 2009, Dodd, Mackie et al. 2011, Larsbrink, Rogers et al. 2014, Despres, Forano et al. 2016, Rogowski, Briggs et al. 2016, Wang, Dong et al. 2016).

Within this framework, the recent advances in sequencing throughput together with the development of improved bioinformatic tools will probably lead to the identification of

additional PULs involved in the degradation of plant cell wall polysaccharides and will facilitate a better understanding of these gene clusters. This is especially relevant as many members of the *Bacteroidetes* are difficult to cultivate in the lab and therefore genome- and metagenome-based analyses are important tools to study this bacterial phylum.

In conclusion, the omics-based analyses performed in this study revealed distinct differences in the hydrolysis of carbohydrates between the natural systems and the technical system. Based on the findings in this study, it is likely that the importance and potential of bacteria of the phylum *Bacteroidetes* for the degradation of lignocellulosic biomass in biogas fermenters were underestimated in the past. It is further tempting to speculate that by increasing the abundance of the *Bacteroidetes* in biogas fermenters, an increase in the hydrolytic performance of agricultural biogas plants might be achieved. A higher degradation rate of plant derived biomass in a biogas fermenter will in turn result in a higher overall methane yield. However, this hypophysis has to be verified in additional experiments and suitable approaches to increase the abundance of the *Bacteroidetes* in biogas fermenters have to be elucidated. The simple addition of bacteroidetal isolates at high levels will presumably not result in a lasting increase in the abundance of the *Bacteroidetes* in biogas fermenters. It is likely that the added organisms will quickly be outcompeted again by members of the *Firmicutes.* Therefore, further research is required to identify the crucial factors which are responsible for the better adaption of the *Bacteroidetes* to natural habitats like the gastrointestinal tracts of animals. In a next step, it might be possible to adapt these factors for biotechnical systems such as biogas fermenters and thereby increase the efficiency of these systems for the fermentation of lignocellulosic substrates.

8 References

Ahlgren S. and L. Di Lucia (2016). "Erratum to: Indirect land use changes of biofuel production - a review of modelling efforts and policy developments in the European Union." Biotechnol Biofuels **9**: 40.

Bergmann I., E. Nettmann, K. Mundt and M. Klocke (2010). "Determination of methanogenic Archaea abundance in a mesophilic biogas plant based on 16S rRNA gene sequence analysis." Can J Microbiol **56**.

Berlemont R. and A. C. Martiny (2013). "Phylogenetic Distribution of Potential Cellulases in Bacteria." Applied and Environmental Microbiology **79**(5): 1545-1554.

Bian G., L. Ma, Y. Su and W. Zhu (2013). "The microbial community in the feces of the white rhinoceros (Ceratotherium simum) as determined by barcoded pyrosequencing analysis." PLoS One **8**.

Blum D. L., I. A. Kataeva, X. L. Li and L. G. Ljungdahl (2000). "Feruloyl esterase activity of the Clostridium thermocellum cellulosome can be attributed to previously unknown domains of XynY and XynZ." Journal of Bacteriology **182**(5): 1346-1351.

Braga C. M., S. Delabona Pda, D. J. Lima, D. A. Paixao, J. G. Pradella and C. S. Farinas (2014). "Addition of feruloyl esterase and xylanase produced on-site improves sugarcane bagasse hydrolysis." Bioresour Technol **170**: 316-324.

Brown M. E. and M. C. Y. Chang (2014). "Exploring bacterial lignin degradation." Current Opinion in Chemical Biology **19**: 1-7.

Christopherson M. R., G. Suen, S. Bramhacharya, K. A. Jewell, F. O. Aylward, D. Mead and P. J. Brumm (2013). "The Genome Sequences of Cellulomonas fimi and "Cellvibrio gilvus" Reveal the Cellulolytic Strategies of Two Facultative Anaerobes, Transfer of "Cellvibrio gilvus" to the Genus Cellulomonas, and Proposal of Cellulomonas gilvus sp nov." Plos One **8**(1).

Cordonnier C., G. Le Bihan, J. G. Emond-Rheault, A. Garrivier, J. Harel and G. Jubelin (2016). "Vitamin B12 Uptake by the Gut Commensal Bacteria Bacteroides thetaiotaomicron Limits the Production of Shiga Toxin by Enterohemorrhagic Escherichia coli." Toxins (Basel) **8**(1).

Degnan P. H., N. A. Barry, K. C. Mok, M. E. Taga and A. L. Goodman (2014). "Human Gut Microbes Use Multiple Transporters to Distinguish Vitamin B-12 Analogs and Compete in the Gut." Cell Host & Microbe **15**(1): 47-57.

Demain A. L., M. Newcomb and J. H. D. Wu (2005). "Cellulase, clostridia, and ethanol." Microbiology and Molecular Biology Reviews **69**(1): 124-154.

Despres J., E. Forano, P. Lepercq, S. Comtet-Marre, G. Jubelin, C. Chambon, C. J. Yeoman, M. E. Berg Miller, C. J. Fields, E. Martens, N. Terrapon, B. Henrissat, B. A. White and P.

Mosoni (2016). "Xylan degradation by the human gut Bacteroides xylanisolvens XB1A(T) involves two distinct gene clusters that are linked at the transcriptional level." BMC Genomics **17**: 326.

Despres J., E. Forano, P. Lepercq, S. Comtet-Marre, G. Jubelin, C. J. Yeoman, M. E. B. Miller, C. J. Fields, N. Terrapon, C. Le Bourvellec, C. M. G. C. Renard, B. Henrissat, B. A. White and P. Mosoni (2016). "Unraveling the pectinolytic function of Bacteroides xylanisolvens using a RNA-seq approach and mutagenesis (vol 27, 17, 2016)." Bmc Genomics **17**.

Dodd D., R. I. Mackie and I. K. O. Cann (2011). "Xylan degradation, a metabolic property shared by rumen and human colonic Bacteroidetes." Molecular Microbiology **79**(2): 292-304.

Dodd D., Y. H. Moon, K. Swaminathan, R. I. Mackie and I. K. O. Cann (2010). "Transcriptomic Analyses of Xylan Degradation by Prevotella bryantii and Insights into Energy Acquisition by Xylanolytic Bacteroidetes." Journal of Biological Chemistry **285**(39): 30261-30273.

Doi R. H. and A. Kosugi (2004). "Cellulosomes: Plant-cell-wall-degrading enzyme complexes." Nature Reviews Microbiology **2**(7): 541-551.

Eckert K. and E. Schneider (2003). "A thermoacidophilic endoglucanase (CelB) from Alicyclobacillus acidocaldarius displays high sequence similarity to arabinofuranosidases belonging to family 51 of glycoside hydrolases." European Journal of Biochemistry **270**(17): 3593-3602.

Ferrer M., M. Martinez-Martinez, R. Bargiela, W. R. Streit, O. V. Golyshina and P. N. Golyshin (2016). "Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends." Microbial Biotechnology **9**(1): 22-34.

Flint H. J., E. A. Bayer, M. T. Rincon, R. Lamed and B. A. White (2008). "Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis." Nat Rev Microbiol **6**.

Fontes C. M. G. A. and H. J. Gilbert (2010). "Cellulosomes: Highly Efficient Nanomachines Designed to Designed to Deconstruct Plant Cell Wall Complex Carbohydrates." Annual Review of Biochemistry, Vol 79 **79**: 655-681.

Gefen G., M. Anbar, E. Morag, R. Lamed and E. A. Bayer (2012). "Enhanced cellulose degradation by targeted integration of a cohesin-fused beta-glucosidase into the Clostridium thermocellum cellulosome." Proceedings of the National Academy of Sciences of the United States of America **109**(26): 10298-10303.

Grosdidier A., V. Zoete and O. Michielin (2011). "SwissDock, a protein-small molecule docking web service based on EADock DSS." Nucleic Acids Research **39**: W270-W277.

Güllert S., M. A. Fischer, D. Turaev, B. Noebauer, N. Ilmberger, B. Wemheuer, M. Alawi, T. Rattei, R. Daniel, R. A. Schmitz, A. Grundhoff and W. R. Streit (2016). "Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial

biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies." Biotechnology for Biofuels **9**(1): 1-20.

He S., N. Ivanova, E. Kirton, M. Allgaier, C. Bergin, R. H. Scheffrahn, N. C. Kyrpides, F. Warnecke, S. G. Tringe and P. Hugenholtz (2013). "Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites." PLoS One **8**.

Hehemann J. H., A. G. Kelly, N. A. Pudlo, E. C. Martens and A. B. Boraston (2012). "Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes." Proceedings of the National Academy of Sciences of the United States of America **109**(48): 19786-19791.

Henderson G., F. Cox, S. Kittelmann, V. H. Miri, M. Zethof, S. J. Noel, G. C. Waghorn and P. H. Janssen (2013). "Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities." PLoS One **8**.

Herrmann C., C. Idler and M. Heiermann (2015). "Improving aerobic stability and biogas production of maize silage using silage additives." Bioresour Technol **197**: 393-403.

Hess M., A. Sczyrba, R. Egan, T. W. Kim, H. Chokhawala, G. Schroth, S. Luo, D. S. Clark, F. Chen and T. Zhang (2011). "Metagenomic discovery of biomass-degrading genes and genomes from cow rumen." Science **331**.

Horn S. J., G. Vaaje-Kolstad, B. Westereng and V. G. Eijsink (2012). "Novel enzymes for the degradation of cellulose." Biotechnol Biofuels **5**(1): 45.

Ilmberger N., S. Güllert, J. Dannenberg, U. Rabausch, J. Torres, B. Wemheuer, M. Alawi, A. Poehlein, J. Chow and D. Turaev (2014). "A comparative metagenome survey of the fecal microbiota of a breast- and a plant-fed Asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes." PLoS One **9**.

Jaenicke S., C. Ander, T. Bekel, R. Bisdorf, M. Droge, K. H. Gartemann, S. Junemann, O. Kaiser, L. Krause and F. Tille (2011). "Comparative and joint analysis of two metagenomic datasets from a biogas fermenter obtained by 454-pyrosequencing." PLoS One **6**.

Kabisch A., A. Otto, S. Konig, D. Becher, D. Albrecht, M. Schuler, H. Teeling, R. I. Amann and T. Schweder (2014). "Functional characterization of polysaccharide utilization loci in the marine Bacteroidetes 'Gramella forsetii' KT0803." Isme Journal **8**(7): 1492-1502.

Kim D. E., D. Chivian and D. Baker (2004). "Protein structure prediction and analysis using the Robetta server." Nucleic Acids Research **32**: W526-W531.

Larsbrink J., T. E. Rogers, G. R. Hemsworth, L. S. McKee, A. S. Tauzin, O. Spadiut, S. Klinter, N. A. Pudlo, K. Urs, N. M. Koropatkin, A. L. Creagh, C. A. Haynes, A. G. Kelly, S. N. Cederholm, G. J. Davies, E. C. Martens and H. Brumer (2014). "A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes." Nature **506**(7489): 498-502.

Li A., Y. Chu, X. Wang, L. Ren, J. Yu, X. Liu, J. Yan, L. Zhang, S. Wu and S. Li (2013). "A pyrosequencing-based metagenomic study of methane-producing microbial community in solid-state biogas reactor." Biotechnol Biofuels **6**.

Limayem A. and S. C. Ricke (2012). "Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects." Progress in Energy and Combustion Science **38**(4): 449-467.

Liu F. and R. Conrad (2011). "Chemolithotrophic acetogenic H2/CO2 utilization in Italian rice field soil." ISME J **5**.

Lombard V., H. G. Ramulu, E. Drula, P. M. Coutinho and B. Henrissat (2014). "The carbohydrate-active enzymes database (CAZy) in 2013." Nucleic acids Res **42**.

Lynd L. R., P. J. Weimer, W. H. van Zyl and I. S. Pretorius (2002). "Microbial cellulose utilization: Fundamentals and biotechnology." Microbiology and Molecular Biology Reviews **66**(3): 506-577.

Mackenzie A. K., P. B. Pope, H. L. Pedersen, R. Gupta, M. Morrison, W. G. Willats and V. G. Eijsink (2012). "Two SusD-like proteins encoded within a polysaccharide utilization locus of an uncultured ruminant Bacteroidetes phylotype bind strongly to cellulose." Appl Environ Microbiol **78**.

Magnusdottir S., D. Ravcheev, V. de Crecy-Lagard and I. Thiele (2015). "Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes." Front Genet **6**: 148.

Martens E. C., N. M. Koropatkin, T. J. Smith and J. I. Gordon (2009). "Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm." Journal of Biological Chemistry **284**(37): 24673-24677.

Martens E. C., E. C. Lowe, H. Chiang, N. A. Pudlo, M. Wu, N. P. McNulty, D. W. Abbott, B. Henrissat, H. J. Gilbert and D. N. Bolam (2011). "Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts." PLoS Biol **9**.

Martens E. C., E. C. Lowe, H. Chiang, N. A. Pudlo, M. Wu, N. P. McNulty, D. W. Abbott, B. Henrissat, H. J. Gilbert, D. N. Bolam and J. I. Gordon (2011). "Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts." Plos Biology **9**(12).

Martens J. H., H. Barg, M. J. Warren and D. Jahn (2002). "Microbial production of vitamin B12." Appl Microbiol Biotechnol **58**(3): 275-285.

Mechaly A., H. P. Fierobe, A. Belaich, J. P. Belaich, R. Lamed, Y. Shoham and E. A. Bayer (2001). "Cohesin-dockerin interaction in cellulosome assembly - A single hydroxyl group of a dockerin domain distinguishes between nonrecognition and high affinity recognition." Journal of Biological Chemistry **276**(13): 9883-9888.

Naas A. E., A. K. Mackenzie, J. Mravec, J. Schuckel, W. G. Willats, V. G. Eijsink and P. B. Pope (2014). "Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation?" MBio **5**(4): e01401-01414.

Nettmann E., I. Bergmann, S. Pramschufer, K. Mundt, V. Plogsties, C. Herrmann and M. Klocke (2010). "Polyphasic analyses of methanogenic archaeal communities in agricultural biogas plants." Appl Environ Microbiol **76**.

Pope P. B., A. K. Mackenzie, I. Gregor, W. Smith, M. A. Sundset, A. C. McHardy, M. Morrison and V. G. Eijsink (2012). "Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci." PLoS One **7**.

Rakotoarivonina H., B. Hermant, B. Chabbert, J. P. Touzel and C. Remond (2011). "A thermostable feruloyl-esterase from the hemicellulolytic bacterium Thermobacillus xylanilyticus releases phenolic acids from non-pretreated plant cell walls." Applied Microbiology and Biotechnology **90**(2): 541-552.

Roggenbuck M., C. Sauer, M. Poulsen, M. F. Bertelsen and S. J. Sorensen (2014). "The giraffe (Giraffa camelopardalis) rumen microbiome." FEMS Microbiol Ecol **90**.

Rogowski A., J. A. Briggs, J. C. Mortimer, T. Tryfona, N. Terrapon, E. C. Lowe, A. Basle, C. Morland, A. M. Day, H. J. Zheng, T. E. Rogers, P. Thompson, A. R. Hawkins, M. P. Yadav, B. Henrissat, E. C. Martens, P. Dupree, H. J. Gilbert and D. N. Bolam (2016). "Glycan complexity dictates microbial resource allocation in the large intestine (vol 6, 7481, 2015)." Nature Communications **7**.

Rubin E. M. (2008). "Genomics of cellulosic biofuels." Nature **454**(7206): 841-845.

Scheller H. V. and P. Ulvskov (2010). "Hemicelluloses." Annual Review of Plant Biology, Vol 61 **61**: 263-289.

Shallom D. and Y. Shoham (2003). "Microbial hemicellulases." Current Opinion in Microbiology **6**(3): 219-228.

Simon C. and R. Daniel (2011). "Metagenomic Analyses: Past and Future Trends." Applied and Environmental Microbiology **77**(4): 1153-1161.

Solli L., O. E. Havelsrud, S. J. Horn and A. G. Rike (2014). "A metagenomic study of the microbial communities in four parallel biogas reactors." Biotechnol Biofuels **7**.

Sonnenburg E. D., H. J. Zheng, P. Joglekar, S. K. Higginbottom, S. J. Firbank, D. N. Bolam and J. L. Sonnenburg (2010). "Specificity of Polysaccharide Use in Intestinal Bacteroides Species Determines Diet-Induced Microbiota Alterations." Cell **141**(7): 1241-U1256.

Stolze Y., M. Zakrzewski, I. Maus, F. Eikmeyer, S. Jaenicke, N. Rottmann, C. Siebner, A. Puhler and A. Schluter (2015). "Comparative metagenomics of biogas-producing microbial communities from production-scale biogas plants operating under wet or dry fermentation conditions." Biotechnol Biofuels **8**.

Streit W. R. and R. A. Schmitz (2004). "Metagenomics - the key to the uncultured microbes." Current Opinion in Microbiology **7**(5): 492-498.

Tenenbaum D. J. (2008). "Food vs. fuel: diversion of crops could cause more hunger." Environ Health Perspect **116**(6): A254-257.

Terrapon N., V. Lombard, H. J. Gilbert and B. Henrissat (2015). "Automatic prediction of polysaccharide utilization loci in Bacteroidetes species." Bioinformatics **31**.

Thomas F., J. H. Hehemann, E. Rebuffet, M. Czjzek and G. Michel (2011). "Environmental and gut Bacteroidetes: the food connection." Frontiers in Microbiology **2**.

Thompson P. B. (2012). "The agricultural ethics of biofuels: climate ethics and mitigation arguments." Poiesis Prax **8**(4): 169-189.

Venditto I., A. S. Luis, M. Rydahl, J. Schuckel, V. O. Fernandes, S. Vidal-Melgosa, P. Bule, A. Goyal, V. M. R. Pires, C. G. Dourado, L. M. A. Ferreira, P. M. Coutinho, B. Henrissat, J. P. Knox, A. Basle, S. Najmudin, H. J. Gilbert, W. G. T. Willats and C. M. G. A. Fontes (2016). "Complexity of the Ruminococcus flavefaciens cellulosome reflects an expansion in glycan recognition." Proceedings of the National Academy of Sciences of the United States of America **113**(26): 7136-7141.

Wang C., D. Dong, H. S. Wang, K. Muller, Y. Qin, H. L. Wang and W. X. Wu (2016). "Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition." Biotechnology for Biofuels **9**.

Wang L. L., A. Hatem, U. V. Catalyurek, M. Morrison and Z. T. Yu (2013). "Metagenomic Insights into the Carbohydrate-Active Enzymes Carried by the Microorganisms Adhering to Solid Digesta in the Rumen of Cows." Plos One **8**(11).

Weiland P. (2010). "Biogas production: current state and perspectives." Appl Microbiol Biotechnol **85**.

Wirth R., E. Kovacs, G. Maroti, Z. Bagi, G. Rakhely and K. L. Kovacs (2012). "Characterization of a biogas-producing microbial community by short-read next generation DNA sequencing." Biotechnol Biofuels **5**.

Wong D. W. (2006). "Feruloyl esterase: a key enzyme in biomass degradation." Appl Biochem Biotechnol **133**.

Wong D. W. S., V. J. Chan, S. B. Batt, G. Sarath and H. Liao (2011). "Engineering Saccharomyces cerevisiae to produce feruloyl esterase for the release of ferulic acid from switchgrass." Journal of Industrial Microbiology & Biotechnology **38**(12): 1961-1967.

Yaniv O., G. Fichman, I. Borovok, Y. Shoham, E. A. Bayer, R. Lamed, L. J. Shimon and F. Frolow (2014). "Fine-structural variance of family 3 carbohydrate-binding modules as extracellular biomass-sensing components of Clostridium thermocellum anti-sigmaI factors." Acta Crystallogr D Biol Crystallogr **70**(Pt 2): 522-534.

Zakrzewski M., A. Goesmann, S. Jaenicke, S. Junemann, F. Eikmeyer, R. Szczepanowski, W. A. Al-Soud, S. Sorensen, A. Puhler and A. Schluter (2012). "Profiling of the metabolically active community from a production-scale biogas plant by means of high-throughput metatranscriptome sequencing." J Biotechnol **158**.

Zeng B., S. Han, P. Wang, B. Wen, W. Jian, W. Guo, Z. Yu, D. Du, X. Fu and F. Kong (2015). "The bacterial communities associated with fecal types and body weight of rex rabbits." Sci Rep **5**.

Zheng Y., J. Zhao, F. Q. Xu and Y. B. Li (2014). "Pretreatment of lignocellulosic biomass for enhanced biogas production." Progress in Energy and Combustion Science **42**: 35-53.

Zverlov V., D. Köck and W. Schwarz (2015). The role of cellulose-hydrolyzing bacteria in the production of biogas from plant biomass. Microorganisms in biorefineries. B. Kamm. Berlin, Springer.

9 Acknowledgments

Most importantly, I would like to thank Prof. Dr. Wolfgang Streit for giving me the opportunity to conduct my dissertation in his group and for his constant support. He was always there to give advice and guidance when needed but also gave me the freedom to shape this project and to pursuit my own ideas. For this, I am really grateful.

I further would like to thank Prof. Dr. Wolfgang Liebl and all other members of my examination committee for taking the time to evaluate my work.

Beyond that, I would like to thank all my co-workers of the microbiology and biotechnology department for their support, the nice working atmosphere, and the good times we had inside and outside the lab.

My special gratitude goes to Annina for all her interest in my work, her support and affection. I am very much looking forward to the next big chapter in our life together.

Finally, I would like to thank my parents for supporting me all these years, believing in me and making it possible for me to get to this point.