Aalborg Universitet

AALBORG UNIVERSITY

Novel Microorganisms and Their Function on Anaerobic Digesters

Kirkegaard, Rasmus Hansen

DOI (link to publication from Publisher):
10.5278/vbn.phd.engsci.00180

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Kirkegaard, R. H. (2016). *Novel Microorganisms and Their Function on Anaerobic Digesters*. Aalborg Universitetsforlag. https://doi.org/10.5278/vbn.phd.engsci.00180

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: July 05, 2025

NOVEL MICROORGANISMS AND THEIR FUNCTION IN ANAEROBIC DIGESTERS

BY
RASMUS H. KIRKEGAARD

DISSERTATION SUBMITTED 2016

AALBORG UNIVERSITY
DENMARK

NOVEL MICROORGANISMS AND THEIR FUNCTION IN ANAEROBIC DIGESTERS

by

Rasmus H. Kirkegaard

AALBORG UNIVERSITY
DENMARK

Dissertation submitted December 21 2016

Dissertation submitted:	December 21, 2016
PhD supervisor:	Prof. Per Halkjær Nielsen Aalborg University
Assistant PhD supervisor:	Postdoc Mads Albertsen Aalborg University
PhD committee:	Associate Professor Lars Haastrup Pedersen (chairman) Aalborg University
	Professor Anna Schnürer Swedish University of Agricultural Sciences
	Associate Professor David Berry University of Vienna
PhD Series:	Faculty of Engineering and Science, Aalborg University

ISSN (online): 2246-1248
ISBN (online): 978-87-7112-856-7

Published by:
Aalborg University Press
Skjernvej 4A, 2nd floor
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

© Copyright: Rasmus H. Kirkegaard

Printed in Denmark by Rosendahls, 2017

ENGLISH SUMMARY

Biogas production from wastewater streams represents a resource for replacing fossil fuels while minimising the cost and energy needs for waste treatment. This is a central part of the bio-refinery view of wastewater treatment plants and have the potential to turn them into small power plants as an alternative to costly treatment facilities. It is widely believed that we need to advance our understanding of the microbes to fully realise this potential and improve the performance of the underlying biological processes. To determine whether this is a feasible strategy, we need to grasp how many organisms we need to know. What organisms are present? and are the same bacteria present in all reactors? Alternatively, do they differ between different reactors and why? Beyond identification, we also need to establish the functional roles of the different microbes within the anaerobic digesters.

The aim of this project was to establish an overview of the microorganisms in Danish digesters treating primary and surplus sludge at wastewater treatment plant with a focus on novel microbes. In particular, the objectives were to identify novel and uncharacterised microorganisms important for the processes and then obtain their genomic information to serve as a basis for future investigations.

We analysed the microbial community composition in 32 full-scale anaerobic digesters using 16S rRNA gene amplicon sequencing. The reactors were sampled during a six-year period and we used influent samples to distinguish between microbes that were actively growing and the ones that were influent related. This analysis provided evidence for a short list of “most wanted” organisms that made up the majority of the biomass. To improve the discovery beyond PCR biased amplicon sequencing, we developed and tested methods for obtaining full-length SSU sequences directly from rRNA. This provided evidence for many novel bacterial sequences within anaerobic digesters even from a single sample and holds the potential to expand the reference databases dramatically. Based on the comprehensive survey we identified targets for genome recovery and developed the “mmgenome” toolbox to support reproducible genome binning. Using this tool, we successfully retrieved the first genomic information for members of the Hyd24-12 phylum (now Fermentibacteria), as well as a novel genus within the family Anaerolineaceae.

DANSK RESUME

Biogasproduktion fra spildevand repræsenterer en ressource til at erstatte fossilt brændsel samtidig med at det minimerer udgifterne og energibehovet til spildevandsbehandling. Biogassen er en central del af den fremvindende "bio-refinery" tankegang for renseanlæg og har potentiale til at ændre dem fra omkostningstunge renseanlæg til små kraftværker. Der hersker en udbredt enighed om at der er et behov for at øge vores forståelse for mikroorganismerne for at forbedre performance for de underliggende biologiske reaktioner. For at vurdere hvorvidt det er en realistisk strategi er vi nødt til at bestemme hvor mange forskellige organismer vi skal lære at kende. Hvilke mikroorganismer er tilstede? og er det de samme i alle anlæg? Alternativt, er der forskellige arter i forskellige rådnetanks reaktorer?

Målet med dette projekt var at etablere et overblik over mikroorganismerne i Danske rådnetanke ved renseanlæg hvor de behandler primært og sekundært slam. Dette med et fokus på opdagelse af hidtil ukendte bakterier. Vigtige delmål var at identificere ukendte og ubeskrevne mikroorganismer, der er vigtige for processerne og fremskaffe deres genetiske information, som kan danne grundlag for fremtidige studier.

Vi analyserede sammensætningen af de mikrobielle samfund i 32 fuldskala rådnetanke ved hjælp af amplikon sekventering af 16S rRNA genet. Der var taget prøver fra reaktorerne gennem en seks årig periode og vi analyserede indløbsprøver for at belyse hvilke mikroorganismer, der groede aktivt og hvilke, der bare kommer med ind i indløbet. På baggrund af denne analyse fandt vi at en kort liste udgjorde langt størstedelen af biomassen. For at forbedre mulighederne for at opdage nye mikroorganismer uden PCR bias udviklede og testede vi en metode til at bestemme fuldlængde ribosomale RNA sekvenser. Denne analyse indikerede at der selv i en enkel prøve fra rådnetanke er ukendt diversitet nok til at udvide de eksisterende databaser drastisk. Baseret på den omfattende karakterisering af mikrobiologien i danske rådnetanke identificerede vi vigtige organismer for hel genom sekventering og udviklede "mmgenome" værktøjet til at supportere reproducerbar genom binning fra metagenomer. Ved hjælp af dette værktøj opnåede vi at hive den genomiske information ud af metagenomer for medlemmer af Hyd24-12 rækken, som vi har navngivet Fermentibacteria såvel som for et hidtil ukendt genus af Aanaerolineaceae familien.

ACKNOWLEDGEMENTS

I would like express my deep gratitude to my supervisor Per Halkjær Nielsen for the great opportunity to join the scientific community under his great supervision and as a part of his centre for microbial communities. I have enjoyed his valuable inputs and enthusiastic support during the project as well as our scientific discussions at the university and at the pub after conference dinners.

I would also like to thank my amazing colleagues in the environmental biotechnology group for their support in many aspects of the PhD life ensuring a productive and entertaining time with great scientific discoveries, loads of conferences, cake for any occasion, competitive go-kart racing, relaxed paddle-boarding and much more. I would like to address special thanks to my great mentors, office-mates and friends Mads Albertsen, Søren Karst and Simon J. McIlroy for their invaluable collaboration, knowledge sharing, and discussions about almost anything during my PhD journey, which has taken us watching bacterial mats, bears, birds, wolves, and geysers, bike riding along beautiful Swedish rivers and drinking classy cocktails.

Last but not at least I would like to thank my family and friends for their continuous support and enthusiasm about my work and other crazy projects during the last few years such as my first full distance ironman and buying a boat. They have also been instrumental in keeping me aware that not everything revolves around microbes and DNA sequencing all the time.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

Burning of fossil fuels have powered the industrial revolution and accelerated the technological development. However, along with the benefits from this rapid development comes the problems with pollution, health risks, climate change and resource limitations in a closed system when burning the carbon energy deposits accumulated over millions of years (Schellnhuber *et al.*, 2016). In addition, the fossil energy deposits are not evenly distributed and thus not matching the demand in the local regions. This poses a threat to the global economy as political instabilities in the countries with large production of conventional energy can trigger shocks in the global markets or even military action (Bentley, 2002).

Awareness of the need for sustainability and a circular economy, has spurred the interest for alternative sources of energy and better use of existing resources through recovery and recycling (Rogelj *et al.*, 2016). Biogas production from biological waste products such as manure, food waste, and wastewater have the potential to fit both of these needs as it turns carbon waste streams into valuable energy stored as methane that can be utilised as a replacement for fossil fuels in the existing infrastructure. Bio energy is already widely used as an important piece in the sustainability puzzle but does still have potential for further development (Weiland, 2010).

Biogas production from wastewater has benefits that go beyond mere energy production. The carbon and nutrient removal in conventional aerobic wastewater treatment plants is extremely important to avoid eutrophication of receiving waters but the energy demand due to the need for aeration is extremely costly. Wastewater treated anaerobically does thus not just produce energy; it also saves energy for aeration in the conventional treatment process and anaerobic digesters are thus considered a central part of modern wastewater treatment (**Figure 1**). It is believed among researchers and the industry that wastewater treatment can become at least energy neutral or even act as a net energy producer with biogas production from anaerobic digestion effectively making the facility a power plant (Jenicek *et al.*, 2013; McCarty *et al.*, 2011; Hughes, 2015). In addition, the reduced sludge volume makes it cheaper to transport and get rid of the final product now enriched in nutrients such as nitrogen and phosphorus, thereby increasing its value as fertiliser (McCarty *et al.*, 2011). Pathogens in the wastewater pose a threat to human health as microbes from e.g. excrements can cause serious infections that need antibiotic treatment. Anaerobic digestion has shown potential for effective removal of certain human pathogens such as *E. coli* and *Salmonella* spp. especially in thermophilic reactors, whereas they might survive in mesophilic reactors where the temperatures are similar to that of the human body. This is especially important if the remaining sludge is to be used as a fertiliser on farmland (Chen *et al.*, 2012; Smith *et al.*, 2005; Watanabe *et al.*, 1997).

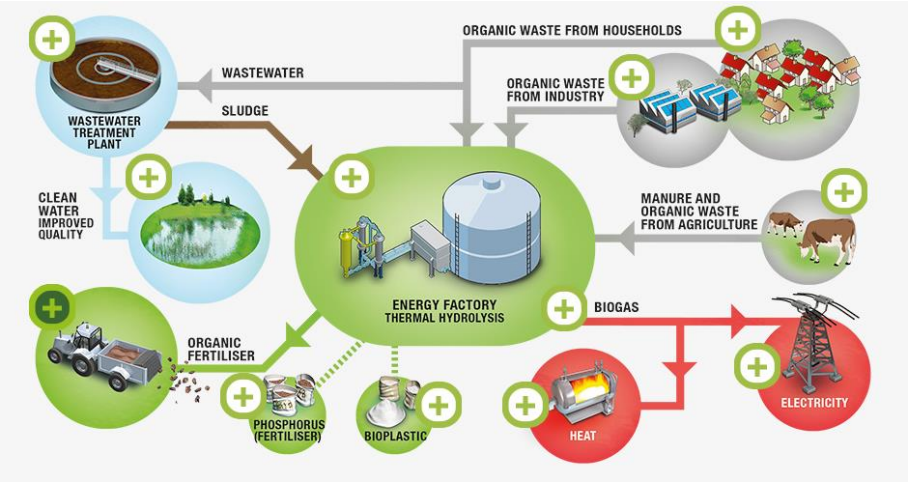

Figure 1 Anaerobic digestion is considered a central part of a modern wastewater treatment plant as it can deliver biogas, bioplastics, and fertilisers while reducing the sludge volume (Source www.billundbiorefinery.dk/en/).

1.1. ANAEROBIC DIGESTION

Biogas production relies on a multifaceted set of microbes to orchestrate the biological conversion of complex biomolecules into methane. This conversion happens as a four-step process beginning with the hydrolysis of proteins, carbohydrates and lipids into sugars, amino acids and fatty acids (**Figure 2**). These molecules are fermented into volatile fatty acids in the acidogenesis and the produced acids are then converted into acetate in the acetogenesis. The acetate as well as the produced hydrogen and carbon dioxide is finally converted into methane in the methanogenesis via either the acetoclastic or hydrogenoclastic pathway (Saady, 2013). Depending on the partial pressure of hydrogen, microbes might also be able to shift between the two paths to methane through syntrophic acetate oxidation or homoacetogenesis (Sun *et al.*, 2014; Saady, 2013).

As no single microbe has been found to carry out the entire process, biogas production relies on well-balanced interactions between groups of microorganisms spanning both the bacterial and archaeal domain. Bacteria are involved in the first three steps of the conversion whereas archaea are solely responsible for the methane production, as no methane-producing bacteria have been identified yet. Organisms capable of some of these reactions and their tight syntrophic interactions have been demonstrated in lab cultures (Patel and Sprott, 1990; Yamada *et al.*, 2006). However, the responsible organisms working at full-scale complexity were often poorly characterised or completely unknown (Weiland, 2010). A number of studies have therefore sought to describe the microbial communities in full-scale anaerobic digesters using DNA based surveys and by recovery of genomes from some of these organisms (Sundberg *et al.*,

2013; Vrieze *et al.*, 2015b; Treu *et al.*, 2016; Vanwonterghem *et al.*, 2016). It is widely believed that an improved biological understanding of the identified organisms is going to help ensure that the biological process runs well in an engineered system. The direct benefits will be the increase in both process performance and stability (Vanwonterghem *et al.*, 2014; Carballa *et al.*, 2015). Fundamental knowledge of hitherto uncharacterised microorganisms will be obtained as an extra benefit from the process of optimising performance from biological insight by adding some new entries to the microbial encyclopaedia and expanding what is already known about the microbial tree of life and microbial ecology (Rinke *et al.*, 2013). The microbial surveys of anaerobic digesters have highlighted that we know little or nothing about many of the bacteria present in these systems or their interactions with other microbes (Sundberg *et al.*, 2013; Vrieze *et al.*, 2015b).

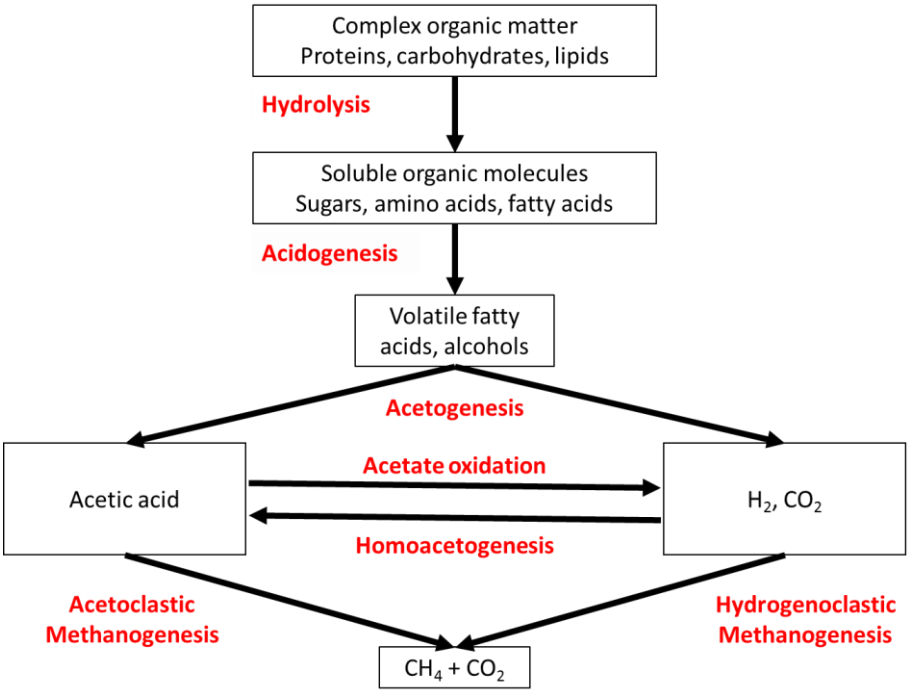

Figure 2 Conversion of complex biomass into methane and carbon dioxide involves four steps hydrolysis, acidogenesis, acetogenesis and methanogenesis. Methane is created through two different pathways depending on the microbes present and bacteria can shift between the pathways by performing either acetate oxidation or homoacetogenesis. Adapted from (Saady, 2013).

1.2. AIM

The aim of this project was to discover novel abundant microorganisms in anaerobic digesters at wastewater treatment plants and obtain their genomes as a basis to describe their functional potential and ultimately understand their role in the ecosystem. More specifically the two main objectives of this work can be stated:

1. To identify novel and abundant microorganisms that are active in anaerobic digesters through a survey spanning multiple plants and years to make a prioritised “most wanted list”.
2. To obtain genomic information for the high priority target microbes in anaerobic digesters through binning of sequencing data from metagenomes and subsequent metabolic reconstruction.

1.3. IDENTIFICATION OF TARGET MICROBES IN ANAEROBIC DIGESTERS

Microorganisms have been studied extensively since their discovery in the 17th century and the techniques have changed dramatically from "simple" morphology based microscopy to whole genome sequencing, gene expression profiling, and DNA based identification of individual cells. The methods are thus providing information with increasing resolution and often proving that microbes are far more capable and complex than we previously imagined (Gest, 2004; Koch *et al.*, 2014; van Kessel *et al.*, 2015; Daims *et al.*, 2015).

1.3.1. RAPID IDENTIFICATION OF MICROBES

The current standards provide identification within a few days of all microorganisms in complex samples. This is achieved through DNA sequencing of marker genes that are classified based on curated reference databases to reveal the composition in many samples with high throughput (Caporaso *et al.*, 2011). This type of microbial analysis have been used in the context of anaerobic digestion to identify the key organisms involved in the processes and to assess whether these were shared among different reactors independently of location providing a basis for generalised biological insight (Rivière *et al.*, 2009).

The analysis of archaea and bacteria involved in the anaerobic digestion relies on sequencing of the universal 16S rRNA gene that contains regions with evolutionary divergence as well as more conserved regions that make up primer targets for producing amplicons (Woese and Fox, 1977; Ashelford *et al.*, 2005). The short read length of the most common DNA sequencing platforms have limited high-throughput analysis of complex samples to include only a subset of the 16S rRNA gene. Thus only covering one or a few of the variable regions. The lack of a perfect region for covering all organisms has led to studies applying different primers and thus introducing different PCR biases that make comparison between studies less tractable (Klindworth *et al.*, 2013).

1.3.2. LINKING SEQUENCES WITH NAMES

The link between the DNA sequence from an amplicon and a biological name is provided by the comparison with a curated taxonomy. However, there is currently no consensus on how this should be developed and maintained and there are thus a number of alternative reference databases (Silva, RDP, Greengenes and NCBI) providing different names for the same DNA sequences (Mcdonald *et al.*, 2012; Quast *et al.*, 2013; Cole *et al.*, 2009; Federhen, 2012). A stable link between sequence and a name is key to biological understanding as it allows researchers to draw upon the pool of existing knowledge about the organisms (McIlroy *et al.*, 2015). To benefit from this link it is vital to get taxonomic classifications with a resolution that ensures that

at least some phenotypic features are conserved within the phylogenetic group. To match this criteria classifications to genus or species level is needed whereas higher level classifications are less likely to show trait conservation within the groups (Martiny *et al.*, 2013).

1.3.3. COMPARISON BETWEEN STUDIES

The use of different DNA extraction protocols, primers and reference databases have been a problem in the context of anaerobic digestion where it makes it difficult to compare findings between studies and generalise knowledge (**Table 1**) (Debelius *et al.*, 2016). In addition, the introduction of active biomass into biogas reactors poses the question of whether detected organisms are active, dead or dormant. DNA can be extracted and amplified by PCR from cells in all of these stages but that does not mean that the organisms attribute to the function of the process. Solutions to inhibit PCR from dead cells using chemical addition and light activation have been suggested. However, they were found not to be suited for complex and dark samples such as anaerobic digester sludge due to limited penetration and complex interactions (Wagner *et al.*, 2008). The problem of inactive organisms has been identified in engineered biological treatment systems including anaerobic digesters using microbial analysis of influent streams as an indicator of whether organisms are actively growing in the system (Saunders *et al.*, 2015; Mei *et al.*, 2016).

The use of different protocols for characterisation of microbes in anaerobic digesters hampers the establishment of general knowledge, as findings in one study are often impossible to validate with another. Surveys of anaerobic sludge digesters applying 16S rRNA gene amplicon sequencing has led to the current consensus that microorganisms are generally shared among different plants with similar configuration (Rivière *et al.*, 2009; Sundberg *et al.*, 2013; Vrieze *et al.*, 2015b). Furthermore, it has provided insight into the main players of anaerobic digestion at full-scale. Revealing the acetoclastic genus *Methanosaeta* as the dominant group of methanogens, followed by lower abundances of hydrogenoclastic methanogens such as *Methanobrevibacter*, *Methanobacterium*, *Methanoculleus* and others. Most archaeal OTUs had close matches in the databases and thus received genus level classification for most of the sequences (Rivière *et al.*, 2009; Sundberg *et al.*, 2013; Vrieze *et al.*, 2015b). However, the bacterial communities have presented more novel sequences and a larger fraction, 73% (Sundberg *et al.*, 2013) and 67% (Vrieze *et al.*, 2015b), of the reads were not assigned a genus level classification compared to only 3% of the archaeal reads (Sundberg *et al.*, 2013). The lack of a genus name for the bacteria makes it impossible to establish the link between the sequence information and previous knowledge obtained from other studies as the conservation of traits is less pronounced at higher taxonomic levels (Martiny *et al.*, 2013).

To mitigate the problems with poor classifications of bacterial OTUs and incoming microbes in anaerobic digesters we have analysed the microbial community

composition of more than 300 samples from thirty-two anaerobic digesters during six years along with the samples from the influent streams of primary and surplus sludge (**Paper 1**). We have used the most abundant OTUs from this comprehensive microbial analysis in the anaerobic digesters to guide an extensive curation of the Silva taxonomy to expand our similar efforts from activated sludge, namely the MiDAS database (**Paper 2**). With this work, we have been able to classify 80% of the bacterial and more than 90% of the archaeal reads to the genus level and identify organisms such as *Microthrix* and *Tetrasphaera* that were clearly influent derived and likely inactive in the digesters (**Paper 1 & 2**). Our survey confirms the separate clustering of samples from the different reactor conditions (**Figure 3**) and indicate that the microbial communities were dominated by less than 300 OTUs that account for 80% of the reads (**Paper 1**).

The dominant methanogens in the mesophilic reactors were *Methanosaeta* followed by hydrogenoclastic methanogens such as *Methanolinea*, *Methanospirillum* and *Methanobrevibacter*. The thermophilic reactors were dominated by *Methanothermobacter* followed by *Methanosarcina* (**Paper 1**). The most abundant bacteria in the mesophilic reactors were the previously uncharacterised T78 phylotype belonging to Chloroflexi accounting on average for 2-13% of the reads in the anaerobic digesters indicating an obvious target for further studies. The thermophilic reactors had a high abundance of the G35_D8 phylotype, which represented on average 4-9% of the reads and thus represents another target for further studies (**Paper 1**). To allow for comparison between the microbial analysis of sludge digesters and the extensive collection of data from activated sludge systems we have used the same DNA extraction kit, primers, PCR settings as provided by the MiDAS initiative to characterise the bacterial community (Albertsen *et al.*, 2015; McIlroy *et al.*, 2015). We also tested multiple primer pairs for targeting the archaea and evaluated their performance based on a comparison to a PCR free metagenome dataset.

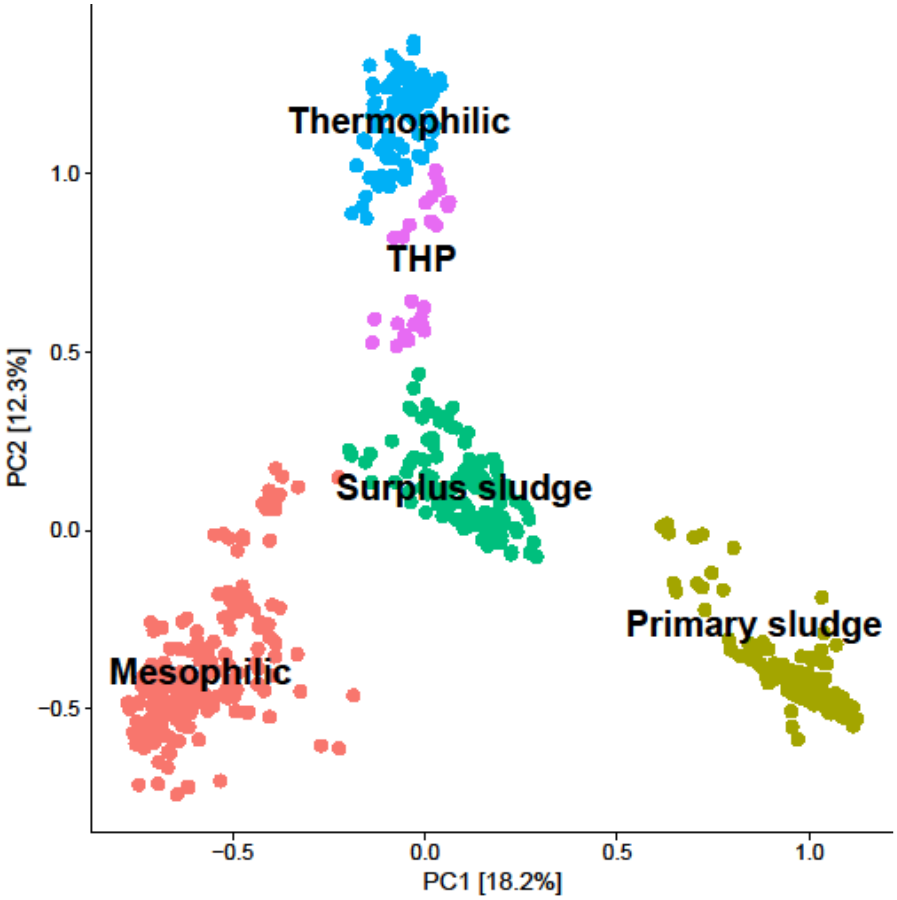

Figure 3 *"Principal component analysis of the bacterial communities analysed in this study highlighting samples by process type information. Mesophilic (), thermophilic (), and mesophilic with thermal hydrolysis pre-treatment (THP), primary sludge (), and surplus sludge ()" (**Paper 1**).*

Table 1 *Overview of some of the differences in community profiling of anaerobic digesters that make comparison between studies flawed. *"Universal" means that the primers should target both bacteria and archaea.*

Target	DNA extraction	PMA	Region	Database	Reference
Bacteria and archaea separately	Fast DNA SPIN Kit for Soil	no	bV1-V9 aV1-V9	RDP	(Ariesyady *et al.*, 2007)
Bacteria and archaea separately	Enzymatic lysis, phenol/chloroform	no	bV1-V8 aV1-V8	RDP	(Rivière *et al.*, 2009)
Bacteria	Phenol/chloroform method	no	bV1-V2	Greengenes	(Werner *et al.*, 2011)
Bacteria	PowerSoil DNA extraction kit	no	bV1-V3	RDP	(Lee *et al.*, 2012)
"Universal"	Fast DNA SPIN Kit for Soil	no	V6-V8	Greengenes	(Ho *et al*., 2013)
"Universal"	Fast DNA SPIN Kit for Soil	no	V3-V4	RDP	(Sundberg *et al*., 2013)
"Universal"	PowerSoil DNA extraction kit	no	V3	RDP,SILVA, Greengenes	(Kougias *et al*., 2014)
Bacteria and archaea separately	E.Z.N.A Soil DNA kit	no	bV1-V3 aV3-V5	SILVA	(Yu *et al*., 2014)
"Universal"	Fast DNA SPIN Kit for Soil	no	V4-V6	Greengenes	(Narihiro *et al*., 2015)
"Universal"	Fast DNA SPIN Kit for Soil	no	V4	Greengenes	(Vrieze *et al*., 2015b)
Archaea	PowerSoil DNA extraction kit	no	aV1-V2	NCBI	(Wilkins *et al*., 2015)

Bacteria and archaea separately	PowerSoil DNA extraction kit	no	bV1V3 aV3-V6	Greengenes	(Abendroth *et al.*, 2015)
"Universal"	Fast DNA SPIN Kit for Soil	no	V5-V6	SILVA	(Vrieze *et al.*, 2015a)
"Universal"	Fast DNA SPIN Kit for Soil	no	V4-V5	Greengenes	(Mei *et al.*, 2016)
"Universal"	Fast DNA SPIN Kit for Soil	no	V4	Greengenes	(Mosbæk *et al.*, 2016)
"Universal"	Fast DNA SPIN Kit for Soil	no	V4	Greengenes, RDP and NCBI	(Seib *et al.*, 2016)

1.3.4. DISCOVERY OF NOVEL ORGANISMS

To discover novel organisms it is important that identification methods do not rely too heavily on prior knowledge. This is an inherent problem when it comes to amplicon sequencing as it relies on primers designed based on what is already known and present in the reference databases. Databases from which we already know that primers cannot be designed to be truly universal (Klindworth *et al.*, 2013). The sequences deposited in the reference databases are mostly PCR derived and it is thus likely that the conservation of primer sites is overestimated (Brown *et al.*, 2015).

To circumvent this, we have developed a method that allows the sequencing of the 16S rRNA gene sequence without the need for primers and through molecular tagging also allows the sequencing of not just a fragment, but the entire sequence spanning all the variable regions on short read sequencers (**Paper 4**). Furthermore, we demonstrated a solution using the molecular tags to sequence the entire 16S rRNA gene and provide high accuracy sequences by consensus sequencing on the error prone long read sequencer MinION. These methods, provides a solution with much better taxonomic resolution, due to the increased sequence length and comes with no discovery bias, as no prior information about the sequence composition is needed. In addition, this method also provides sequences from all the three known domains of life unlike the "universal" primers that generally targets subsets of bacteria and archaea but misses eukaryotes (**Table 1**) (Karst *et al.*, 2016a). We demonstrated this method in an anaerobic digester among other environments (**Figure 4**), and found that

even in just one sample there is considerable bacterial diversity without close matches in the reference databases supporting the need for additional reference sequences.

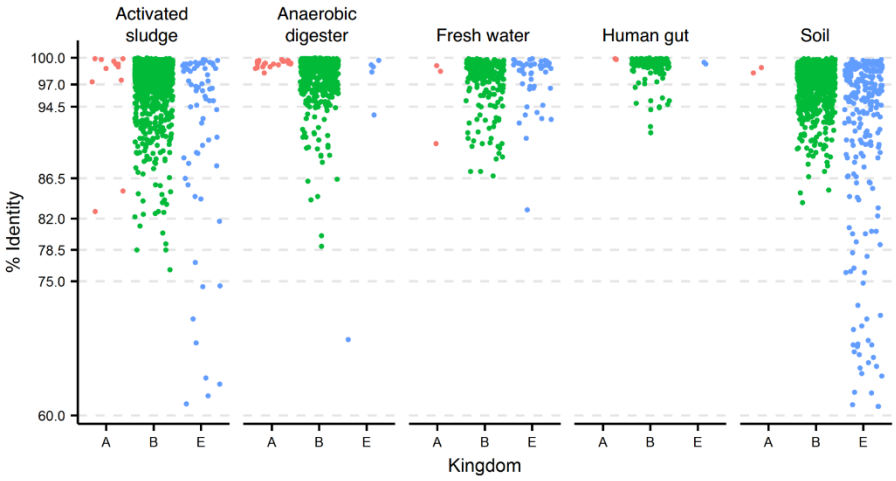

Figure 4 *Sequence identity (%) to the best match between the full-length SSU sequences obtained from the different environments to the Silva database for Archaea (A), Bacteria (B), and Eukarya (E). Each point represents a SSU sequence* (Karst *et al.*, 2016a).

To obtain novel genomes efficiently it is key to use the above-mentioned techniques to identify samples enriched with novel microorganisms using high throughput screening. Such screenings will also contribute to improve the general census about microbes and more specifically in anaerobic digesters to establish the list of important organisms (Schloss *et al.*, 2016; Rivière *et al.*, 2009).

1.4. GENOME RECOVERY FROM ANAEROBIC DIGESTERS

Following the identification of samples with novel and active target organisms, it is necessary to employ strategies and tools to obtain the genomes from the complex mixtures of organisms (**Figure 5**). This has led to two very different strategies enrichment towards ultimately isolation in the lab or sequencing of DNA from a mixed community and subsequent isolation of the sequences using bioinformatics.

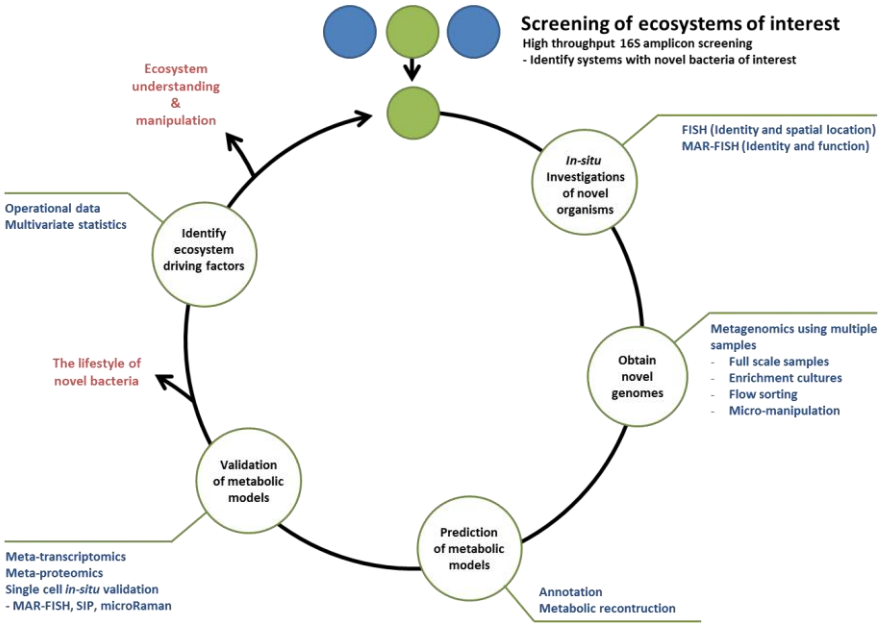

Figure 5 *Framework for identifying novel microorganisms and studying their way of life through a genomic approach.*

1.4.1. PURE CULTURE AND SINGLE CELL SEQUENCING

Classical microbiology has relied on isolation of microbes into pure culture and since the first bacterial genome was sequenced in 1995 plenty of microbial genomes have followed and it is now routine to sequence pure culture genomes (Fleischmann *et al.*, 1995; Fraser *et al.*, 1995; Land *et al.*, 2015). However, even though many organisms can now be cultured in the lab using various strategies to mimic the environment it is still very time consuming and somewhat luck based (Stewart, 2012).

The need for genomes without the requirement for culturing has spurred the development of single cell sequencing techniques to obtain genomes from mixed cultures relying on enrichment of the target species by sorting of cells and colonies or even sequencing the DNA of a single cell (Blainey, 2013; Lim *et al.*, 2015). However, this process is also relatively time consuming and the estimated completeness of the genomes derived with these methods is often very low with a notable risk of contamination (Parks *et al.*, 2015; Blainey, 2013).

1.4.2. GENOME RECOVERY FROM METAGENOMIC SEQUENCING

Metagenomic sequencing is an alternative approach to obtain genomes, from a complex sample, without the need for tedious and time-consuming enrichment in the

lab. DNA from all the organisms is sequenced, assembled and subsequently separated into the original genomes in a process termed genome binning. The first genomic binning strategies relied on species-specific signatures in the DNA such as GC content, k-mer distributions etc. (Mande *et al.*, 2012). However, the real breakthrough in recovery of genomes from metagenomic sequencing came with the sequence independent binning utilising the relationship between abundance of an organism and sequence coverage in multiple samples (**Figure 6**) (Sharon *et al.*, 2013; Albertsen *et al.*, 2013). This has led to a dramatic increase in available software tools that allow supervised to completely automated binning of metagenomic data into genomes based on differential abundance in multiple samples (Karst *et al.*, 2016b; Imelfort *et al.*, 2014; Alneberg *et al.*, 2014; Kang *et al.*, 2015). However, with the explosion of genome bins derived from metagenomic sequencing there is a need for quality control to ensure that the databases are not flooded with incomplete and contaminated genomes (Parks *et al.*, 2015). The quality control often depends on "universal" single copy genes that can be used to estimate completeness and in the case of duplicate genes indicate contamination (Parks *et al.*, 2015). Metagenomic surveys of anaerobic digesters have deposited genome bins to the public databases with very varying quality determined by these methods and completeness estimates spanning from 8-100% as well as contamination levels from 0-2902% (**Table 2**). This observation highlights that the genome quality problem is very real if genome references should be a resource to the scientific community (Parks *et al.*, 2015; Markowitz *et al.*, 2015).

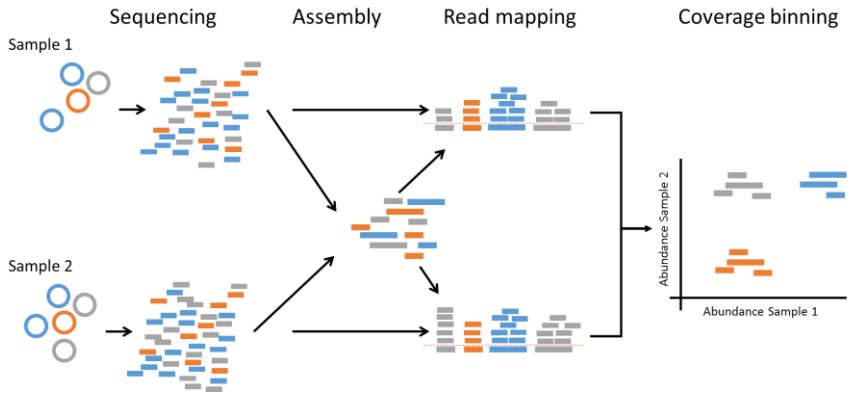

Figure 6 *Concept of genome recovery from complex samples. Metagenomic sequencing, assembly, read mapping and coverage based genome binning.*

The genome bins often consists of multiple contigs or scaffolds rather than a single contiguous sequence. This fragmentation of the genome sequence is the result of a number of problems with metagenomic sequencing for instance the short read lengths of the most common sequencing platforms. When the short reads do not span repeat regions they cannot resolve where the individual contigs should be connected (Alneberg *et al.*, 2014; Williamson *et al.*, 2016). To circumvent this problem a

multitude of methods have been developed, it is possible to provide linking information between contigs using mate-pair sequencing, make artificial long reads or even simply read longer fragments of DNA (Mardis, 2013; Eisenstein, 2015). These long read and hybrid approaches have shown promising results for closing genomes in pure and simple cultures (Loman *et al.*, 2015; Daims *et al.*, 2015). However, data generation from the first iterations have not been enough for tackling complex systems yet and the current iterations still have to prove their worth in the field (Jain *et al.*, 2016).

Following the retrieval of genomes from novel organisms in complex settings, it is important to test and validate the findings from the genomic potential. Even though it is at present difficult to obtain completely finished genomes, high quality genome bins have been powerful for generation of hypotheses about their metabolism for further testing in complex settings as well as used to design DNA probes for visualisation *in situ* (Sekiguchi *et al.*, 2015; Kirkegaard *et al.*, 2016). The visualisation can be complemented using a suite of isotope labelling techniques to help assist in the process of determining who eats what, where and when (Neufeld *et al.*, 2007).

Automated coverage binning solutions use statistical power to separate genomes and thus need many related samples with different abundance patterns, demonstrated for 11+ samples (Alneberg *et al.*, 2014; Imelfort *et al.*, 2014; Kang *et al.*, 2015). However, in many cases that amount of samples will be difficult or too expensive to get and sequence. In addition, small contigs with repetitive elements such as the 16S rRNA gene will have coverage profiles that are considerably off and are thus often missing in the bins produced by unsupervised methods. We have developed the mmgenome package that allows visual coverage binning of genomes from as little as two samples (Karst *et al.*, 2016b). The toolbox uses the open source environment R and reproducible binning can be achieved and documented using RMarkdown files (R Core Team, 2016). In addition to coverage, the package allows the user to benefit from sequence signatures such as the GC content, colour contigs by phylogenetic classification and use paired-end or mate-pair information to pull in contigs with different coverage profiles such as multi copy genes e.g. the 16S rRNA gene. We have used the mmgenome tool to extract genomes from full-scale anaerobic digesters for organisms representing the first genomic information for the phylum previously known as Hyd24-12 and a novel genus within the Anaerolineaceae (**Paper 5 & 6**). Based on the obtained sequence information we have also designed FISH probes for visualisation of the microbes *in situ*.

Table 2 Overview of some of the genome recovery studies from anaerobic digesters

No. of bins	Completeness/contamination estimates (method)	Sequencing technology	Binning method	Reference
2	93-94% / NA (111 marker genes)	Illumina Nextera and Nextera matepair	Metagenomic sequencing and coverage binning using GrooPM (Imelfort *et al.*, 2014)	(Sekiguchi *et al.*, 2015)
10	54-118% / NA (139 conserved genes)	Ion Torrent	Metagenomic sequencing and binning using Metawatt (Strous *et al.*, 2012)	(Nolla-Ardèvol *et al.*, 2015)
3	86-91% / 1-2% (CheckM)	Illumina TruSeq PCR free	Metagenomic sequencing and coverage binning using mmgenome package (Karst *et al.*, 2016b)	(Kirkegaard *et al.*, 2016)
5	83-97% / 2-28% (CheckM)	KAPA-Illumina library creation	Metagenomic sequencing and binning using MetaBAT (Kang *et al.*, 2015)	(Stolze *et al.*, 2016)
106	15-99%/ 0-16% (CheckM)	Illumina TruSeq PCR free, Nextera	Metagenomic sequencing and coverage binning	(Campanaro *et al.*, 2016)
157	26-97% (107 "essential" genes)	Illumina Nextera XT	Metagenomic sequencing and coverage binning	(Treu *et al.*, 2016)
43	8-100%/ 0-2902% (CheckM)	Illumina & pacbio	Metagenomic sequencing and binning using PhyloPythia S+ (Gregor *et al.*, 2016)	(Hagen *et al.*, 2016)
101	59-100% / 0-11% (CheckM)	illumina	Metagenomic sequencing and coverage binning using GrooPM (Imelfort *et al.*, 2014)	(Vanwonterghem *et al.*, 2016)

1.5. CONCLUSION

A comprehensive survey of the microbiology of Danish digesters at wastewater treatment plants have been carried out using 16S rRNA gene amplicon sequencing. The analysis revealed that, in agreement with other studies, microbial communities clustered based on process temperature and pre-treatment **(Paper 1**). In addition, the survey revealed that most of the abundant organisms were shared among all the plants of similar operation and that only 300 OTUs were needed to account for 80% of the reads. The fact that the majority of the biomass was made up of a set of relatively few organisms that were shared is key as it makes it feasible to study the individual organisms in detail and generalise findings. A notable impact of influent microbes on the apparent microbial community in anaerobic digesters was detected in the survey. Some of the seemingly most abundant microorganisms were deemed as inactive in the digesters (**Paper 1**). With this knowledge a short list of "most wanted" organisms that are abundant, uncharacterised, and active in anaerobic digesters were identified and forms the basis for prioritising future efforts of studying the microbes in depth (**Paper 1**). Based on the survey, digester relevant taxa were added to the MiDAS database efforts curating the taxonomy and expanding the knowledge database with methanogens as well as digester specific bacteria (**Paper 2**).

An improved method for primer free discovery using full-length sequencing of the small subunit rRNA has been developed and tested in an anaerobic digester (**Paper 3**). It revealed that digesters still contain notable diversity within bacteria without good matches in the public databases. Even from just a single sample, sequences were obtained that may represent new genera, families, orders and classes based on the sequence identities of less than 94.5%.

The mmgenome package, an open source software tool for visual supervised genome extraction, was developed to recover genome bins from metagenomes and tested in the context of full-scale anaerobic digesters (**Paper 4, 5 & 6**). Genomes were recovered from novel and abundant organisms in mesophilic anaerobic digesters representing the first genomic information about members of the Candidate phylum Hyd24-12. This lead to the development of FISH probes, their visualisation *in situ* and prediction of metabolic capabilities (**Paper 5**). A closed genome was recovered from one of the most abundant taxa within the Chloroflexi in the mesophilic digesters previously only known from the 16S rRNA gene sequence. Based on the genomic information, FISH probes were designed, they were visualised and shown to co-locate *in situ* with *Methanosaeta* spp. (**Paper 6**).

1.6. PERSPECTIVES

Genomes assembled from short read metagenome sequencing are generally fragmented in the public databases and thus present as draft genome bins of contigs and scaffolds. However, microbial genomes need to be of high quality to have lasting value for the scientific community. We do not know what is missing in between the contigs and it could be of biological importance (Parks *et al.*, 2015; Williamson *et al.*, 2016). To ensure better quality of reference genomes in public databases they need high accuracy and to be present as a number of closed contigs that matches the biological entity of circular or linear genomes. Long reads will provide much easier assembly and the resulting closed genomes will form the foundation for improving biological understanding (Koren *et al.*, 2016). High sequence accuracy can be obtained by polishing the genome assemblies with complementary sequencing data with different error profiles leading to high quality reference genomes (Walker *et al.*, 2014; Sović *et al.*, 2016).

Getting the genome catalogue is just the first step towards fundamental biological understanding and ultimately process control. The genomes form the foundation for generating hypotheses about what the microbes are capable of, but to go beyond genetic potential the hypotheses need subsequent experimental validation and can be tested using a palette of techniques in combination. This includes various types of microscopy, gene expression profiling, protein profiling, characterization of enzymes and the entire suite of "omics" technologies (Neufeld *et al.*, 2007; Segata *et al.*, 2013; Gerlt *et al.*, 2011; Neufeld, 2016).

The path to better operation of digesters at wastewater treatment plants based on biologically informed control relies on the development of rapid identification of microbes, functional interpretation and supervised decision on actions needed to intervene. Close to real-time DNA sequencing using cheap and disposable sequencers will bring the rapid identification and quantification on location (Jain *et al.*, 2016). Long read sequencing will provide high quality genome references, *in situ* methods the functional information and database efforts will help link identity with function but all aspects need to be in place before the full potential can be realised (Jain *et al.*, 2016; Koren *et al.*, 2016; Neufeld *et al.*, 2007; McIlroy *et al.*, 2015).

CHAPTER 2. CONTRIBUTIONS TO OTHER PUBLICATIONS

Albertsen M, Karst SM, Ziegler AS, **Kirkegaard RH** and Nielsen "PH. Back to basics – the influence of DNA extraction and primer choice on phylogenetic analysis in activated sludge communities".
PLoSONE, July 2015. DOI: 10.1371/journal.pone.0132783

Daims, H , Lebedeva, E , Pjevac, P , Han, P , Herbold, C , Albertsen, M , Jehmlich, N , Palatinszky, M , Vierheilig, J , Bulaev, A , **Kirkegaard, RH** , Bergen, MV, Rattei, T , Berndinger, B , Nielsen, PH, Wagner M. "Complete nitrification by Nitrospira bacteria".
Nature, Nov. 2015. DOI: 10.1038/nature16461

McIlroy, S, Karst, S, Nierychlo, M, Dueholm, M, Albertsen, M, **Kirkegaard, RH**, Seviour, R, Nielsen, PH. "Genomic and in situ investigations of the novel uncultured Chloroflexi associated with 0092 morphotype filamentous bulking in activated sludge".
ISME, Feb. 2016. DOI: 10.1038/ismej.2016.14

Law, Y, **Kirkegaard, RH,** Cokro AA, Liu, X, Arumugam, K, Xie, C, Bjerregaard, MS, Moses, DD, Nielsen, PH, Wuertz, S, Williams, RH. "Integrative microbial community analysis reveals full-scale enhanced biological phosphorus removal under tropical conditions".
Scientific reports, May. 2016. DOI: 10.1038/srep25719

Karst, SM, Albertsen, M, **Kirkegaard, RH,** Dueholm, MS, Nielsen, PH. Molecular Methods. Chapter 8 (p. 285-323) in "Experimental Methods In Wastewater Treatment" edited by van Loosdrecht, M.C.M., Nielsen, P.H. Lopez-Vazquez, C.M. and Brdjanovic, D.
ISBN: 9781780404745 (Hardback), ISBN: 9781780404752 (eBook).
Published by IWA Publishing, London, UK.

BIBLIOGRAPHY

Abendroth C, Vilanova C, Günther T, Luschnig O, Porcar M. (2015). Eubacteria and archaea communities in seven mesophile anaerobic digester plants in Germany. *Biotechnol Biofuels* **8**: 87.

Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31**: 533–8.

Albertsen M, Karst SM, Ziegler AS, Kirkegaard RH, Nielsen PH. (2015). Back to basics - The influence of DNA extraction and primer choice on phylogenetic analysis of activated sludge communities. *PLoS One* **10**: e0132783.

Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, *et al.* (2014). Binning metagenomic contigs by coverage and composition. *Nat Methods* **11**: 1144–6.

Ariesyady HD, Ito T, Okabe S. (2007). Functional bacterial and archaeal community structures of major trophic groups in a full-scale anaerobic sludge digester. *Water Res* **41**: 1554–1568.

Ashelford KE, Chuzhanova N a., Fry JC, Jones AJ, Weightman AJ. (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* **71**: 7724–7736.

Bentley RW. (2002). Global oil and gas depletion: An overview. *Energy Policy* **30**: 189–205.

Blainey PC. (2013). The future is now: Single-cell genomics of bacteria and archaea. *FEMS Microbiol Rev* **37**: 407–427.

Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, *et al.* (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**: 208–211.

Campanaro S, Treu L, Kougias PG, De Francisci D, Valle G, Angelidaki I. (2016). Metagenomic analysis and functional characterization of the biogas microbiome using high throughput shotgun sequencing and a novel binning strategy. *Biotechnol Biofuels* **9**: 26.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, *et al.* (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* **108 Suppl**: 4516–4522.

Carballa M, Regueiro L, Lema JM. (2015). Microbial management of anaerobic digestion: exploiting the microbiome-functionality nexus. *Curr Opin Biotechnol* **33**: 103–11.

Chen Y, Fu B, Wang Y, Jiang Q, Liu H. (2012). Reactor performance and bacterial pathogen removal in response to sludge retention time in a mesophilic anaerobic digester treating sewage sludge. *Bioresour Technol* **106**: 20–26.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, *et al.* (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**: D141–D145.

Daims H, Lebedeva EVEV, Pjevac P, Han P, Herbold C, Albertsen M, *et al.* (2015). Complete nitrification by Nitrospira bacteria. *Nature* **528**: 504–509.

Debelius J, Song SJ, Vazquez-Baeza Y, Xu ZZ, Gonzalez A, Knight R. (2016). Tiny microbes, enormous impacts: what matters in gut microbiome studies? *Genome Biol* **17**: 217.

Eisenstein M. (2015). Startups use short-read data to expand long-read sequencing market. *Nat Biotechnol* **33**: 433–435.

Federhen S. (2012). The NCBI Taxonomy. *Nucleic Acids Res* **40**: D136--D143.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, *et al.* (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269**: 496–512.

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, *et al.* (1995). The Minimal Gene Complement of Mycoplasma genitalium. *Science (80-)* **270**: 397–403.

Gerlt JA, Allen KN, Almo SC, Armstrong RN, Babbitt PC, Cronan JE, *et al.* (2011). The enzyme function initiative. *Biochemistry* **50**: 9950–9962.

Gest H. (2004). The discovery of microorganisms by Robert Hooke and Antoni van Leeuwenhoek, Fellows of The Royal Society. *Notes Rec R Soc* **58**: 187–201.

Gregor I, Dröge J, Schirmer M, Quince C, McHardy AC. (2016). PhyloPythiaS+ : a self-training method for the rapid reconstruction of low-ranking taxonomic bins from metagenomes. *PeerJ* **4**: e1603.

Hagen LH, Frank JA, Zamanzadeh M, Eijsink VGH, Pope PB, Horn SJ, *et al.* (2016). Quantitative metaproteomics highlight the metabolic contributions of uncultured

phylotypes in a thermophilic anaerobic digester. *Appl Environ Microbiol* AEM.01955-16.

Ho DP, Jensen PD, Batstone DJ. (2013). Methanosarcinaceae and acetate-oxidizing pathways dominate in high-rate thermophilic anaerobic digestion of waste-activated sludge. *Appl Environ Microbiol* **79**: 6491–6500.

Hughes L. (2015). Water Management Is Highly Energy Consuming — But It Doesn't Have To Be. The Huffington Post.

Imelfort M, Parks D, Woodcroft BJ, Dennis P, Hugenholtz P, Tyson GW. (2014). GroopM: an automated tool for the recovery of population genomes from related metagenomes. *PeerJ* **2**: e603.

Jain M, Olsen HE, Paten B, Akeson M, Branton D, Daniel B, *et al.* (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* **17**: 239.

Jenicek P, Kutil J, Benes O, Todt V, Zabranska J, Dohanyos M. (2013). Energy self-sufficient sewage wastewater treatment plants: Is optimized anaerobic sludge digestion the key? *Water Sci Technol* **68**: 1739–1743.

Kang DD, Froula J, Egan R, Wang Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**: e1165.

Karst SM, Dueholm MS, McIlroy SJ, Kirkegaard RH, Nielsen PH, Albertsen M. (2016a). Thousands of primer-free, high-quality, full-length SSU rRNA sequences from all domains of life. *bioRxiv* **10**: 70771.

Karst SM, Kirkegaard RH, Albertsen M. (2016b). Mmgenome: a Toolbox for Reproducible Genome Extraction From Metagenomes. *bioRxiv* 59121.

van Kessel MAHJ, Speth DR, Albertsen M, Nielsen PH, Op den Camp HJM, Kartal B, *et al.* (2015). Complete nitrification by a single microorganism. *Nature* 1–17.

Kirkegaard RH, Dueholm MS, Mcilroy SJ, Nierychlo M, Karst SM, Albertsen M, *et al.* (2016). Genomic insights into members of the candidate phylum Hyd24-12 common in mesophilic anaerobic digesters. *ISME J* 1–13.

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.

Koch H, Galushko A, Albertsen M, Schintlmeister A, Spieck E, Richter A, *et al.* (2014). Growth of nitrite-oxidizing bacteria by aerobic hydrogen oxidation. *Science* **345**: 1052–4.

Koren S, Walenz BP, Berlin K, Miller JR, Phillippy AM. (2016). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *bioRxiv* 71282.

Kougias PG, Francisci D De, Treu L, Campanaro S, Angelidaki I, De Francisci D, *et al.* (2014). Microbial analysis in biogas reactors suffering by foaming incidents. *Bioresour Technol* **167**: 24–32.

Land M, Hauser L, Jun S-R, Nookaew I, Leuze MR, Ahn T-H, *et al.* (2015). Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics* **15**: 141–61.

Lee S-H, Kang H-J, Lee YH, Lee TJ, Han K, Choi Y, *et al.* (2012). Monitoring bacterial community structure and variability in time scale in full-scale anaerobic digesters. *J Environ Monit* **14**: 1893–905.

Lim SW, Tran TM, Abate AR. (2015). PCR-activated cell sorting for cultivation-free enrichment and sequencing of rare microbes. *PLoS One* **10**: e0113549.

Loman NJ, Quick J, Simpson JT. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data.

Mande SS, Mohammed MH, Ghosh TS. (2012). Classification of metagenomic sequences: Methods and challenges. *Brief Bioinform* **13**: 669–681.

Mardis ER. (2013). Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto Calif)* **6**: 287–303.

Markowitz VM, Chen I-MA, Chu K, Pati A, Ivanova NN, Kyrpides NC. (2015). Ten years of maintaining and expanding a microbial genome and metagenome analysis system. *Trends Microbiol* **23**: 730–41.

Martiny AC, Treseder K, Pusch G. (2013). Phylogenetic conservatism of functional traits in microorganisms. *ISME J* **7**: 830–8.

McCarty PL, Bae J, Kim J. (2011). Domestic wastewater treatment as a net energy producer - can this be achieved? *Environ Sci Technol* **45**: 7100–6.

Mcdonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, *et al.* (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* **6**: 610–618.

McIlroy SJ, Saunders AM, Albertsen M, Nierychlo M, Mcilroy B, Hansen AA, *et al.* (2015). MiDAS: the field guide to the microbes of activated sludge. *Database* **2015**: bav062.

Mei R, Narihiro T, Nobu MK, Kuroda K, Liu W-T. (2016). Evaluating digestion efficiency in full-scale anaerobic digesters by identifying active microbial populations through the lens of microbial activity. *Sci Rep* **6**: 34090.

Mosbæk F, Kjeldal H, Mulat DG, Albertsen M, Ward AJ, Feilberg A, *et al.* (2016). Identification of syntrophic acetate-oxidizing bacteria in anaerobic digesters. *ISME J* **2**: 1–14.

Narihiro T, Nobu MK, Kim NK, Kamagata Y, Liu WT. (2015). The nexus of syntrophy-associated microbiota in anaerobic digestion revealed by long-term enrichment and community survey. *Environ Microbiol* **17**: 1707–1720.

Neufeld JD. (2016). Migrating SSU rRNA gene surveys to the metagenomics era. *Environ Microbiol Rep* **0**: in press.

Neufeld JD, Wagner M, Murrell JC. (2007). Who eats what, where and when? Isotope-labelling experiments are coming of age. *ISME J* **1**: 103–110.

Nolla-Ardèvol V, Peces M, Strous M, Tegetmeyer HE. (2015). Metagenome from a Spirulina digesting biogas reactor: analysis via binning of contigs and classification of short reads. *BMC Microbiol* **15**: 277.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–55.

Patel GB, Sprott D. (1990). 'Methanosaeta concilii gen. nov. sp. nov. ("Methanothrix concilii") and Methanosaeta thermoacetophila nom. rev., comb. Nov'. *Int J Syst Bacteriol* **40**: 79–82.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, *et al.* (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res* **41**: 590–596.

R Core Team. (2016). R: A language and environment for statistical computing.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, *et al.* (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431–437.

Rivière D, Desvignes V, Pelletier E, Chaussonnerie S, Guermazi S, Weissenbach J, *et al.* (2009). Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. *ISME J* **3**: 700–714.

Rogelj J, den Elzen M, Höhne N, Fransen T, Fekete H, Winkler H, *et al.* (2016). Paris Agreement climate proposals need a boost to keep warming well below 2 °C. *Nature* **534**: 631–639.

Saady NMC. (2013). Homoacetogenesis during hydrogen production by mixed cultures dark fermentation: Unresolved challenge. *Int J Hydrogen Energy* **38**: 13172–13191.

Saunders AMAM, Albertsen M, Vollesen J, Nielsen PHPH, Vollertsen J, Nielsen PHPH. (2015). The activated sludge ecosystem contains a core community of abundant organisms. *ISME J* 1–10.

Schellnhuber HJ, Rahmstorf S, Winkelmann R. (2016). Why the right climate target was agreed in Paris. *Nat Clim Chang* **6**: 649–653.

Schloss PD, Girard RA, Martin T, Edwards J, Thrash JC. (2016). Status of the archaeal and bacterial census: an update. *MBio* **7**: e00201-16.

Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS, Huttenhower C. (2013). Computational meta'omics for microbial community studies. *Mol Syst Biol* **9**: 666.

Seib MDD, Berg KJJ, Zitomer DHH. (2016). Influent wastewater microbiota and temperature influence anaerobic membrane bioreactor microbial community. *Bioresour Technol* **216**: 446–452.

Sekiguchi Y, Ohashi A, Parks DH, Yamauchi T, Tyson GW, Hugenholtz P. (2015). First genomic insights into members of a candidate bacterial phylum responsible for wastewater bulking. *PeerJ* **3**: e740.

Sharon I, Morowitz MJ, Thomas BC, Costello EK, Relman DA, Banfield JF. (2013). Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Res* **23**: 111–120.

Smith SR, Lang NL, Cheung KHM, Spanoudaki K. (2005). Factors controlling pathogen destruction during anaerobic digestion of biowastes. *Waste Manag* **25**: 417–425.

Sović I, Križanović K, Skala K, Šikić M. (2016). Evaluation of hybrid and non-hybrid methods for de novo assembly of nanopore reads. *Bioinformatics* 30437.

Stewart EJ. (2012). Growing unculturable bacteria. *J Bacteriol* **194**: 4151–4160.

Stolze Y, Bremges A, Rumming M, Henke C, Maus I, Pühler A, *et al.* (2016). Identification and genome reconstruction of abundant distinct taxa in microbiomes from one thermophilic and three mesophilic production-scale biogas plants. *Biotechnol Biofuels* **9**: 156.

Strous M, Kraft B, Bisdorf R, Tegetmeyer HE. (2012). The binning of metagenomic contigs for microbial physiology of mixed cultures. *Front Microbiol* **3**: 1–11.

Sun L, Müller B, Westerholm M, Schnürer A. (2014). Syntrophic acetate oxidation in industrial CSTR biogas digesters. *J Biotechnol* **171**: 39–44.

Sundberg C, Al-Soud WA, Larsson M, Alm E, Yekta SS, Svensson BH, *et al.* (2013). 454 Pyrosequencing analyses of bacterial and archaeal richness in 21 full-scale biogas digesters. *FEMS Microbiol Ecol* **85**: 612–626.

Treu L, Kougias PG, Campanaro S, Bassani I, Angelidaki I. (2016). Deeper insight into the structure of the anaerobic digestion microbial community; the biogas microbiome database is expanded with 157 new genomes. *Bioresour Technol* **216**: 260–266.

Vanwonterghem I, Jensen PD, Ho DP, Batstone DJ, Tyson GW. (2014). Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. *Curr Opin Biotechnol* **27**: 55–64.

Vanwonterghem I, Jensen PD, Rabaey K, Tyson GW. (2016). Genome-centric resolution of microbial diversity, metabolism and interactions in anaerobic digestion. *Environ Microbiol* **June**: 3144–3158.

Vrieze J De, Gildemyn S, Vilchez-vargas R, Jáuregui R, Pieper DH, Verstraete W, *et al.* (2015a). Inoculum selection is crucial to ensure operational stability in anaerobic digestion. *Appl Microbiol Biotechnol* **99**: 189–199.

Vrieze J De, Saunders AM, He Y, Fang J, Nielsen PH, Verstraete W, *et al.* (2015b). Ammonia and temperature determine potential clustering in the anaerobic digestion microbiome. *Water Res* **75**: 312–23.

Wagner AO, Malin C, Knapp BA, Illmer P. (2008). Removal of free extracellular DNA from environmental samples by ethidium monoazide and propidium monoazide. *Appl Environ Microbiol* **74**: 2537–9.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, *et al.* (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome

assembly improvement. *PLoS One* **9**. e-pub ahead of print, doi: 10.1371/journal.pone.0112963.

Watanabe H, Kitamura T, Ochi S, Ozaki M. (1997). Inactivation of pathogenic bacteria under mesophilic and thermophilic conditions. In: Vol. 36. *Water Science and Technology*. pp 25–32.

Weiland P. (2010). Biogas production: Current state and perspectives. *Appl Microbiol Biotechnol* **85**: 849–860.

Werner JJ, Knights D, Garcia ML, Scalfone NB, Smith S, Yarasheski K, *et al.* (2011). Bacterial community structures are unique and resilient in full-scale bioenergy systems. *Proc Natl Acad Sci U S A* **108**: 4158–4163.

Wilkins D, Lu XY, Shen Z, Chen J, Lee PKH. (2015). Pyrosequencing of mcrA and archaeal 16s rRNA genes reveals diversity and substrate preferences of methanogen communities in anaerobic digesters. *Appl Environ Microbiol* **81**: 604–613.

Williamson CHD, Sanchez A, Vazquez A, Sahl JW. (2016). Bacterial genome reduction as a result of short read sequence assembly. *bioRxiv*. e-pub ahead of print, doi: 10.1101/091314.

Woese CR, Fox GE. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**: 5088–5090.

Yamada T, Sekiguchi Y, Hanada S, Imachi H, Ohashi A, Harada H, *et al.* (2006). Anaerolinea thermolimosa sp. nov., Levilinea saccharolytica gen. nov., sp. nov. and Leptolinea tardivitalis gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes Anaerolineae classis nov. and Caldilineae classis nov. in the . *Int J Syst Evol Microbiol* **56**: 1331–40.

Yu H, Wang Q, Wang Z, Sahinkaya E, Li Y, Ma J, *et al.* (2014). Start-up of an anaerobic dynamic membrane digester for waste activated sludge digestion: temporal variations in microbial communities. *PLoS One* **9**: e93710.

CHAPTER 3. PAPER 1 - IDENTIFYING THE ABUNDANT AND ACTIVE MICROORGANISMS COMMON TO FULL-SCALE ANAEROBIC DIGESTERS

Kirkegaard, RH, McIlroy, SJ, Kristensen, JM, Nierychlo, M, Karst, SM, Albertsen, M, Dueholm, MS, Nielsen, PH. *"Identifying the abundant and active microorganisms common to full-scale anaerobic digesters"*
(In preparation)

CHAPTER 4. PAPER 2 - MIDAS 2.0: A SITE-SPECIFIC TAXONOMY AND ONLINE DATABASE FOR THE ORGANISMS OF WASTEWATER TREATMENT SYSTEMS EXPANDED FOR ANAEROBIC DIGESTER GROUPS

McIlroy, SJ, **Kirkegaard, RH**, McIlroy, B, Nierychlo, M, Kristensen, JM, Karst, SM, Albertsen, M, Nielsen, PH. *"MiDAS 2.0: A site-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups"*
(Submitted to *Database*)

CHAPTER 5. PAPER 3 - THOUSANDS OF PRIMER-FREE, HIGH-QUALITY, FULL-LENGTH SSU RRNA SEQUENCES FROM ALL DOMAINS OF LIFE

Karst SM, Dueholm, MS, McIlroy, SJ, **Kirkegaard, RH**, Nielsen, PH, Albertsen, M. *"Thousands of primer-free, high-quality, full-length SSU rRNA sequences from all domains of life".*
BioRxiv, August. 2016. DOI: 10.1101/070771

CHAPTER 6. PAPER 4 - MMGENOME: A TOOLBOX FOR REPRODUCIBLE GENOME EXTRACTION FROM METAGENOMES

Karst, SM, **Kirkegaard, RH,** Albertsen, M. *"mmgenome: a toolbox for reproducible genome extraction from metagenomes"*.
BioRxiv, June. 2016. DOI: 10.1101/059121

CHAPTER 7. PAPER 5 - GENOMIC INSIGHTS IN MEMBERS OF THE CANDIDATE PHYLUM HYD24-12 COMMON IN MESOPHILIC ANAEROBIC DIGESTERS

Kirkegaard, RH, Dueholm, MS, McIlroy, SJ, Nierychlo, M, Karst, S, Albertsen, M, Nielsen, PH. *"Genomic insights in members of the Candidate phylum Hyd24-12 common in mesophilic anaerobic digesters"*
Published ISME, Mar. 2016. DOI: 10.1038/ismej.2016.43

CHAPTER 8. PAPER 6 - GENOME ANALYSIS OF A REPRESENTATIVE OF THE UNCULTURED A6 PHYLOTYPE WITHIN THE ANAEROLINEACEAE REVEALS THEIR LIKELY ROLE AS PRIMARY FERMENTERS ABUNDANT IN MESOPHILIC ANAEROBIC DIGESTERS

McIlroy, SJ, **Kirkegaard, RH**, Dueholm, MS, Karst, SM, Albertsen, M, Nielsen, PH. *"Genome analysis of a representative of the uncultured A6 phylotype within the Anaerolineaceae reveals their likely role as primary fermenters abundant in mesophilic anaerobic digesters"*
(In preparation)

ISSN (online): 2246-1248
ISBN (online): 978-87-7112-856-7

AALBORG UNIVERSITY PRESS