รายงานวิจัยฉบับสมบูรณ์

โครงการ การพัฒนาวิธีทางชีวสารสนเทศและเมตาจีโนมิก สำหรับการปรับปรุงระบบบำบัดน้ำเสียจากโรงงานยาง

Development of bioinformatic and metagenomic approaches for the improvement of latex wastewater treatment system

โดย ดร.กานต์ธิดา กุศลมโน และคณะ

พฤษภาคม 2563

สัญญาเลขที่ MRG6080183

รายงานวิจัยฉบับสมบูรณ์

โครงการ การพัฒนาวิธีทางชีวสารสนเทศและเมตาจีโนมิก สำหรับการปรับปรุงระบบบำบัดน้ำเสียจากโรงงานยาง

Development of bioinformatic and metagenomic approaches for the improvement of latex wastewater treatment system

คณะผู้วิจัย สังกัด

1	ดร. กานต์ธิดา กุศลมโน	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
2	ดร. พิชาภัค อุทัยไพศาลวงศ์	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
3	นายธนาวุธ ศรีสุข	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
4	น.ส. นิมารดี บุญอาพัทธิ์เจริญ	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
5	ดร. พรพรรณ พาณิชย์นำสิน	ศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพแห่งชาติ
6	รศ.ดร. สุภาภรณ์ ชีวะธนรักษ์	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

สนับสนุนโดยสำนักงานคณะกรรมการการอุดมศึกษา และสำนักงานกองทุนสนับสนุนการวิจัย

บทคัดย่อ

การย่อยสลายสารอินทรีย์แบบไร้อากาศถูกแนะนำให้เป็นทางเลือกที่มีความหวังสำหรับระบบบำบัดน้ำเสียในอุตสาหกรรม โดยไม่เพียงแต่ใช้สำหรับการบำบัดน้ำเสียแต่สามารถให้ก๊าซชีวภาพสำหรับเป็นพลังงานทางเลือกได้อีกด้วย สำหรับน้ำเสียจากโรงงานยางเป็นน้ำเสียที่มีความเข้มข้นของซัลเฟตสูง เป็นสาเหตุให้การบำบัดน้ำเสียและการผลิตก๊าซมีเทนไม่มีประสิทธิภาพ การเข้าใจถึงกลุ่มจุลินทรีย์ซึ่งเป็นตัวขับเคลื่อนการย่อยสลายสารอินทรีย์แบบไร้อากาศเป็นสิ่งสำคัญอย่างยิ่งในการปรับปรุงระบบบำบัดน้ำเสีย การศึกษาแบบเมตาจีโนมิกเป็นการใช้เทคโนโลยีการหาลำดับนิวคลีโอไทด์ในปริมาณมากในการอ่านสารพันธุกรรมของจุลินทรีย์ในสภาวะแวดล้อมได้โดยตรงโดยไม่ต้องอาศัยการเลี้ยงเชื้อ ในงานวิจัยนี้ ผู้จัดทำได้ใช้เทคนิคแบบเมตาจีโนมิกศึกษาจุลินทรีย์ในระบบบำบัดน้ำเสียโรงงานยางแบบ Multi-stage treatment system โดยไปป์ไลน์ทางชีวสารสนเทศสำหรับการวิเคราะห์ข้อมูลเมตาจีโนมิกได้ถูกพัฒนาขึ้น ซึ่งให้ผลลัพธ์เป็นรูปแบบและการทำงานของกลุ่มจุลินทรีย์ การศึกษาแสดงให้เห็นว่ารูปแบบกลุ่มจุลินทรีย์มีการเปลี่ยนไปในสภาวะที่ระบบมีประสิทธิภาพสูงสุดเมื่อเทียบกับสภาวะอ้างอิง โดยการศึกษานี้พบกลุ่มแบคทีเรียรีดิวซ์ซัลเฟต ได้แก่ *Desulfovibrionaceae* *Desulfomicrobiaceae* และ *Desulfobacteraceae* และแบคทีเรียในกลุ่มซัลไฟด์ออกซิไดซิงค์ ได้แก่ *Spirochaetaceae* *Rhodobacteraceae* *Campylobacteraceae* *Comamonadaceae* และ *Burkholderiaceae* เป็นต้น เมทาโนเจนที่ถูกค้นพบในปริมาณมากเป็นชนิดที่เปลี่ยนอะซิเตตให้เป็นมีเทน ได้แก่ *Methanosaetaceae* และ *Methanosarcinaceae* เป็นที่น่าสนใจที่หนึ่งในแบคทีเรียรีดิวซ์ซัลเฟต คือ *Desulfovibrio vulgaris* ถูกพบในปริมาณมากในถังหมักที่ให้ประสิทธิภาพสูง แบคทีเรียนี้เป็นกลุ่มที่ออกซิไดซ์ซัลเฟตอย่างไม่สมบูรณ์ ทำให้ได้อะซิเตตเป็นผลผลิต ซึ่งเป็นสารอาหารสำหรับเมทาโนเจนในการผลิตก๊าซมีเทน โครงการนี้ให้ผลลัพธ์เป็นไปป์ไลน์และแนวทางปฏิบัติสำหรับการวิเคราะห์ข้อมูลเมตาจีโนมิก รวมถึงองค์ความรู้เชิงลึกของกลุ่มจุลินทรีย์ในระบบบำบัดน้ำเสียที่มีซัลเฟตสูง โดยกลุ่มจุลินทรีย์เหล่านี้สามารถนำไปศึกษาต่อและประยุกต์ใช้สำหรับการจัดการกลุ่มจุลินทรีย์เพื่อเพิ่มเสถียรภาพและประสิทธิภาพของระบบบำบัดน้ำเสียแบบไร้อากาศจากโรงงานยางได้

คำสืบค้น: การวิเคราะห์ข้อมูลแบบเมตาจีโนมิก ไมโครไบโอม ชีวสารสนเทศ การย่อยสลายสารอินทรีย์แบบไร้อากาศ ระบบบำบัดน้ำเสียโรงงานยาง น้ำเสียที่มีซัลเฟตสูง แบคทีเรียรีดิวซ์ซัลเฟต

Abstract

Anaerobic digestion (AD) has been introduced as a promising solution for industrial wastewater treatment systems, not only for wastewater treatment but also biogas production as renewable energy. Particularly, latex wastewater containing high concentration of sulfate causes both inefficient water treatment and methane production. Understanding the microbial communities driving AD process is a key success for improving the treatment systems. Metagenomics employs high-throughput technologies to direct sequence genetic materials and reveal all microbes from a particular environment without cultivation. In this work, we conducted shotgun metagenomic approach to study microbiome in multi-stage latex wastewater treatment system. Bioinformatics pipeline for analyzing microbiome from metagenomic data has been developed providing microbial and functional profiles. The study showed a shift of microbial profiles from control to the optimal performance reactors. A group of sulfate reducing bacteria (SRB) was found in our study such as *Desulfovibrionaceae*, *Desulfomicrobiaceae* and *Desulfobacteraceae*. Diverse families of sulfide oxidizing bacteria (SOB) were discovered, for example, *Spirochaetaceae*, *Rhodobacteraceae*, *Campylobacteraceae*, *Comamonadaceae* and *Burkholderiaceae*. Acetoclastic methanogens, *Methanosaetaceae Methanosarcinaceae*, were prevalently revealed in the system playing an important role in methane production. Interestingly, one of the discovered SRB, *Desulfovibrio vulgaris*, was found with higher abundance in the optimal performance reactor. It is a group of incompletely-oxidizing sulfate reducer providing acetate as a product, which could be a substrate of methanogens. This project provides a pipeline and practical guideline for metagenomic analysis as well as insight information of microbial community in a sulfate-rich wastewater treatment system. The communities could be further studied and applied for microbial management and manipulation to increase the stability and efficiency of the anaerobic latex wastewater treatment.

Keywords: Metagenomic analysis, Microbiome, Bioinformatics, Anaerobic digestion, Latex wastewater treatment, Sulfate-rich wastewater, Sulfate reducing bacteria

Executive Summary

This research project provides insight scientific knowledge in bioinformatics and biotechnology applied for anaerobic digestion (AD) for wastewater treatment and biogas production as a source of renewable energy. Even though several industries are interested in utilization of the technology, the instability of the system is still an issue. Several factors causing the problem including sulfate-rich condition in e.g. latex wastewater treatment system. A research focusing on the problem is needed and a key success is the understanding of microbial communities driving the AD process. Gaining such knowledge provides better control of the systems. With the high-throughput technology nowadays, a community of microbes can be studied though genetic materials without cultivation. The technology is known as metagenomics, allowing a discovery of massive numbers of microbes. Nevertheless, such data cannot be analyzed by hand and needed deep specific knowledge to extract information called bioinformatics. It is important to build a workflow to analyze the data to get informative and accurate results. Thus, this project serves two main purposes of (1) developing a bioinformatic pipeline for analyzing metagenomic data, and (2) analyzing microbiome data in latex wastewater treatment system. For the first purpose, the constructed pipeline facilitates the analysis of microbial communities through the high-throughput genetic sequences. The pipeline cannot only be used in this project but also provides a practical guideline for analyzing other communities in similar or different area of microbiome research. Secondly, the analysis of microbiome in latex wastewater treatment system provides a high-resolution of microbes inside the system. Several sulfate reducing bacteria (SRB) and sulfide oxidizing bacteria (SOB) were discovered in the system. Acetoclastic methanogens were found as main methane producers. Interestingly, one of detected SRB, *Desulfovibrio vulgaris*, was found with higher abundance in the high-performance system compared to control. This is interesting as SRB is usually found as a competitor to methanogens, microbes producing methane. However, *D. vulgaris* was found as a group of incompletely-oxidizing sulfate reducer providing acetate as a product, which could be a substrate of methanogens instead (mainly *Methanosaeta harundinacea* in this system). The study suggests a pair or group of microbial communities living together in the AD system providing good performance of both sulfate removal and methane production. The discovery could be further studied and applied in the industry towards environmental friendly treatments and the use of renewable energy. In addition, the project provides related scientific outputs publicly for research communities as paper, conference proceedings and presentations (please see appendix).

Acknowledgement

We acknowledge the financial support of this project (MRG6080163) from the Thailand Research Fund as well as Office of the Higher Education Commission. We thank facilities and high performance computing supports at King Mongkut's University of Technology Thonburi. We would like to express our great appreciation for fruitful scientific discussion and friendly environments from the teams at Systems Biology and Bioinformatics and Excellent Center of Waste Utilization and Management, King Mongkut's University of Technology Thonburi.

Table of contents

List of tables

List of figures

Chapter 1
Introduction

1.1 Introduction to the research problem and its significance

Anaerobic digestion (AD) has become a promising solution for wastewater treatment since it enables both pollution control and generation of renewable energy. It provides biomaterials degradation by microorganisms in the absence of oxygen. The degradation processes result formation of biogas (i.e. methane and carbon dioxide), which has been considered to contribute as sustainable energy. In addition, the AD system consumes low energy that causes lower cost for the water treatment comparing to other systems that require high energy such as the aerated lagoon system. The AD systems have been constructed especially for agricultural and industrial wastewater, which contain high levels of biodegradable materials. Nevertheless, the operational stability of the AD system is still poor e.g. yielding low biogas production, which prevents the technique from being widely applied and commercialized.

In addition, wastewater from the latex factories contains high organic, ammonia and sulfate concentration. With the high sulfate concentration, wastewater treatment systems of latex factories contain more pollution problems both air and water pollutions, and also system instabilities according to the toxic of the sulfide producing during the process. Sulfate reducing bacteria (SRB) is a known group of microbes changing sulfate to sulfide, which is toxic to environment and microbes in the AD system especially methanogen. In addition, SRB is a competitor to methanogen as final oxidizers in the AD process. These problems cause low efficiency of water treatment and biogas production.

Understanding the microbial communities and their functions is a key step for microbial management in AD wastewater system in order to optimize the system performance and prevent system failure. Even though the AD is a well-known technology, the functioning microbiomes and their interactions are still not completely understood. This could be because of the complexity of the anaerobic microbiome and the high number of uncharacterized microorganisms. Discovery of the fundamental knowledge of the microbial communities and their behaviors corresponding to environmental changes and process disturbances would lead a way to monitor and control microbial community in the AD system. For example, microbial indicators can be set for detecting optimal performance of a reactor or warning of process failure. Discovery of a microbial community enhancing biogas production could also be applied. The

efficient management of the system would lead to a significant improvement of the system performance especially in biogas production.

With the availability of the next generation sequencing (NGS) technology, living cells can be studied in molecular levels in a high-throughput manner. The technology provides an opportunity to read all genetic materials (i.e. DNA or RNA) in a collection of tons of short fragments from living cells, including microbial communities. The later one is known as metagenomics, the study of overall DNAs of a particular sample from an environment. The study is not limited to microbiomes in wastewater systems. Unlike the conventional microbial techniques that are culture-dependent techniques, metagenomics extracts DNAs directly from a sample of the study without cultivation. The technique provides more complete picture of microorganism structures (diversity and abundance), as so far only a small part of microorganisms could be cultivated. In addition, by performing whole-genome shotgun sequencing, the technique yields not only the inside structures of microbial communities but also information of functional genes in the system. The metagenomics using whole-genome shotgun sequencing so far is a powerful tool to study microbial communities, functions and dynamics to disturbances.

With the high amount of sequences that will be read from the metagenomic study (could reach to millions of short read sequences), it is too far to interpret the data by hands. There is a strong need of bioinformatics, which includes several branches of sciences e.g. statistics, computer science, biology, and other related disciplines, to manage and analyze the data. Development of a pipeline that is a flow of integrated bioinformatic methods and tools is essential. The pipeline will enable the extraction of biological knowledge from the DNA community data.

In order to achieve efficient management of the sulfate-rich wastewater system for optimal performance in wastewater treatment and high biogas production, this project will be the insight study of microbial community structures, functions and dynamics through metagenomics. The project will include the development of a bioinformatic pipeline to analyze the metagenomic data for extracting biological information leading better understanding of the anaerobic digestion in the industrial latex wastewater system.

1.2 Objectives

1. To develop a bioinformatic pipeline for analyzing metagenomic data of anaerobic wastewater treatment systems

2. To analyze metagenomic data of anaerobic microbiomes in latex wastewater for better understanding of the microbial community structures, functions and dynamisms

1.3 Scope of research

In this project, the bioinformatic analysis of metagenomic data from sulfate rich wastewater system will be performed. The pipeline and analysis is based on high-throughput sequencing technology of short reads. The study includes two main steps as following.

1. Development of a bioinformatic pipeline for analyzing metagenomic data from a wastewater treatment system

    1.1 Development of modules for analyzing metagenomic data, which are taxonomic analysis (reads-based methods), functional analysis and comparative metagenomics

    1.2 Development of databases for taxonomic and functional analyses of metagenomic data from AD systems

    1.3 Modules integrations into an analytical pipeline linking to the developed database for analyzing of metagenomic data

    1.4 Use of open source software/tools

2. Metagenomic analysis of anaerobic metagenomes in sulfate rich wastewater system by conducting the developed bioinformatic pipeline

    2.1 Data pre-processing of each metagenomic sample

    2.2 Taxonomic analysis of each metagenomic sample

    2.3 Functional analysis of each metagenomic sample

    2.4 Comparative metagenomics between different samples

1.4 Expected output

1. Bioinformatic pipeline for analyzing metagenomic data of anaerobic wastewater treatment systems

2. Microbial and functional profiles of microbial communities in the studied latex wastewater treatment system

1.5 Expected outcome

The understanding of the microbial structures, functions and dynamics at different bioreactor performance from this study could provide knowledge guiding better microbial management and monitoring the sulfate-rich wastewater system. For example, abundance of specific species or functional genes, or microbial diversity patterns could be used to detect the system at optimal performance or warn before failure. This project could provide a preliminary study for further studying microbiome responding to different disturbances or factors, helping to control the communities for obtaining higher efficiency of anaerobic wastewater system.

Chapter 2
Literature review

2.1 Anaerobic digestion and biogas production

Anaerobic digestion (AD) provides biogas as a final product which can be use as rewable energey. The use of the biogas to replace fossil fuels could reduce generation of greenhouse gases simultaneously with wastewater treatment. The biogas production occurs by digestion of various organic compounds which are controlled by diverse microorganisms under the anoxic environment comprising four main processes of hydrolysis, acidogenesis, acetogenesis and methanogenesis [1, 2].

A primary step of AD is hydrolysis that complex insoluble polymers are decomposed into the soluble monomers (e.g. amino acids, small molecule sugar and fatty acids) by hydrolytic enzymes [1]. These enzymes are secreted by specific microorganisms belonging in the class of *Clostridia* and *Bacilli*, for example, *Clostridium thermocellum* and *Enterococcus faecalis* which were reported as cellulose-degrading bacteria [3].

The second process is acidogenesis. this step is an acid-forming step that fermentative microorganisms referred to as acidobacteria playing an important role to convert hydrolysis resulting products to acidic intermediates such as volatile fatty acids (VFAs) or alcohols [1]. VFAs is a short-chain fatty acids which structure comprises of two- to five-carbon atoms such as acetic acid (C_2), propionic acid (C_3), butyric acid (C_4), valeric acid (C_5) and others. The representative bacteria of this process were reported as *Clastridium acetibutilicum*, *Clostridium perfingens, Enterococcusfaecium, Lactobacillus helveticus* and etc [3].

Acetogenesis is a conversion step of acidic intermediates produced from hydrolysis and acidogenesis processes, to generate acetic acid (C_2), CO_2 and H_2, which are consequently utilized by biogas producing microbes [1]. Various anaerobes were reported as key organisms to syntrophically produce acetate and H_2, for example, *Syntrophobacter*, *Syntrophomonas*, *etc.* which are propionate- and butyrate-decomposers [4, 5], respectively. Moreover, acetate could be produced by homoacetogenic bacteria or homoacetogen which utilize H_2/CO_2 as a carbon source [6]. An example of the homoacetogen is *Acetobacterium* sp [7].

The final step is methanogenesis, an important process to yield methane and CO_2. The biogas producing microbes, also known as methanogens, belong mainly in the archaea domain that play a crucial role in methane production. Methanogens could be categorized into two groups of acetoclastic methanogen (AM) and hydrogenotrophic methanogen (HM). According to the

process of methane production, acetate is converted by AM, whereas CO_2 and H_2 are syntrophically converted by HM with contribution of bacterial partner called syntrophs [1]. Several AMs have been reported, for instance, *Methanosaeta* sp. and *Methanosarcina* sp. while *Methanobactereium sp., Methanobrevibacter sp., Methanomicrobium sp. Methanococcus sp.* have been reported as HMs [2].

2.2 Anaerobic digestion in sulfate-rich wastewater system

Anaerobic digestion is a promising solution for wastewater treatment since it enables both pollution control and generation of renewable energy [8]. It provides biomaterials degradation by microorganisms in the absence of oxygen. The degradation processes result formation of biogas (i.e. methane and carbon dioxide), which has been considered to contribute as sustainable energy. In addition, the AD system consumes low energy that causes lower cost for the water treatment comparing to other systems that require high energy such as the aerated lagoon system. The AD systems have been constructed especially for agricultural and industrial wastewater, which contain high levels of biodegradable materials.

A wide variety of substances have been reported as inhibitors causing anaerobic process failure, for example, sulfide, ammonia, light metal ions heavy metals and organics. High concentration of the substances could toxic the microbial community and lead to process failure. This project will focus mainly on the effects of high concentration of sulfate, which is a common constituent of many industrial wastewaters and also latex factories. In anaerobic reactors, the sulfate is reduced to sulfide by the sulfate reducing bacteria (SRB). The SRB is a competitor to other anaerobic microorganisms including methane producing bacteria (MPB) as they share common substrates. Also, the produced sulfide is toxic to the MPB, other anaerobic microorganisms, and SRB itself. These factors lead low biogas production, inefficient treatment and/or failure of the system. Furthermore, sulfide oxidizing bacteria (SOB) are also often found in the system. They oxidize dissolved hydrogen sulfide and other sulfur compounds.

2.3 Next-Generation Sequencing (NGS)

Sequencing technologies revolutionized the molecular biology and microbiology studies in a few decades and have become the most frequently used tool to cultured-independently investigate the potential microorganisms [9]. The technology helps to determine the order of nucleotide bases (A, T, C, G) of DNA or transcribed RNA sequences. The instruments and materials for sequencing have been developed to decrease the cost and accelerate the speed as a

large-scale sequencing. The NGS technology, which is a platform for high-throughput sequencing, has been widely used for genomics and transcriptomics studies. This sequencing technology has many advantages (e.g. highly-parallel sequencing, which can perform many reactions at the same time) and chip-scale process (which means all processes can be done on a tiny place called "sequencing chip"). The product of sequencing technology also known as "read" is short (approximately 50-700 bases), consequently the platform referred to as short-read sequencing platform [9].

Currently, NGS technologies comprise several platforms such as Illumina, Roche and LifeTechnologies. However, Illumina is well-known, widely and frequently used platform for short-read sequencing. The instruments of Illumina platform include NextSeq, MiSeq and HiSeq, etc. Their very performances are run time, the number of maximum output (Gb), resulting-read length and the number of reads in parallelized run. For example of Illumina's instrument, MiSeq could generate 132 to 15 Gb of 300x2 paired-end read while HiSeq2500 could generate in wide range throughput from 180 up to 500 Gb of the read length 50 to 125 bases paired-end read [10]. The application of each instrument based on the purpose of a study, for instance, target-specific sequencing, whole-genome sequencing of large genome (human, plat, animal model) or small genome (microbe, virus), etc. Moreover, recently technology could generate paired-end sequence including forward and reverse reads of each sequencing region performing more accurate than single-ended read sequences [9, 11].

2.4 Metagenomic data analysis

With the development of NGS technologies, high-throughput DNA sequencing has become economically affordable for studying living organisms including sequencing of environmental DNA samples that provides metagenomic data [12]. There are two main approaches of NGS-based methods to study microbiomes, which are amplicon-based (e.g. 16S rRNA) sequencing and shotgun sequencing. Both approaches provide several thousand or even millions of short DNA sequences as results, called reads. Lengths of the DNA sequences vary according to the technologies capacities. The first approach is a target sequencing method requiring specific primers to amplified specific target genes. 16S rRNAs is the most popular genes for target sequencing to study microbial diversity and abundance, as it is highly conserved sequences over evolutionary time. Another approach is whole-genome shotgun sequencing providing short reads of all DNA materials from the studied samples. This method requires DNA library construction using ramdom primers. Even through the whole-genome shotgun is more

expensive than the 16S rRNAs as more DNA fragments will be sequenced, the approach provides more information of functional genes in the studied communities. Since the whole-genome shotgun sequencing will be utilized in this project, an overview of the method will be described as below.

Two main bioinformatic approaches can be utilized for analyzing whole-genome shotgun metagenomic data [13]. The methods are depending on whether an assembly process will be performed. The assembly process is a process to extend short reads from sequencing technologies to longer sequences by recursively finding overlapped regions of the short reads. For the assembly-free method, the produced reads will be analyzed directly after the pre-processing process to reduce sequencing errors. Each read will be categorized in taxonomy binning process. Generally, similarity-based taxonomy assignment is utilized by comparing to references databases such as NCBI RefSeq [14], KEGG [15] and EggNOG [16]. Another type of methods for taxonomy binning considers sequence compositions, for example, GC content, codon usage or frequency of short oligomers (k-mers). Several computational frameworks have been developed to perform the taxonomy binning, for instance, MEGAS, PhymmBL, MetaPhlAn, and mOTU. Then, the taxonomic abundance and microbiome diversity can be measured.

For the assembly-based strategy, reads from the same genomes are merged into longer sequences called contigs. The functional diversity of the studied microbial communities can be quantified from functional annotations of these metagenomic sequences. Two main steps will be conducted for the gene annotation process, which are gene prediction and functional annotation. For the gene prediction step, the encoding genes will be identified from the assembled contigs. Then, functional annotation can be performed after deriving the identified genes. Several databases are available for inferring functions e.g. KEGG [15], COG [17], Pfam [18], EggNOG [16], MG-RAST [19], and CAMERA [20]. In addition, functional enrichment analysis of the pathways in a sample or between samples can be conducted. Comparing metabolic gene abundance between samples can also be studied.

Chapter 3
Materials and Methods

The development of bioinformatic pipeline for metagenomic data analysis, followed by analytical step of anaerobic microbiomes for better understanding from high sulfate wastewater systems were performed. The project was under collaboration with a team at Excellent Center of Waste Utilization and Management (ECoWaste), Pilot Plant Development and Training Institute (PDTI), King Mongkut's University of Technology Thonburi (KMUTT), Bangkhuntien, Bangkok, Thailand. The DNAs of the microbiomes were extracted from the ECoWaste center, and then outsourced for shotgun metagenome sequencing. The metagenomic data were analyzed in this project through the developed bioinformatic pipeline. The details of all materials and methods are described below.

3.1 Materials

3.1.1 Metagenomic samples of the studied latex wastewater treatment system

The microbiome samples were collected from the AD multi-stage treatment system aiming for better performance of sulfate removal and higher methane production. The system comprises (1) Acidogenic sulfate reducing reactor; SRB_R, (2) Sulfide oxidizing reactor; SOB_R, and (3) Methanogenic reactor; MT_R, which are connected orderly (Figure 3.1). The reactors carry three main functions of reducing sulfate, oxidizing sulfide and producing methane, respectively. In order to improve the treatment performance of reducing sulfate while still maximizing methane production, there will be a recirculation between SRB_R and SOB_R reactors. The metagenomic samples for shotgun sequencing were derived from SRB_R and SOB_R reactors at the control and recirculation rate of 70%, which is an optimal rate providing high sulfate removal and biogas production. The optimal performance was at organic loading rate (OLR) 5 kg COD/m^3 day, biogas production of 0.234 m^3/kg COD, and chemical oxygen demand (COD) removal at 83%. The anaerobic sludge was derived from AD wastewater system of a concentrated latex factory.

In total, four metagenomic samples were sequenced using Illumina MiSeq platform of 150 bases paired-end reads. The samples were from SRB_R and SOB_R reactors at control stage, and SRB_R and SOB_R reactors when performed recirculation rate at 70%.

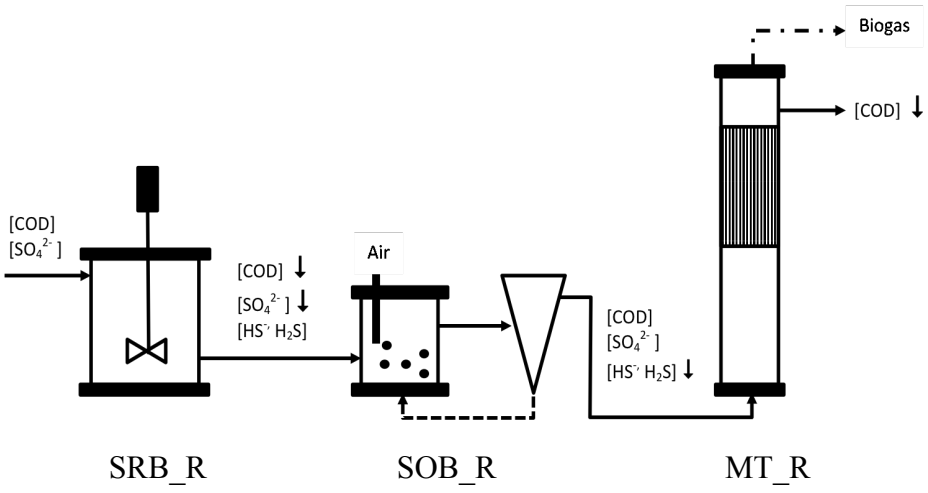


Figure 3.1 The design of AD multi-stage treatment system. The system comprises three reactors of acidogenic sulfate reducing reactor (SRB_R), Sulfide oxidizing reactor (SOB_R), and Methanogenic reactor (MT_R), respectively.

3.1.2 Software for metagenomic analysis

A list of tools or software was integrated to develop a pipeline for metagenomic data analysis (Table 3.1). These are open-source software which are freely available. Each tool has its own function of data analysis. In addition, a list of databases was collected for each step of analyses as shown in Table 3.2.

Table 3.1 A list of the software used for the developed metagenomic pipeline.

Method	Software/tools	References
Data pre-processing		
Quality checking/visualization	FastQC	[21]
Trimming out low-quality bases and adapter removal	Trimmomatic	[22]
Taxonomic analysis		
Taxonomic classification	Centrifuge	[23]
Taxonomic profile visualization	Krona, R, RAM package	[24], [25], [26]
Functional analysis		
Sequences assembly	IDBA-UD, BLAST	[27], [28]
Gene prediction	Prodigal	[29]
Functional assignment	eggNOG-Mapper	[30]
Reads mapping	Bowtie	[31]
Pathway visualization	KEGG Mapper	[32]
Others		
Parsing data files	Python, R	[33], [25]
Normalization	R	[25]
Statistic calculation	Python, R, Shell script	[33], [25]

Table 3.2 A list of the databases for metagenomic analysis.

Method	Database name	References
Taxonomic analysis	RefSeq	[14]
Functional analysis	EggNOG, KEGG	[16], [34]
Interpretation	MiDAS	[35]

3.2 Methods

In order to analyze metagenomic data of sulfate rich wastewater, the bioinformatic pipeline for analyzing metagenomic data was firstly developed, and then utilized in an analytical step of the anaerobic microbiomes.

3.2.1 Development a bioinformatic pipeline for analyzing metagenomic data of anaerobic wastewater treatment systems

Bioinformatic pipeline for analyzing metagenomic data was developed, focusing for microbiomes from AD wastewater systems. The pipeline development can be divided into six main steps (Figure 3.2), which are (1) pipeline design, (2) modules development, (3) modules integration, (4) database development, (5) pipeline testing, and (6) pipeline deployment.

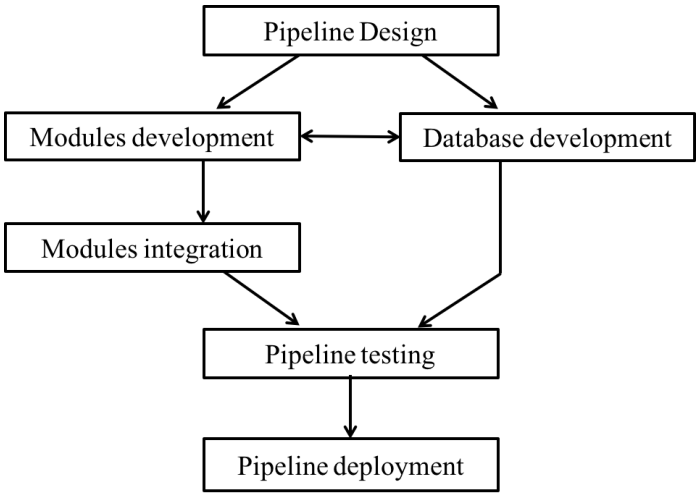


Figure 3.2 An overview of bioinformatic pipeline development for analyzing metagenomic data.

From the first step of pipeline development, the pipeline structure for metagenomic data analysis will be designed by dividing into analytical modules. Here the pipeline was designed for shotgun sequencing data (mainly for Illumina or short reads platform). The analysis of the metagenomic data can be separated into four main parts, which are data pre-processing, taxonomy analysis, functional analysis, and followed by comparative metagenomics. After developing each module, the tested modules were integrated into flowing pipeline. In parallel, the database for analyzing metagenomic data emphasizing on anaerobic microbial was constructed. After combining the integrated modules and the developed database, the pipeline was tested for deployment and used for next step analysis of latex wastewater samples.

3.2.1.1 Modules development and integration

For the development of each module (data pre-processing, taxonomic analysis, functional analysis, and comparative metagenomics), the existing bioinformatic tools were chosen based on their efficiency or accuracy of analyses, requirement of computational power, time consuming and ease of use (Table 3.1). The selection was based on literature review and in-house evaluation [36]. In addition, the synchronization of the tools in the pipeline was also considered. The selected tools/scripts/software were installed and finally tested. The flows of the pipeline were mainly implemented using Python, R programming and shell scripts.

3.2.1.2 Databases development

Each step of the metagenomic analysis requires external databases for analyzing the derived DNA sequences, mainly to identify the taxonomy or function of the studied samples (in taxonomic or functional analysis, respectively). The studied DNAs from metagenomic data can compared to find most similar sequences in publicly available databases for inferring taxonomy and functions (Table 3.2). As these databases have different data formats and were separately available from various sources, databases integration was then performed by downloading in the local storage (hard drive) and linking to the implemented module in order to facilitate the analysis. For taxonomic database, the database was required to be costumed and synchronized to the selected tool, Centrifuge [23].

3.2.2 Metagenomic analysis of anaerobic microbiomes in sulfate rich wastewater system

The developed bioinformatic pipeline from the previous step was utilized for the metagenomic data analysis of anaerobic microbiomes in sulfate rich wastewater system derived from ECoWaste center as described above. Metagenomic data at different conditions will be analyzed, which are from SRB_R and SOB_R reactors at control stage and optimal recirculation rate, respectively. The metagenomic data analysis was performed through the following steps (see an overall workflow in Figure 4.1).

1. Data pre-processing. The raw sequences data (FASTQ format) were evaluated and visualized using FastQC program [21]. Low-base-quality ($Q < 30$) and adapter were trimmed using Trimmomatic [22]. After trimming, sequences with length < 36 base pairs were removed. All sequences passing the filtering criteria were then used for downstream analyses.

2. Taxonomic analysis. The pre-processed read sequences were classified for their microbial taxonomy using Centrifuge [23] and the custom in-house database containing complete genomes of bacteria, archaea, fungi and virus derived from RefSeq database [14].

Relative abundance of the identified microbes were calculated and normalized using scaling method [37]. Microbes with >=1% relative abundance in at least one sample were displayed in taxonomic profiles.

3. Functional analysis. For each sample, the pre-processed short reads were assembled into contigs using IDBA-UD [27]. The in-house Python script was then used to merge contigs between samples (Overlapped sequences > 200 base pairs and identity of overlapped region is > 98% using standalone BLASTN [28]) making reference sequences of all samples. Based on the assembled sequences, gene prediction were performed using Prodigal software [29]. The functional genes were annotated using eggNOG-Mapper [30] and EggNOG database [16]. All gene abundances were normalized using Reads Per Kilobase of gene, per Million mapped reads (RPKM) method. Gene with RPKM value >=1 was used for further analyses.

4. Comparative metagenomics between samples of different reactors and different stages. Taxonomic and functional data of all samples were compared using the normalized abundance values. The comparisons were aimed to reveal dynamisms of microbial communities corresponding to the reactor types and stages. For functional results, a pair of samples were compared to find differential abundance genes and visualize in functional pathways using KEGG Mapper [32]. MiDAS database, Global Database of Microbes in Wastewater Treatment Systems and Anaerobic Digesters [35], was used as a reference and interpretation for microbes detected in wastewater treatment system.

Chapter 4
Results and Discussion

In this project, a bioinformatic pipeline for analyzing metagenomic data of anaerobic wastewater treatment systems was developed. The pipeline was based on the input of high-throughput short read technology, comprising four main analytic modules of data pre-processing, taxonomy analysis, functional analysis, and comparative metagenomics. The developed pipeline was then applied to analyze shotgun metagenomic data of microbial communities in latex wastewater treatment system, which contain high sulfate concentration. Microbial profiles and their functions including the dynamisms between reactor types (SRB and SOB reactors from AD multi-stage treatment system) and running stages (control and optimal conditions) were revealed. The results provide insight information of anaerobic microbes which could be further studied and applied for enhancing biogas production in sulfate-rich wastewater treatment systems.

4.1 Bioinformatic pipeline for analyzing metagenomic data of anaerobic wastewater treatment systems

The developed pipeline comprises four main analytic modules of data pre-processing, taxonomy analysis, functional analysis, and comparative metagenomics. The overall workflow is shown in Figure 4.1 and the tools or software used in the integrated pipeline were listed as mentioned in Table 3.1 (Section 3.1.2). Databases used for metagenomic analysis were also integrated into the pipeline (Table 3.2). The data pre-processing was firstly needed to be performed for quality control of derived sequences. This step helps to alleviate technical errors from sequencing or sampling processes. After the pre-processing step, the taxonomic and functional analyses can be performed. The taxonomic analysis will enable studies of diversity and abundance of the microbiomes of interest. The functional analysis provides discovery of functional genes of microbiomes and their metabolic abilities. Finally, the comparative metagenomics can be performed for comparing microbial communities and their functions between samples. This step reveals dynamic of microbiomes in different conditions. Details of each step with its utilized databases are elucidated below.

4.1.1 Data pre-processing module

A data pre-processing module is a crucial step for quality control of the derived raw sequences (FASTQ format) by evaluating technical errors from sequencing process. In this

pipeline, FastQC [21] software is employed. Low-quality bases and adapters from the sequencing process will be trimmed from the read sequences using Trimmomatic [22]. The cut-off base quality score is usually at quality score >= 20 (99% base call accuracy). However, quality score >=30 (99.9% base call accuracy) is recommended if the derived raw sequences contain high-quality bases. Higher stringency criteria provide better accuracy of downstream analyses of taxonomic classification and functional annotation.

4.1.2 Taxonomic analysis module

Taxonomic analysis reveals diversity and abundance of studied microbiome samples. The processed reads from the data pre-processing step will be used as an input of this step. The input reads will be classified and assigned the taxonomy using Centrifuge [23] by comparing with the taxonomic database. Tools for taxonomic classification were evaluated in term of detected microbial richness and diversity [36]. Centrifuge showed accurate numbers of richness and well-correlated diversity to the mock dataset. Even through the tool could give some numbers of detected artifact sequences, it facilitates better in term of time consuming and required executing memories. In this pipeline, a custom database containing complete genomes of bacteria, archaea, fungi and virus derived from RefSeq database [14] was constructed and formatted for using with the Centrifuge software. The selected genomes in the databases reflect organisms found in wastewater treatment systems. Relative abundance, a ratio of read counts of a particular taxonomy and total reads of a sample, can then be calculated after classification and normalized using scaling method [37]. The normalized values provide comparable abundance between samples. Krona [24] can be used for interactive display of a taxonomic profile.

4.1.3 Functional analysis module

Functional analysis provides a discovery of functional genes of microbiomes and their metabolic abilities. The module takes the input read as pre-processed reads from the data pre-processing step. As read inputs for this pipeline are from short read technology (approximately 100-300 base pairs), it is important to do the sequence assembly before finding functional genes. IDBA-UD [27], which was developed to the specific characteristic of metagenomic data, was selected for assembling reads of each sample. After deriving assembled reads or referred to as contigs, these contigs were then merged between samples in order to make reference contigs for further analyses. The merging process were based on BLASTN with overlapped region between two sequences > 100 base pairs and sequence identity > 98%. We did not perform polled reads assembly among all samples because it would lead more misassembled contigs between similar

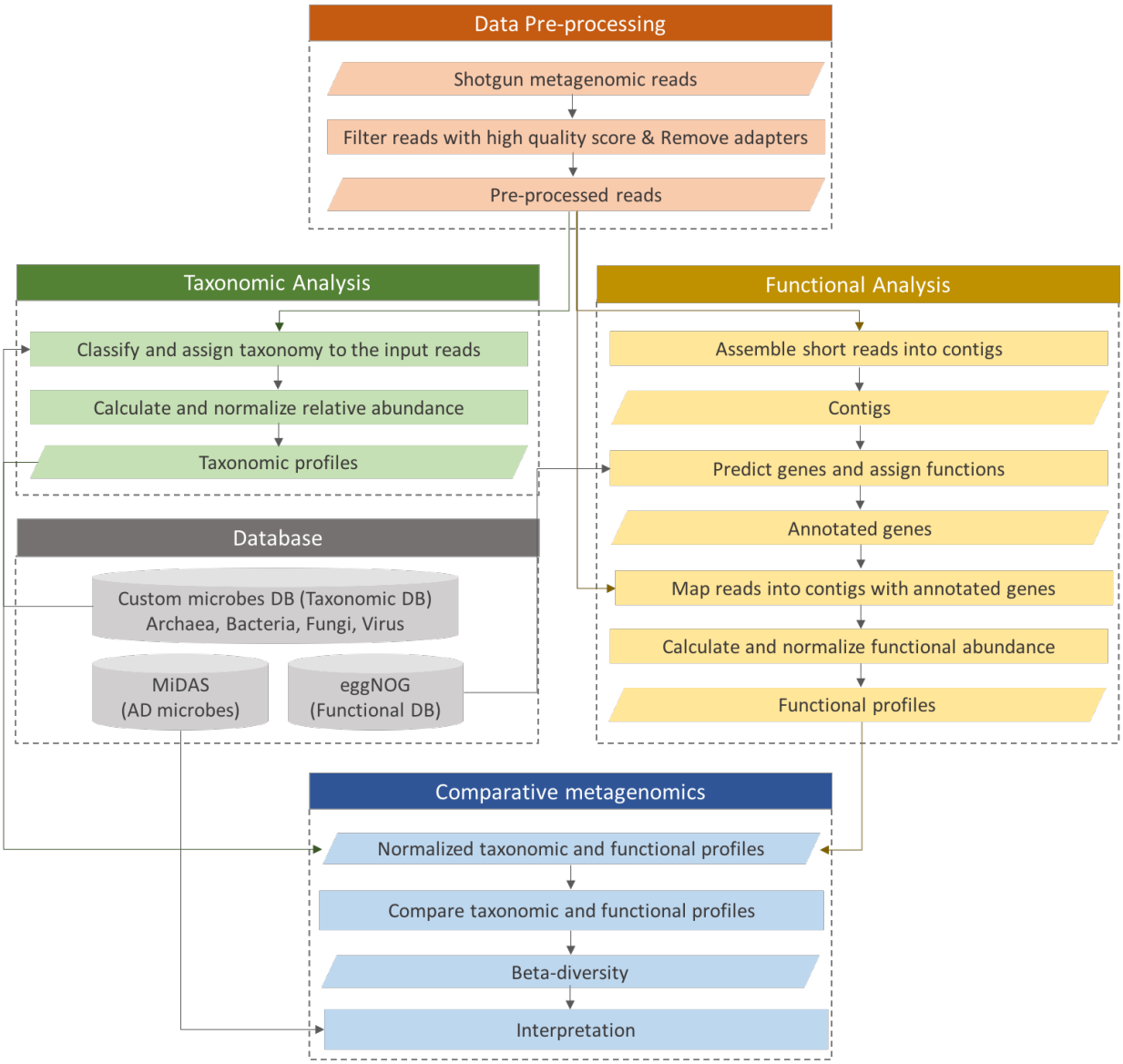


Figure 4.1 An overall of a pipeline for shotgun metagenomic analysis of microbiome from anaerobic wastewater treatment systems. The pipeline was developed as four main integrated modules of data pre-processing, taxonomy analysis, functional analysis, and comparative metagenomics. The databases for taxonomic profiling, functional analysis and interpretation were constructed. AD: Anaerobic digestion; DB: Database.

reads in different samples. The gene regions can then be detected. Here well-known gene prediction tool, Prodigal [29], was utilized. To calculate gene abundance, reads mapped to the predicted genes will be counted using Bowtie tool [31]. For functional annotations of the predicted genes, eggNOG-Mapper [30] was chosen with EggNOG database [16]. The database

contains functional categories including KEGG Orthology (KO), which can be visualized by KEGG Mapper [32]. RPKM method, Reads Per Kilobase of gene per Million mapped reads, was chosen for normalization of gene abundances as it also considers the gene length. Genes with bigger size have higher chance to get more fragments sequenced.

4.1.4 Comparative metagenomics module

Comparative metagenomics can be performed for comparing microbial communities between samples for both taxonomic diversity and functional profiles. The step lead to understanding of the dynamic of microbiomes in different conditions/samples. In order to compare the data between samples, normalized abundance values are needed. This is because the total reads of each sample are not the same. Comparing the abundance using read counts directly would be unfair e.g. compare the abundance between samples with low and high numbers of total reads. Differential abundance genes between two samples can be identified. KEGG Mapper [32] could be used to visualize genes present in a specific pathway. In addition, MiDAS database [35] can be used as a reference of microbe collections in wastewater treatment systems. This is a specific database attempting to collect microbes in wastewater treatment systems from different countries around the world. However, the database still incorporates data from a limited number of countries.

4.2 Microbial communities in the latex wastewater treatment system

The developed pipeline was applied to analyze the shotgun metagenomic data of microbial communities in latex wastewater treatment system containing high sulfate concentration. In this project, the samples were derived from AD multi-stage wastewater treatment system aiming to improve performance of sulfate removal and higher methane production. The system was designed as three consecutive reactors of acidogenic sulfate reducing reactor (SRB_R), Sulfide oxidizing reactor (SOB_R), and Methanogenic reactor (MT_R), respectively (More details in section 3.1.1). The samples were from SRB_R and SOB_R reactors at control (referred to as **SRB_C**, **SOB_C**) and optimal recirculation rate (70%) between the two reactors (referred to as **SRB70**, **SOB70**), respectively. The results of the metagenomic data pre-processing, taxonomic analysis and functional analysis including the community comparisons between samples are shown and discussed below.

4.2.1 Raw and pre-processed metagenomic data

The metagenomic raw data were retrieved from Illumina short read sequencing platform of original paired-end sequence length of 150 base pairs. After quality control or data pre-processing step, about 10% of original sequences were removed. The rest of the sequences were then utilized for downstream analyses. Statistic of the analyzing sequences of the studied samples were shown in Table 4.1

Table 4.1 Statistic of metagenomic sequences of the studied samples

Sample	Number of raw sequences	Number of pre-processed sequences	Sequence length	% GC content
SRB_C	23,419,540	21,648,077 (92.44%)	36-150	49
SRB70	20,554,231	18,416,920 (89.60%)	36-150	45
SOB_C	23,339,401	21,063,247 (90.25%)	36-150	48
SOB70	23,323,966	21,085,072 (90.40%)	36-150	46

4.2.2 Taxonomic profiles of the microbial communities in the latex wastewater treatment system

4.2.2.1 A large number of unknown microbes detected

The high-quality reads were then classified for taxonomy to identify microbial profiles of the samples. The numbers of taxonomic assignable sequences are shown in Table 4.2 varying between 18.29-28.27% of the total pre-processed sequences. The results show high numbers of unknown organisms in the samples. The situation is often found in environmental samples, showing a need for more complete database to provide better information of the analytic environment. In the last decade, high-throughput sequencing technologies have been utilized to study microbial community allowing detection of microbes without cultivation. A large number of organisms are detected though the technologies. This propagates more information to study microbial communities.

4.2.2.2 Microbial community profiles of microbes from the sulfate-rich wastewater treatment system

Based on the taxonomic assignable sequences, microbial profiles of the samples from latex wastewater treatment system (AD multi-stage) with high concentration of sulfate were drawn. The microbial profiles at different taxonomic levels are displayed in Figure 4.2-4.3.

Three domains of microorganisms were detected which are bacteria, archaea and virus (Figure 4.2A). Bacteria is the most abundance organism (88.47 – 94.95%) in this experiment followed by archaea and virus (<1% relative abundance). At a phylum level (Figure 4.2B), Proteobacteria (30.01 – 42.98%), Firmicutes (14.80 – 46.08%), Bacteroidetes (7.09 – 14.03%) and Actinobacteria (7.25 – 8.72%) were found as the top four abundance bacteria having >1% relative abundance in all samples, respectively. Euryarchaeota was found as the most abundance archaea (5.01 – 11.68%), which is a phylum of methanogens. The revealed proportions of archaea are corresponding to the reactor performances. At the optimal recirculation rate of 70% between two reactors of SRB_R and SOB_R, the reactor performances of methane production were higher than at control stage (no recirculation between the reactors). In addition, the archaea proportions were higher in SRB_R than SOB_R reactors at the same stage. This is due to high concentration of sulfate and resulting sulfide of the digestion which are toxic to methanogens.

Several SRB and SOB were detected in the studied system, which corresponds to the condition of latex wastewater containing high concentration of sulfate. Figure 4.3 shows microbial profiles at family and species levels, respectively. For example, *Desulfovibrionaceae*, *Desulfomicrobiaceae* and *Desulfobacteraceae* were found as dominant SRB at a family level. *Desulfovibrionaceae* were found with higher relative abundance at control stage (4.24% and 3.29% in SRB_C and SOB_C, respectively) than optimal recirculation stage (2.91% and 1.96% in SRB70 and SOB70, respectively). On the other hands, *Desulfomicrobiaceae* (0.42%, 0.65%, 0.44% and 1.04% in SRB_C, SRB70, SOB_C and SOB70, respectively) and *Desulfobacteraceae* (0.24%, 2.36%, 0.43% and 1.01% in SRB_C, SRB70, SOB_C and SOB70, respectively) showed higher abundance at the optimal recirculation rate stage. In addition, the relative abundances were slightly higher in SOB_R than SRB_R reactors, which are opposite from *Desulfovibrionaceae*.

Diverse families of SOB were revealed, for instance, *Spirochaetaceae*, *Rhodobacteraceae*, *Campylobacteraceae*, *Comamonadaceae* and *Burkholderiaceae*. These microbes were found with relatively abundance compared to SRB and methanogens.

Table 4.2 Numbers of taxonomic assignable sequences of the studied samples.

Sample	Total mapped sequences (Percent from total)	Number of assignable sequences at each taxonomic level						
		Domain	Phylum	Class	Order	Family	Genus	Species
SRB_C	6,119,494 (28.27%)	5,945,362	5,783,182	5,630,388	5,590,430	5,541,880	5,499,317	5,149,475
SRB70	3,267,881 (17.74%)	3,099,606	2,966,066	2,847,267	2,831,923	2,795,582	2,748,313	2,561,354
SOB_C	4,539,445 (21.60%)	4,326,255	4,157,784	3,977,561	3,935,180	3,886,434	3,851,109	3,578,262
SOB70	3,857,484 (18.29%)	3,649,342	3,484,982	3,325,620	3,294,018	3,246,529	3,182,599	2,945,186

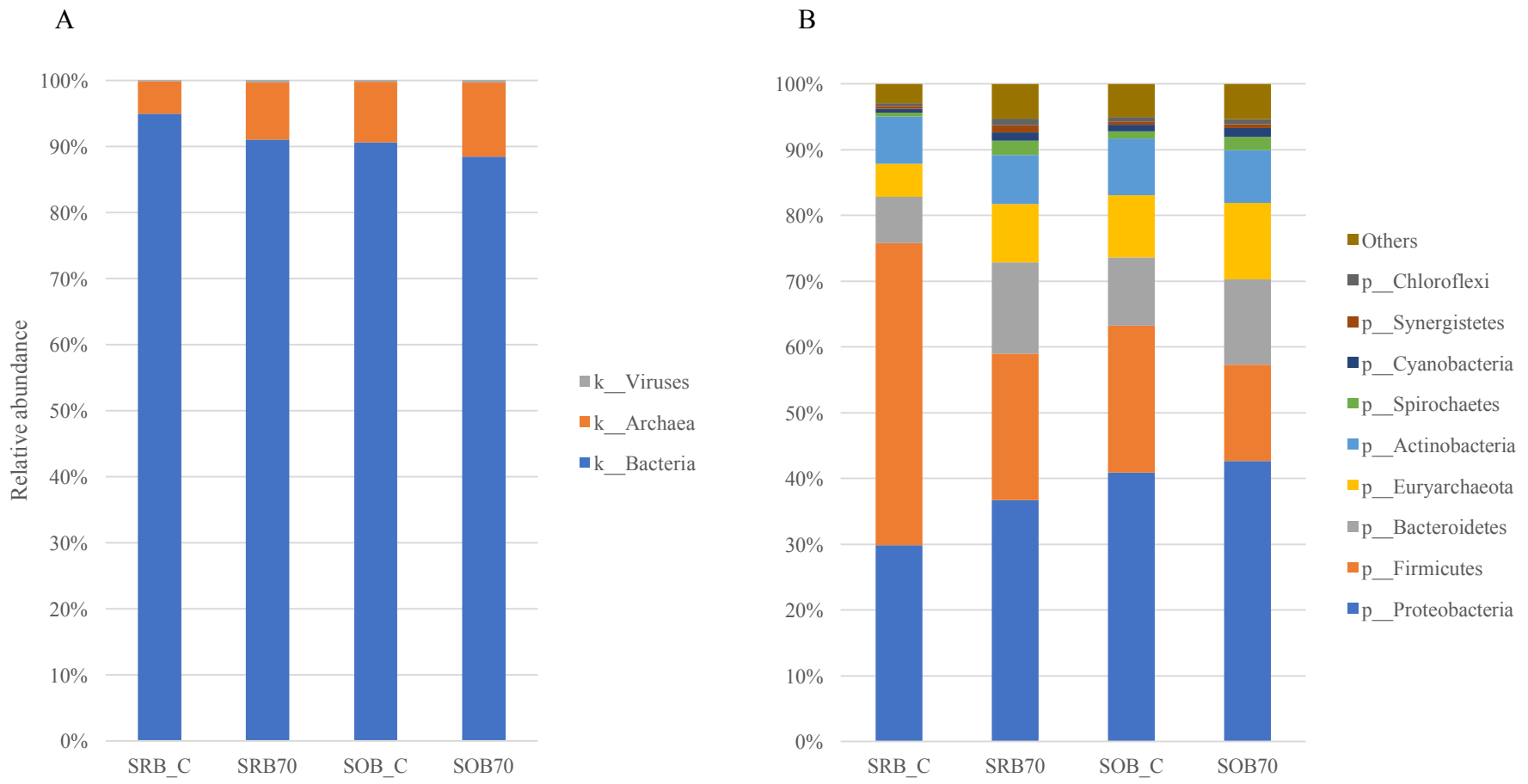


Figure 4.2 Microbial community profiles showing diversities of microbes from latex wastewater treatment system at (**A**) domain and (**B**) phylum levels, respectively. At phylum level, microbes with >=1% relative abundance are shown. The samples are from sulfate reducing (SRB) and Sulfide oxidizing (SOB) reactors at control (C) and recirculation rate of 70% (70) stages.

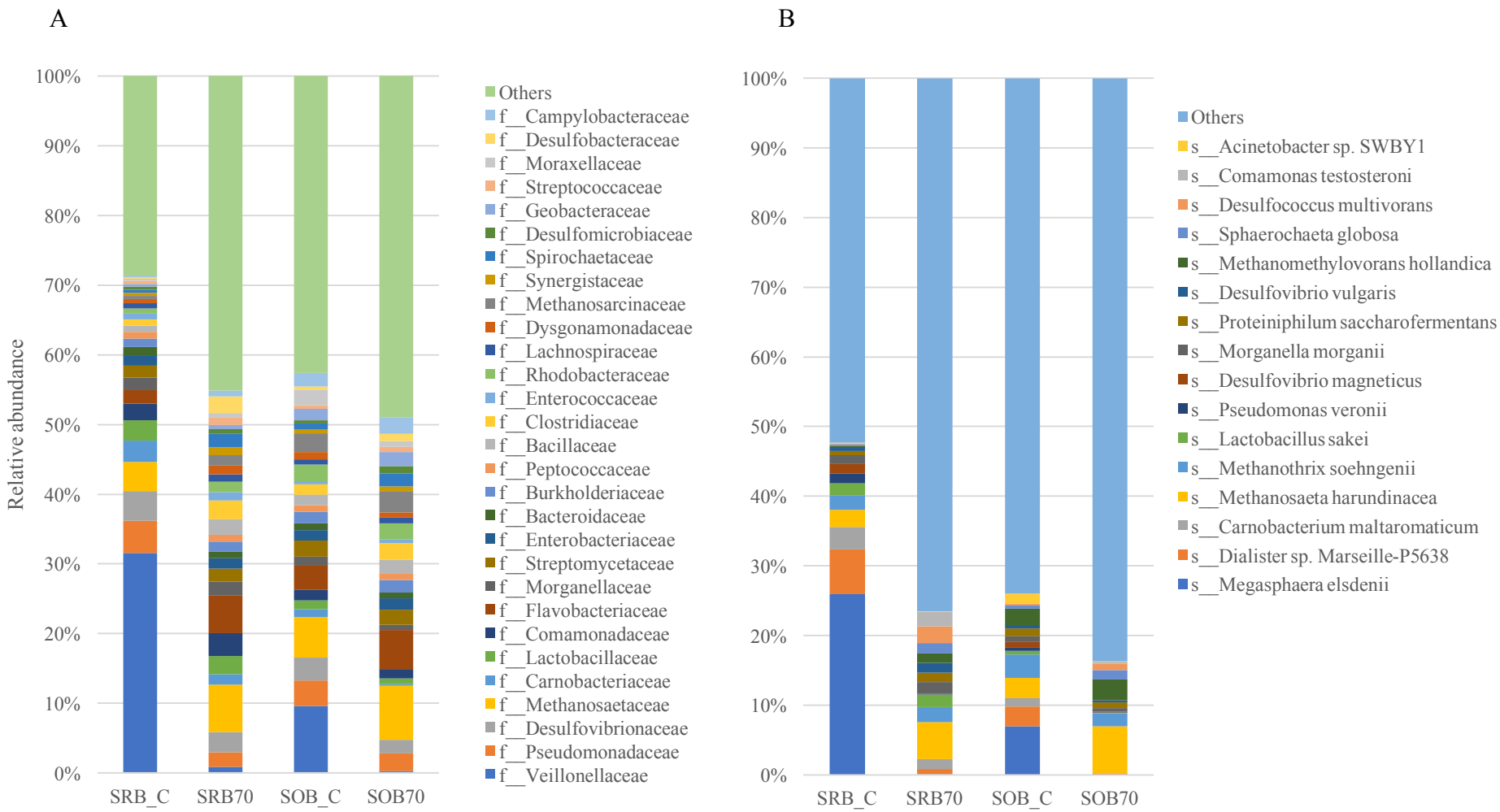


Figure 4.3 Microbial community profiles showing diversities of microbes (>=1% relative abundance) from latex wastewater treatment system at (**A**) family and species (**B**) levels, respectively. The samples are from sulfate reducing (SRB) and Sulfide oxidizing (SOB) reactors at control (C) and recirculation rate of 70% (70) stages.

Methanosaetaceae (4.20 – 7.76%) and *Methanosarcinaceae* (0.48 – 3.09%) were found as prevalent families of methanogens. These are acetoclastic methanogens utilizing acetate as substrate to produce methane. The resulting abundances show consistency to the results at the phylum level (Figure 4.2B) that are higher in the optimal recirculation stage than no circulation stage and higher in SOB_R reactor than SRB_R reactor.

At a species level, *Megasphaera elsdenii* was revealed as the most prevalent bacteria and highly present at the control stages (25.95% and 6.94% in SRB_R and SOB_R reactors, respectively). It has been reported as lactate degrader to propionate, butyrate and acetate [REF], which could be substrates for SRB and methanogens. *Methanosaeta harundinacea* was the most dominant methanogen in the treatment system. Interestingly, *Desulfovibrio vulgaris* and *Desulfococcus multivorans* were found as dominant SRB. These two microbes were found with higher relative abundance at the optimal condition of wastewater treatment. The discovery suggests SRB that can utilize sulfate in wastewater and could work together with methanogens. *D. vulgaris* is a group of incompletely-oxidizing sulfate reducer providing acetate as a product [38], which could be a substrate of methanogens. Further study of these bacteria could be a recommendation.

4.2.2.3 Dissimilarity of microbial profiles between different reactor types and performance stages

Figure 4.4 shows dissimilarity of microbial profiles between different reactor types (SRB: Sulfate reducing reactor; SOB: Sulfide oxidizing reactor) and performance stages (C:Control; 70: Optimal recirculation rate at 70%). The microbiome profiles at the optimal recirculation rate (SRB70 and SOB70) show the most similar profiles than others. This could be because of the recirculation between the two reactors, which make the microbes mixed together. In addition, the profiles at different stages reflect different performance of the treatment system. The microbial profiles of the optimal condition could be preferred than the control stage. The sample of SRB_R reactor at control stage provided the most distinct profile compared to others. This would be mainly because of the *Megasphaera elsdenii* which has very high abundance than other species (25.95% in SRB_C). Nevertheless, this bacterium was dramatically decreased in the SOB_R reactor at the same condition (6.94%), and found with low abundance at the optimal stage (0.10% and 0.06% in SRB70 and SOB70, respectively). It could be possible that *M. elsdenii* degrades some products which could be substrates for SRB that completing with methanogens. Further study would give more insight information.

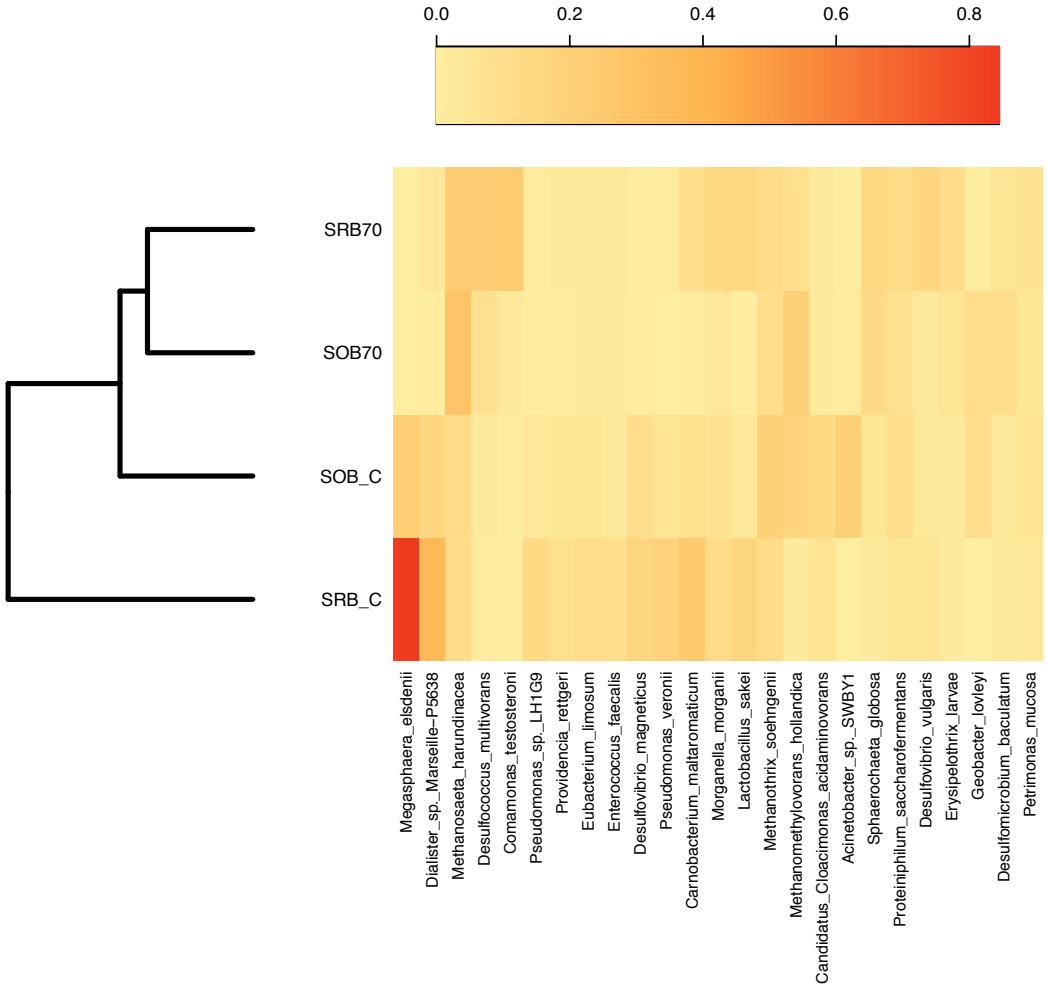


Figure 4.4 A heatmap showing dissimilarity of microbial profiles between different reactor types (SRB: Sulfate reducing reactor; SOB: Sulfide oxidizing reactor) and performance stages (C:Control; 70: Optimal recirculation rate at 70%). Top 25 abundance microbes at a species level are shown.

4.2.3 Functional profiles of microbial communities in the latex wastewater treatment system

4.2.3.1 Numbers of functional genes predicted in the latex wastewater treatment system

For functional analysis, gene predictions were firstly performed based on assembled reference contigs of all samples. The numbers of predicted open reading frames or genes in each sample are shown in Figure 4.5. 305665 genes are found in all studied samples. Some genes are uniquely found in each sample with the numbers of 12320, 15156, 14770 and 14666 in SRB_C, SRB70, SOB_C and SOB_70, respectively.

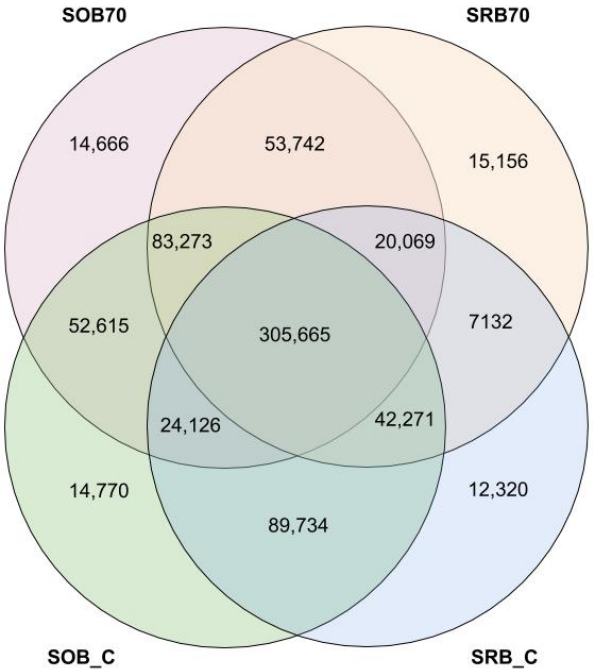


Figure 4.5 A Venn diagram showing numbers of functional genes predicted from microbial communities in the latex wastewater treatment system.

4.2.3.2 Functional genes of microbiome in the latex wastewater treatment system

All detected genes were then annotated for functions. Only genes having RPKM >1 were considered. Genes in the same functional pathway were grouped and displayed in Figure 4.6 and Table 4.3. In this experiment, abundance of functional genes of the microbiome in each sample are slightly different. Pathways of metabolite and short-chain fatty acid degradation and are among the top 20 pathways as well as methane metabolism. These pathways are important in AD process and biogas production.

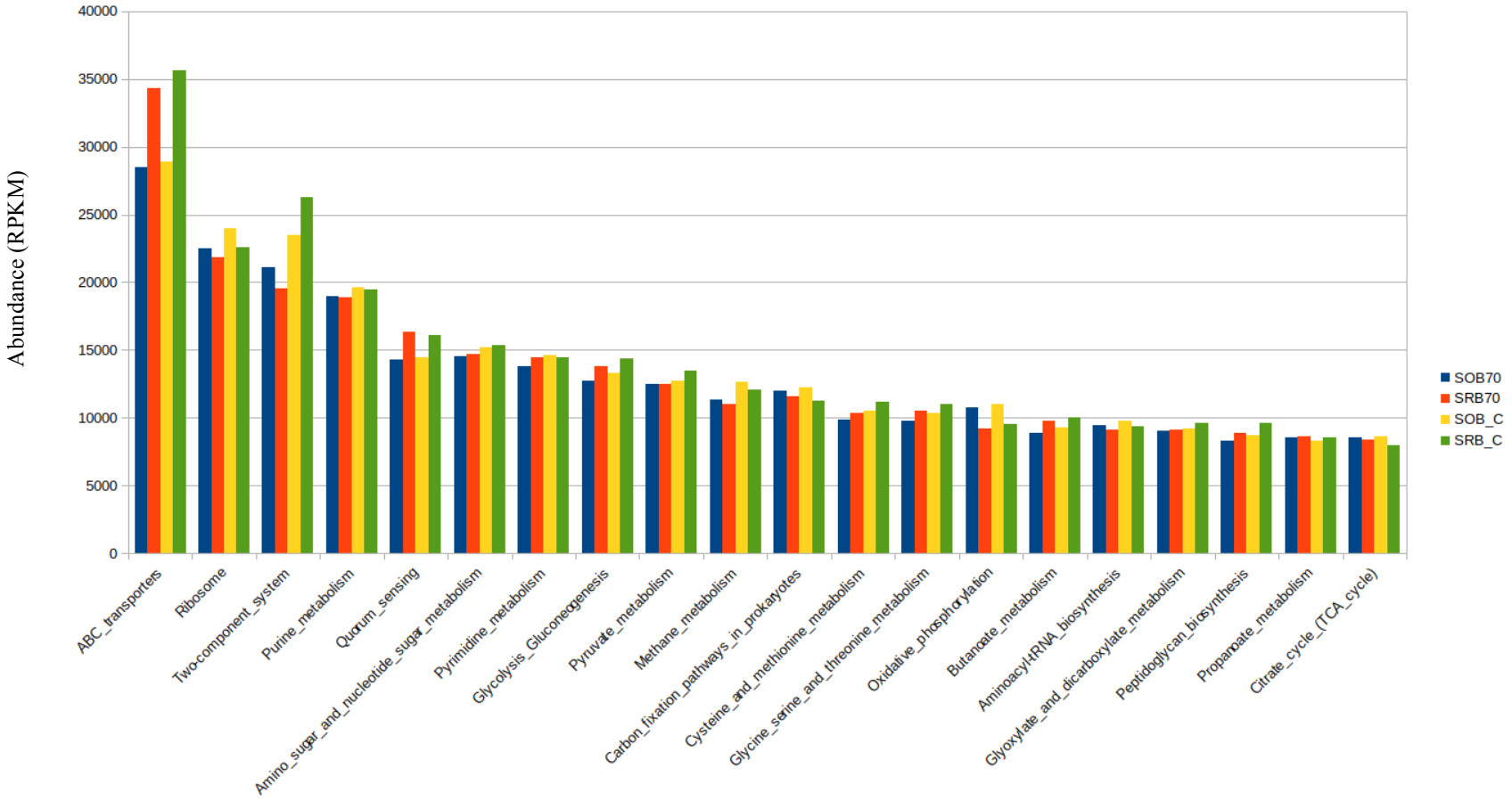


Figure 4.6 Top 20 abundance functional pathways found in the microbiome from the latex wastewater treatment system.

Table 4.3 Top 20 abundance functional pathways of microbiome from the latex wastewater treatment system. Sum RPKM of genes in each detected pathway are shown as abundance.

Functional pathway	**Sum abundance (RPKM)**			
	SRB_C	**SRB70**	**SOB_C**	**SOB70**
ABC_transporters	35624	34340.7	28890.8	28488.3
Ribosome	22557.9	21874.3	23947.2	22538.2
Two-component_system	26262	19582.8	23502.7	21140
Purine_metabolism	19479.3	18915.3	19644.7	18955.7
Quorum_sensing	16110.4	16334.3	14439.6	14276.1
Amino_sugar_and_nucleotide_sugar_metabolism	15367.3	14669.7	15199.1	14511.2
Pyrimidine_metabolism	14463	14466.6	14623.9	13836.3
Glycolysis_Gluconeogenesis	14335.2	13811.3	13346.4	12698.4
Pyruvate_metabolism	13491.4	12489.2	12752.3	12474.2
Methane_metabolism	12093.1	11028.5	12680.1	11306.7
Carbon_fixation_pathways_in_prokaryotes	11288.9	11562.1	12273.3	11971.9
Cysteine_and_methionine_metabolism	11157.8	10314.8	10517.2	9865.45
Glycine_serine_and_threonine_metabolism	11009.7	10521.4	10378.5	9782.25
Oxidative_phosphorylation	9499.01	9218.56	11033.1	10782.1
Butanoate_metabolism	10059.3	9768.28	9273.37	8879.25
Aminoacyl-tRNA_biosynthesis	9349.05	9118.35	9781.63	9488.34
Glyoxylate_and_dicarboxylate_metabolism	9613.77	9124.54	9242.96	9067.73
Peptidoglycan_biosynthesis	9616.94	8862.03	8689.5	8289.89
Propanoate_metabolism	8559.09	8613.5	8278.47	8511.09
Citrate_cycle_(TCA_cycle)	7995.43	8373.18	8628.18	8563.41

Differential abundance genes between each pair of samples were calculated in order to compare differences of microbial functions in each reactor. For example, differential abundance genes in methane metabolism pathway between SRB_C and SRB70 were displayed in Figure 4.7. Most of the genes discovered in this pathway are found with higher abundance in SRB70 (represented in blue color in Figure 4.6) which is corresponding to the better performance of the SRB70 sample of methane production.

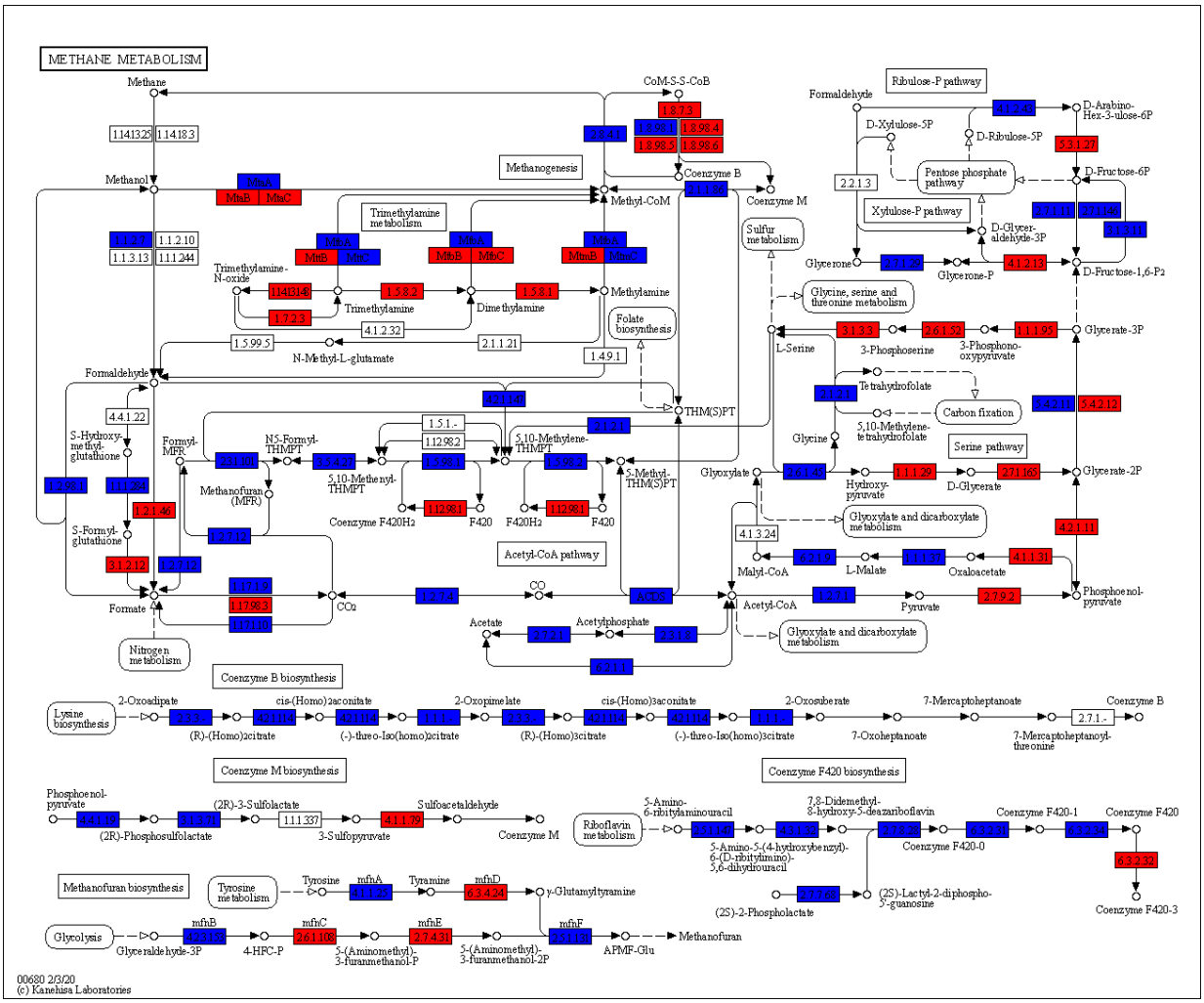


Figure 4.7 Differential abundance genes in methane metabolism pathway between SRB_C and SRB70 samples. Genes with higher abundance in SRB70 were labeled in blue. On the other hands, genes labeled in red have higher abundance in SRB_C sample. The white box is gene that cannot be found in any sample.

Chapter 5
Conclusion and suggestions

5.1 Conclusion

This project conducted the study of microbial communities in latex wastewaster treatment system through the technique called metagenomomics. The study yielded high-resolution of the microbes in the system without cultivation, which is a key information to improve the AD treatment system and enhance methane production. Firstly, in order to perform the metagenomic analysis, the bioinformatic pipeline was developed providing informative and accurate of microbial communities for both profiles (richness and abundance), and functions (what microbes can do). The pipeline is not limited to only project, but could also be applied to other studies of metagenomic data. This would be a practical guideline for biologists or other scientists to analyze metagenomic data. Secondly, the developed pipeline was applied to study the microbiome in a multi-stage treatment system of latex wastewater. Here we reported several of SRB in the system, often found in sulfate-rich wastewater, such as *Desulfovibrionaceae*, *Desulfomicrobiaceae* and *Desulfobacteraceae*. A list of SOB were also found such as *Spirochaetaceae*, *Rhodobacteraceae*, *Campylobacteraceae*, *Comamonadaceae* and *Burkholderiaceae*. *Methanosaetaceae Methanosarcinaceae* were found as prevalent acetoclastic methanogens utilizing acetate to methane. Interestingly, one of the discovered SRB, *Desulfovibrio vulgaris*, was found with higher abundance in the optimal performance reactor. It is a group of incompletely-oxidizing sulfate reducer providing acetate as a product, which could be a substrate of methanogens. The communities could be further studied and utilized for microbial management to increase the stability and efficiency of the anaerobic sulfate-rich wastewater treatment.

5.2 Suggestions for further studies

1. Metatranscriptomics could be performed to reveal functional expressions in the studied system. In this study, metagenomic analysis was performed. The technique allows to study microbial community profiles and functions through DNA materials. This provides the information of capacities of possible functions in the system. However, active genes could be different and revealed through the study of RNA, called metagranscripotimcs. The study could reflect functions that truly active in the system.

2. Further studies of some discovered microoganisms and their interactions could be performed to provide deeper information to improve the wastewater treatment systems. A group of dominant microbes in the high-performance system would be as of interest leading the microbiome engineering or manipulation for enhancing methane production.

References

1. Abdelgadir, A., et al., *Characteristics, process parameters, and inner components of anaerobic bioreactors.* Biomed Res Int, 2014. **2014**: p. 841573.
2. Anderson, K., P. Sallis, and S. Uyanik, *Anaerobic treatment processes.* Handbook of water and wastewater microbiology, 2003: p. 391-426.
3. Wirth, R., et al., *Characterization of a biogas-producing microbial community by short-read next generation DNA sequencing.* Biotechnology for biofuels, 2012. **5**(1): p. 41.
4. Boone, D.R. and M.P. Bryant, *Propionate-degrading bacterium, Syntrophobacter wolinii sp. nov. gen. nov., from methanogenic ecosystems.* Appl. Environ. Microbiol., 1980. **40**(3): p. 626-632.
5. McInerney, M.J., et al., *Syntrophomonas wolfei gen. nov. sp. nov., an Anaerobic, Syntrophic, Fatty Acid-Oxidizing Bacterium.* Appl Environ Microbiol, 1981. **41**(4): p. 1029-39.
6. Diekert, G. and G. Wohlfarth, *Metabolism of homoacetogens.* Antonie van Leeuwenhoek, 1994. **66**(1-3): p. 209-221.
7. Balch, W.E., et al., *Acetobacterium, a new genus of hydrogen-oxidizing, carbon dioxide-reducing, anaerobic bacteria.* International Journal of Systematic and Evolutionary Microbiology, 1977. **27**(4): p. 355-361.
8. Koch, C., et al., *Microbiomes in bioenergy production: from analysis to management.* Curr Opin Biotechnol, 2014. **27**: p. 65-72.
9. Buermans, H. and J. Den Dunnen, *Next generation sequencing technology: advances and applications.* Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 2014. **1842**(10): p. 1932-1941.
10. Goodwin, S., J.D. McPherson, and W.R. McCombie, *Coming of age: ten years of next-generation sequencing technologies.* Nature Reviews Genetics, 2016. **17**(6): p. 333.
11. Glenn, T.C., *Field guide to next-generation DNA sequencers.* Molecular ecology resources, 2011. **11**(5): p. 759-769.
12. Metzker, M.L., *Sequencing technologies - the next generation.* Nat Rev Genet, 2010. **11**(1): p. 31-46.
13. Ji, B. and J. Nielsen, *New insight into the gut microbiome through metagenomics.* Advances in Genomics and Genetics, 2015. **6**: p. 77-91.
14. Haft, D.H., et al., *RefSeq: an update on prokaryotic genome annotation and curation.* Nucleic Acids Res, 2018. **46**(D1): p. D851-D860.
15. Kanehisa, M., et al., *Data, information, knowledge and principle: back to metabolism in KEGG.* Nucleic Acids Res, 2014. **42**(Database issue): p. D199-205.
16. Huerta-Cepas, J., et al., *eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses.* Nucleic Acids Res, 2019. **47**(D1): p. D309-D314.
17. Tatusov, R.L., et al., *The COG database: a tool for genome-scale analysis of protein functions and evolution.* Nucleic Acids Res, 2000. **28**(1): p. 33-6.
18. Finn, R.D., et al., *The Pfam protein families database: towards a more sustainable future.* Nucleic Acids Res, 2015.
19. Wilke, A., et al., *The MG-RAST metagenomics database and portal in 2015.* Nucleic Acids Res, 2015.
20. Sun, S., et al., *Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource.* Nucleic Acids Res, 2011. **39**(Database issue): p. D546-51.

21. *FastQC*. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
22. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-20.
23. Kim, D., et al., *Centrifuge: rapid and sensitive classification of metagenomic sequences.* Genome Res, 2016. **26**(12): p. 1721-1729.
24. Ondov, B.D., N.H. Bergman, and A.M. Phillippy, *Interactive metagenomic visualization in a Web browser.* BMC Bioinformatics, 2011. **12**: p. 385.
25. *The R Project for Statistical Computing*. Available from: https://www.r-project.org/.
26. Chen, W. *R for Amplicon-Sequencing-Based Microbial-Ecology*. Available from: https://www.rdocumentation.org/packages/RAM.
27. Peng, Y., et al., *IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth.* Bioinformatics, 2012. **28**(11): p. 1420-8.
28. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
29. Hyatt, D., et al., *Prodigal: prokaryotic gene recognition and translation initiation site identification.* BMC Bioinformatics, 2010. **11**: p. 119.
30. Huerta-Cepas, J., et al., *Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper.* Mol Biol Evol, 2017. **34**(8): p. 2115-2122.
31. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
32. Kanehisa, M. and Y. Sato, *KEGG Mapper for inferring cellular functions from protein sequences.* Protein Sci, 2020. **29**(1): p. 28-35.
33. *Python*. Available from: https://www.python.org/.
34. Ogata, H., et al., *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic Acids Res, 1999. **27**(1): p. 29-34.
35. Nierychlo, M., Andersen, K.S., Xu, Y., Green, N., Albertsen, M., Dueholm, M.S., Nielsen, P.H. , *Species-level microbiome composition of activated sludge - introducing MiDAS 3 ecosystem-specific reference database and taxonomy.* BioRxiv, 2019.
36. Uthaipaisonwong P., S.T., Laeman K., Cheevadhanarak S., Kusonmano K. *Evaluation of shotgun metagenomic classification performance via microbial richness and diversity*. in *The 9th International Conference on Computational Systems-Biology and Bioinformatics (CSBio 2018)*. 2018. Bangkok, Thailand.
37. Schloss, P.D., *Reintroducing mothur: 10 Years Later.* Appl Environ Microbiol, 2020. **86**(2).
38. Heidelberg, J.F., et al., *The genome sequence of the anaerobic, sulfate-reducing bacterium Desulfovibrio vulgaris Hildenborough.* Nat Biotechnol, 2004. **22**(5): p. 554-9.

Outputs from the research fund (MRG6080183)

Scientific knowledge

- Bioinformatic pipeline for metagenomic data analysis of bacterial communities in anaerobic wastewater treatment system
- Microbial community and functional profiles of microorganisms in sulfate-rich wastewater treatment system

International paper

Puengrang P, Suraraks B, Prommeenate P, Boonapatcharoen N, Cheevadhanarak S, Tanticharoen M, and Kusonmano K*, 2020, "Diverse Microbial Community Profiles of Propionate-Degrading Cultures Derived from Different Sludge Sources of Anaerobic Wastewater Treatment Plants", **Microorganisms**, Vol. 8, No. 2, p. 277., doi: 10.3390/microorganisms8020277 (impact factor 4.167)

International conference proceedings and presentations

Uthaipaisonwong P, Srisuk T, Laeman K, Cheevadhanarak S, Kusonmano K*. Evaluation of shotgun metagenomic classification performance via microbial richness and diversity. The 9th International Conference on Computational Systems-Biology and Bioinformatics (CSBio 2018), December 2018, Bangkok, Thailand. (Oral presentation)

Rangsiwutisak C, Cheevadhanarak S, Kusonmano K*. Hypervariable regions assessment for 16S rRNA gene-based metagenomic analysis in activated sludge and anaerobic digestion. The 30th Annual Meeting of the Thai Society for Biotechnology and International Conference, November 2018, Bangkok, Thailand. (Proceedings and oral presentation)

Laeman K, Uthaipaisonwong P, Rangsiwutisak C, Cheevadhanarak S, Kusonmano K*. SBI Metagenomics: An integrated database and visualization platform for metagenomic studies. The 30th Annual Meeting of the Thai Society for Biotechnology and International Conference, November 2018, Bangkok, Thailand. (Poster presentation)

*Corresponding author

Appendix

International paper, conference proceedings and presentations

microorganisms

MDPI

Article

Diverse Microbial Community Profiles of Propionate-Degrading Cultures Derived from Different Sludge Sources of Anaerobic Wastewater Treatment Plants

Pantakan Puengrang [1], Benjaphon Suraraksa [2,3], Peerada Prommeenate [3], Nimaradee Boonapatcharoen [2], Supapon Cheevadhanarak [4,5], Morakot Tanticharoen [6] and Kanthida Kusonmano [4,7,*]

1 Bioinformatics and Systems Biology Program, School of Bioresources and Technology, and School of Information Technology, King Mongkut's University of Technology Thonburi, Bangkok 10150, Thailand; joejo.pk.p@gmail.com
2 Excellent Center of Waste Utilization and Management, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bangkok 10150, Thailand; benjaphon@biotec.or.th (B.S.); nimaradee.boo@kmutt.ac.th (N.B.)
3 Biochemical Engineering and Systems Biology Research Group, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency at King Mongkut's University of Technology Thonburi, Bangkok 10150, Thailand; peerada.pro@biotec.or.th
4 Systems Biology and Bioinformatics Research Group, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bangkok 10150, Thailand; supapon.che@mail.kmutt.ac.th
5 School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bangkok 10150, Thailand
6 National Science and Technology Development Agency, Pathumthani 12120, Thailand; morakot.tan@nstda.or.th
7 Bioinformatics and Systems Biology Program, School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bangkok 10150, Thailand
* Correspondence: kanthida.kus@mail.kmutt.ac.th; Tel.: +66-2470-7407

Received: 31 December 2019; Accepted: 16 February 2020; Published: 18 February 2020

Abstract: Anaerobic digestion (AD) has been used for wastewater treatment and production of renewable energy or biogas. Propionate accumulation is one of the important problems leading to an unstable system and low methane production. Revealing propionate-degrading microbiome is necessary to gain a better knowledge for alleviation of the problem. Herein, we systematically investigated the propionate-degrading cultures enriched from various anaerobic sludge sources of agro-industrial wastewater treatment plants using 16S rRNA gene sequencing. Different microbial profiles were shown even though the methanogenic activities of all cultures were similar. Interestingly, non-classical propionate-degrading key players *Smithella*, *Syntrophomonas*, and *Methanosaeta* were observed as common prevalent taxa in our enriched cultures. Moreover, different hydrogenotrophic methanogens were found specifically to the different sludge sources. The enriched culture of high salinity sludge showed a distinct microbial profile compared to the others, containing mainly *Thermovirga*, *Anaerolinaceae*, *Methanosaeta*, *Syntrophobactor*, and *Methanospirillum*. Our microbiome analysis revealed different propionate-degrading community profiles via mainly the *Smithella* pathway and offers inside information for microbiome manipulation in AD systems to increase biogas production corresponding to their specific microbial communities.

Keywords: 16S rRNA gene-based sequencing; agro-industrial sludge; anaerobic digestion; microbiome; propionate-degrading cultures

1. Introduction

Biogas is an alternative fuel that can be produced by wastewater treatment under the absence of oxygen, called anaerobic digestion (AD). This process consists of various complex organic degrading sub-processes which are driven by microbial communities [1,2]. Even though the AD system has been considered as a promising solution for wastewater treatment and biogas production, the operational stability in several systems is still poor and yields low biogas production. Various factors have been reported as AD inhibitors causing system instability, such as volatile fatty acids (VFAs), long-chain fatty acids (LCFAs), toxic chemical substances, etc. [3,4]. Many studies have been set up to determine optimal process parameters for gaining high biogas production [5–8].

The anaerobic digestion process entails four steps: hydrolysis, acidogenesis, acetogenesis, and methanogenesis [9]. During hydrolysis, lipids, proteins, polysaccharides, and soluble organic matter are all degraded, with the final products being further treated through acidogenesis to yield volatile fatty acids (VFAs). The acidogenesis step is followed by acetogenesis, during which the VFAs are digested by acetogenic microorganisms producing a smaller molecule, acetate. The last step is methanogenesis, in which methane is generated. This process involves microorganisms called methanogens, which can be categorized into two groups according to their substrates. Acetoclastic methanogens (AMs) use acetate, while hydrogenotrophic methanogens (HMs) use H_2/CO_2 as substrates [10]. Through these AD steps, VFA accumulation often occurs because of the rapid degradation from the acidogenic process and thermodynamically unfavorable degradation [11].

The accumulation of propionic acid, one of the VFAs, has been reported as one of the important reasons for low methane production, as its propagation in the system decreases pH and subsequently inhibits methanogenic activity [6,12]. Enriched cultures of propionic-degrading microorganisms for bioaugmentation have been introduced as a solution to alleviate the acid accumulation, resulting in a more stable system and higher biogas productivity [13–16]. The technique is the practice of adding a particular microbial culture, which can be grown by using specific substrate as a carbon and energy source, to the unstable AD system for enhancing or boosting process performance. This relies on the fact that the propionate-degrading microbes are a key factor for the improvement of stability and efficiency of anaerobic treatment. Understanding the structure and microbial dynamism of the propionic-degrading communities, including mainly propionate degraders and methanogens, is required to better control and manage the microorganisms for reliability of the treatment systems.

A number of propionate-degrading microbes have been reported, with two main pathways of methylmalonyl Co-A (MMC) and dismutation. The MMC pathway was observed with *Syntrophobacter* sp. and *Pelotomaculum* sp. [17,18], and was mostly reported as a route of classical propionate degradation in AD. The overall reaction is: $Propionate^- + 3H_2O \rightarrow Acetate^- + HCO_3^- + H^+ + 3H_2$; $\Delta G^\circ = 76.1$ kJ/mol [19]. *Methanospirillum* sp. has been found as the main HM, required to maintain H_2 partial pressure for syntrophic activities with *Syntrophobacter* sp. [20–22]. On the other hand, the dismutation pathway was found with *Smithella propionica* which dismutates propionate to acetate and a butyrate through a six-carbon intermediate molecule. The overall equation is: $2Propionate^- + 2H_2O \rightarrow 3Acetate^- + H^+ + 2H_2$; $\Delta G^\circ = 48.4$ kJ/mol [23–25], giving more acetate and less hydrogen per one mole propionate compared to the MMC pathway. The *Smithella* was found as syntrophic-oxidizing bacteria with a number of HMs such as *Methanospirillum* sp. [26] and *Methanoculleus* sp. [27]. However, we believe that all related microbes of the processes have not been completely revealed.

Next-generation sequencing (NGS) technologies have been developed, generating a large amount of genetic sequences allowing culture-independent study of living organisms [28–30]. This provides a big advantage to understanding microbial communities as beforehand only a few percent of microorganisms could be studied by cultivation in laboratories. The 16S rRNA gene is a commonly used marker to identify microorganisms from a particular environment using NGS. It has also been applied to explore the AD systems for both lab-scale and full-scale digesters [31,32]. Several microorganisms in the AD process were revealed through NGS-based techniques in different digester conditions [33,34]. To our knowledge, a small number of propionate-degrading community studies

have been reported [35,36]. Variation of the communities as a whole system from different wastewater sources have still not been completely revealed. There is a need to extend the investigation of the microorganisms in propionate-degrading microbial communities, providing insight for microbial monitoring and manipulation to control the system stability and prevent failure.

Here, we observed anaerobic propionate-degrading communities via the enriched cultures inoculated from different sources of agro-industrial wastewater treatment plants. The microbiome profiles were investigated using a 16S rRNA-based sequencing approach. Firstly, we investigated the shift of microbiome profiles from inoculum to enrichment stages for revealing propionate-degrading communities. Then, we identified common and unique propionate-degrading microbes among the different sludge sources. We discuss this and conclude with the possible propionate-degrading communities and pathways specific to the original sludge sources.

2. Materials and Methods

2.1. Microorganisms and Enrichment Process

The propionate-degrading cultures used in this study were enriched from different anaerobic sludge sources. The anaerobic sludge was obtained from six full-scale wastewater treatment plants in Thailand, which treated domestic wastewater (Domestic), fruit juice-processing wastewater (FruitJuice), palm oil mill effluent (PalmOil), starch-processing wastewater (Starch), pig manure waste (PigManure), and seafood-processing wastewater (Seafood). Ten g/L from each sludge was inoculated in a 2-liter reactor-equipped gas counter and mixer at room temperature. To enrich the propionate-degrading cultures, all reactors were fed daily with sodium propionate as the sole carbon source. All reactors were operated for 7 months to increase the organic loading rate (OLR) to 3.0 g chemical oxygen demand (COD)/L/d and the hydraulic retention time (HRT) to 5 days. During the enrichment process, all reactors were evaluated by measuring pH, total volatile acid (TVA), alkalinity, COD reduction, and methane production to control the reactor performance. All enriched cultures were measured for specific methanogenic activity (SMA), using acetic acid as a substrate, with three replications. When operating at propionate loading rate of 3.0 g COD/L/d, the performance of all reactors and the activities of all enriched cultures are shown in Table 1.

Table 1. Performance of six reactors operating at propionate loading rate of 3.0 g COD/L/d and microbial activities of enriched propionate-degrading cultures.

Anaerobic Sludge from Various Anaerobic Wastewater Treatment Plants	Reactor Performance					Specific Methanogenic Activity (SMA) (g COD/g VSS/d)
	pH	TVA/Alkalinity	COD Reduction (%)	Biogas Composition (%)		
				$\%CH_4$	$\%CO_2$	
Domestic	7.50	0.30	86.5	60.0	35.5	0.22 ± 0.016
FruitJuice	7.50	0.30	85.0	62.5	34.0	0.17 ± 0.011
PalmOil	7.49	0.30	86.5	75.0	22.0	0.20 ± 0.009
Starch	7.56	0.25	90.0	73.5	23.5	0.22 ± 0.007
PigManure	7.57	0.27	89.0	75.5	21.0	0.28 ± 0.003
Seafood	7.52	0.35	80.0	80.0	17.5	0.14 ± 0.015

Remark: 1 mole propionate gives 1.75 mole methane and 1.25 mole carbon dioxide [37]. TVA: total volatile acid, COD: chemical oxygen demand.

2.2. Sample Collection and Molecular Analysis

To investigate the microbial communities of the anaerobic sludges obtained from the six full-scale anaerobic digesters (called inoculum) and from the enrichment process (called enriched cultures), DNA from all samples was extracted using DNeasy PowerSoil Kit. The extracted DNA was sequenced with TruSeq PCR-Free library following the manufacturer's protocol designed for the V3-V4 hypervariable region of the 16S rRNA gene. The universal primers, 319F-CCTAYGGGRBGCASCAG and 806-GGACTACNNGGGTATCTAAT, were utilized. The sequencing was based on the Illumina HiSeq

platform generating 250 bases paired-end reads. The obtained 16S rRNA gene-based sequencing data have been deposited at European Nucleotide Archive (ENA) under the accession number ERP113548.

2.3. Microbiome Analysis Based on the 16S rRNA Gene Sequences

The microbiome analysis of the enriched propionate-degrading culture was performed using Mothur software (version 1.39.5) [38] including the processes of data preprocessing, operational taxonomic units (OTUs) clustering, taxonomic assignment, and microbial diversity analysis. For the data preprocessing step, sequencing adapter sequences were removed and then paired-end reads were merged into contiguous sequences or contigs. Low-quality sequences which contained ambiguous bases (N), undesired length, off-target amplicon, or ≥ 8-base homopolymer length were discarded. The derived sequences were denoised using a Precluster algorithm to reduce single-base sequencing errors. The UCHIME algorithm [39] was used to remove chimeric sequences. The qualified sequences were then utilized for downstream analyses. The *de novo* OTU clustering was performed using 97% sequence similarity to identify the OTUs. Singletons (OTUs having only one sequence among all samples) were considered as sequencing errors and discarded. SILVA database version 132 [40] was utilized for taxonomic assignment of each OTU. Alpha diversity was measured to estimate sequencing coverage and microbial richness using Good's coverage and Chao1 indices, respectively. To make comparable microbial profiles, sequence abundances were normalized by a scaling technique based on the number of smallest total sequences among studied samples. OTUs with greater than 1% relative abundance across all samples were displayed in the microbial profiles. For beta-diversity analyses, Bray–Curtis dissimilarities among samples were measured for community comparison and used to visualize the principal coordinate analysis (PCoA) and heatmap. The visualization was performed using R version 3.6.1 (ggplot2 [41] and pheatmap [42] packages). Significant differences of the community profiles were estimated by analysis of similarity (ANOSIM) [43]. Dominant OTUs with greater than 1% relative abundance of each sample were retrieved for the identification of major common and unique organisms in propionate-degrading communities among different sludge sources. OTUs found in at least three out of five samples (excluding the Seafood sample) were reported as common OTUs in propionate-degrading communities.

3. Results

3.1. A Shift of Microbiome Profiles from Inoculums to Enriched Propionate-Degrading Cultures

Microbial communities of sludge inoculums obtained from different full-scale anaerobic wastewater treatment systems and their corresponding propionate-degrading cultures were identified using 16S rRNA gene sequencing. Richness of all samples estimated by Chao1 index vary from 1495.17 to 2811.46 taxa, showing a lower number of enriched cultures than inoculums (Table S1). The similarities of microbial community profiles between the inoculums and enriched propionate-degrading cultures are illustrated via a PCoA plot (Figure 1). Both inoculums and enriched cultures show trends of more similar microbial profiles at the same stages than the same sludge sources, except the Seafood sludge (Tables S2 and S3). The microbial profiles of inoculums are significantly different from the enriched cultures (p = 0.04). Figure 2 displays overall taxonomic profiles of all samples with their relative abundance. Nine out of 57 phyla are prevalent, having greater than 1% relative abundance across all samples (Figure 2A). Euryarchaeota, Proteobacteria, Firmicutes, Chloroflexi, and Synergistetes are found as the top five most abundant phyla. These phyla are dominant in both inoculum and enriched cultures, but their proportions are different in each sample. Figure 2B shows the assigned microbial community profiles at the genus level. The overall profiles and dissimilarity measures suggest a shift from inoculum to enriched stages. *Methanosaeta* is a dominant archaeon in all samples of both inoculums and enriched cultures (5.43%–38.72%), but with higher proportion in the enriched cultures. *Smithella*, one of the most abundant bacteria, increased their relative abundance in the enriched samples (0.51%–9.16% in the inoculums and 0.61%–26.09% in the enriched cultures). *Peptoclostridium* show

high proportion in the inoculums (0.98%–13.16%) but very less in the enriched cultures (0.06%–0.67%), whereas *Syntrophomonas* have low abundance in the inoculums (0.2%–3.03%) but are mostly prevalent in the enriched cultures (0.22%–18.91%). Particularly, the phylum Synergistetes represents high proportion in the Seafood sludge, distinguishing from other samples in both stages (43.76% and 36.10% at inoculum and enriched culture, respectively). These mainly belong to the genus of *Thermovirga* (1.23% in the inoculum and 23.37% in the enriched culture). In addition, *Syntrophobactor* and *Syner*-01 show high abundance in the enriched sample from the Seafood sludge (8.47% and 6.73%, respectively).

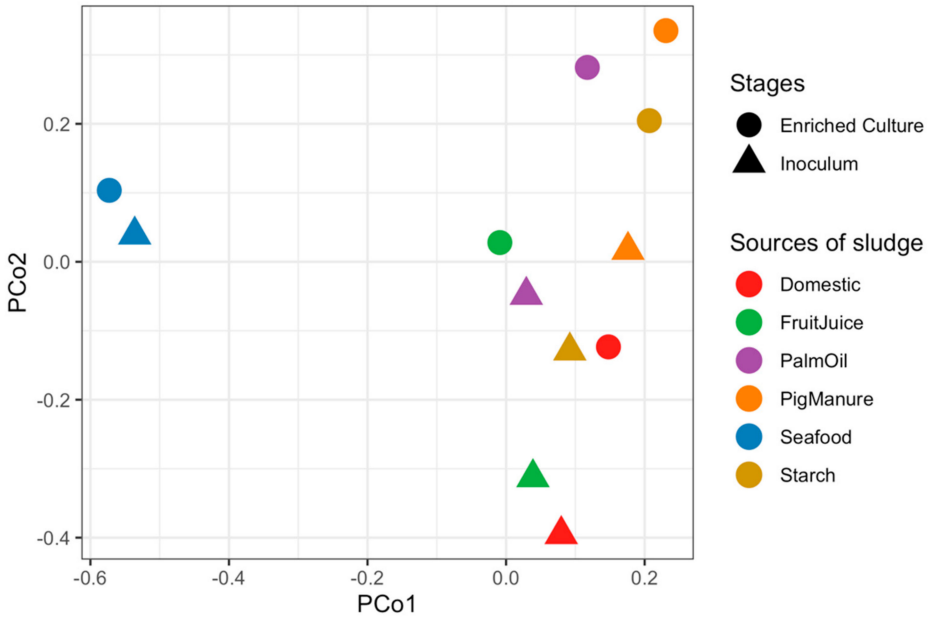


Figure 1. Principal Coordinate Analysis (PCoA) plot showing dissimilar microbial profiles of inoculums and enriched propionate-degrading cultures using the Bray–Curtis measure. Each dot represents an individual anaerobic digestion (AD) sample. Shapes represent stages of the samples: triangles for inoculums and circles for enriched cultures. The colors represent samples from different anaerobic sludge sources.

3.2. Microbiome Profiles of Propionate-Degrading Cultures Enriched from Different Inoculum Sources

Among the enriched cultures, common and distinct patterns of microbial profiles between different inoculum sources were revealed at the OTU resolution. Figure 3 shows a heatmap of dominant OTUs (greater than 1% relative abundance across all enriched samples) labeled at genus level with their relative abundance. The result reveals 52 dominant OTUs among the propionate-degrading cultures enriched from different inoculum sources from a total of 87 OTUs from both stages (Table S4). The enriched cultures of Domestic and FruitJuice show closet profiles among the six enriched cultures, followed by a pair of PigManure and Starch. *Methanosaeta* (OTU00003) and *Syntrophomonas* (OTU00011 and OTU00012) are commonly dominant in the enriched cultures of Domestic and FruitJuice. The Seafood sample showed the most distinguished profile compared to others. *Methanosaeta* (OTU00001) and *Smithella* (OTU00002) occurred with high abundance in all of the enriched cultures except the Seafood sample (5.86%–38.38% and 2.79%–23.77%, respectively). *Thermovirga* (OTU00007, 22.91%) is remarkable as a unique OTU dominant in the enriched culture from the Seafood sludge. In addition, *Syntrophobacter* (OTU00016, 8.13%), *Desulfobacteraceae* (OTU00040, 3.66%), *Methanospirillum* (OTU00025, 2.71%), and *Methanosaeta* (OTU00041, 2.37%) are also shown with higher abundance in the Seafood sample compared to others. Several unculturable dominant taxa of the class *Anaerolineae* were observed, for example, OTU00004 (18.62%) in the Seafood, OTU00009 (8.18%) and OTU00028 (1.75%) in the Starch, OTU00014 (7.14%) and OTU00046 (2.3%) in the PalmOil, and OTU00053 (1.34%) in the PigManure samples.

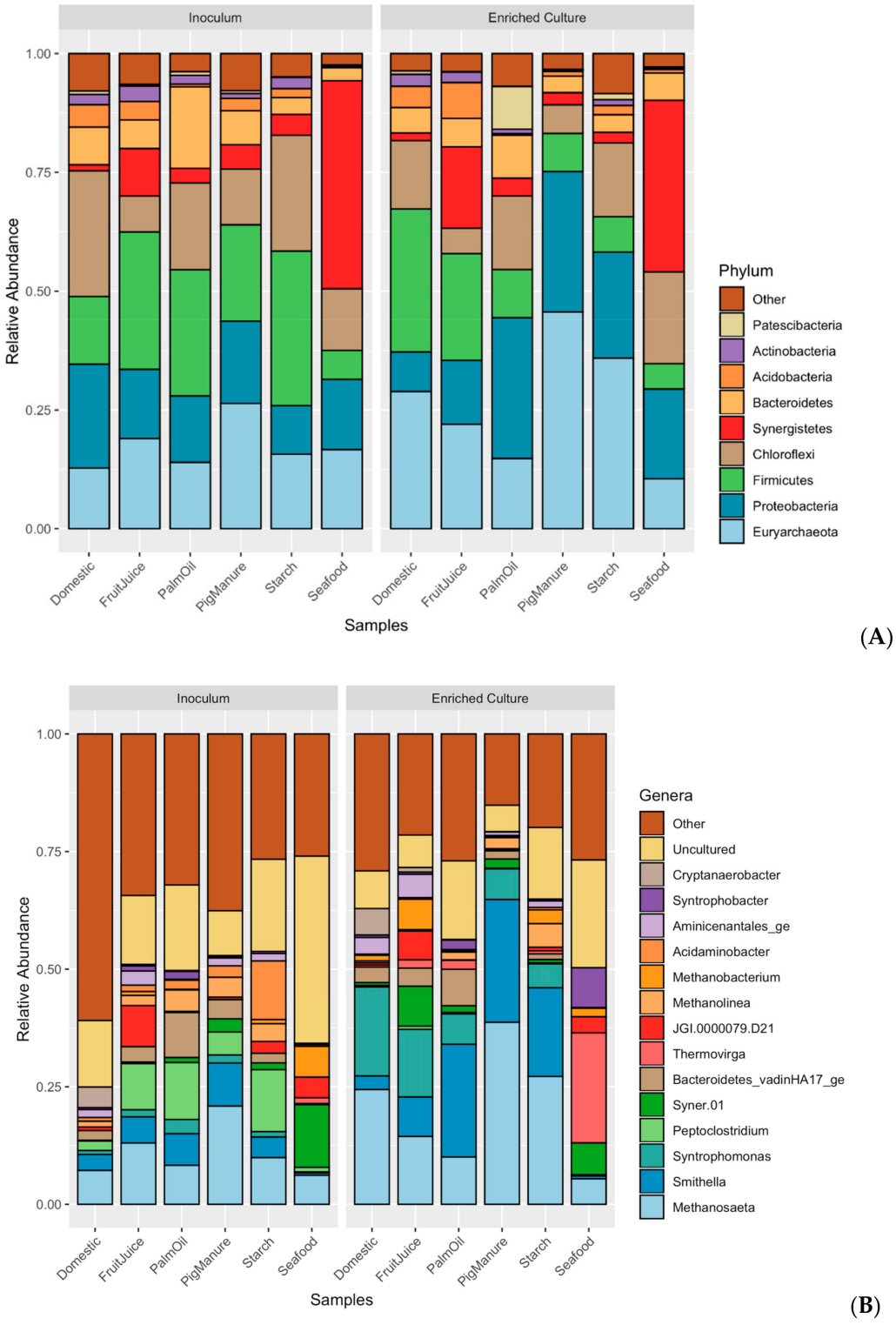


Figure 2. Microbial community profiles showing relative abundance of microbes derived from different anaerobic sludge sources at inoculum and propionate-enriched culture stages, respectively. (**A**) At the phylum level, 19 taxa are dominant (≥1% relative abundance) from the total of 57 assignable phyla. (**B**) At the genus level, 15 dominants of 875 assignable genera are shown.

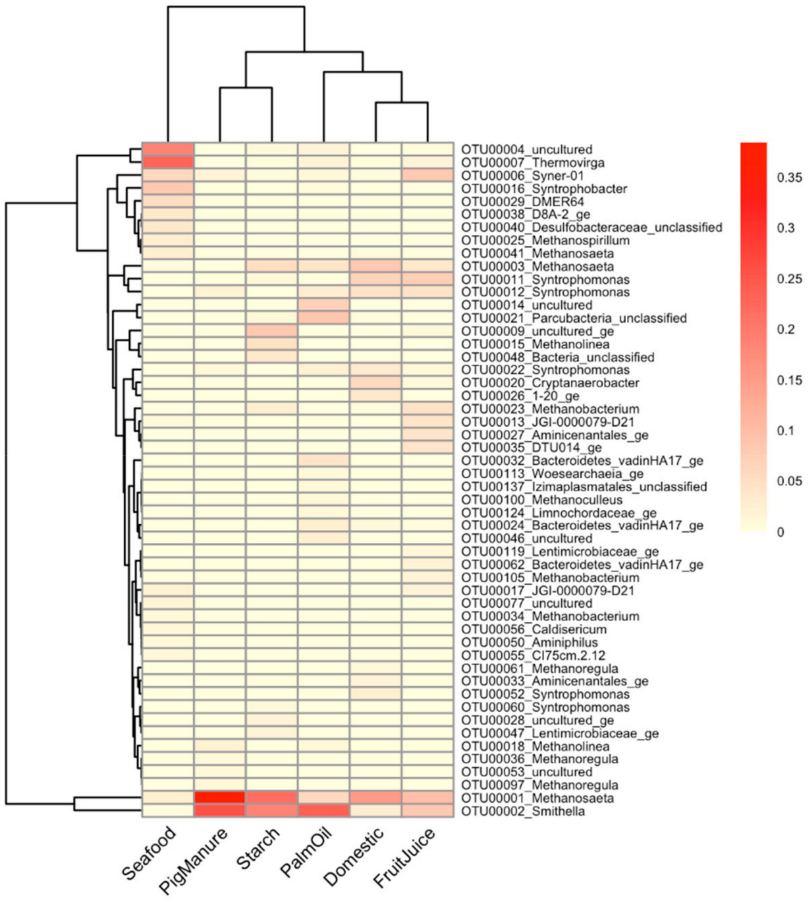


Figure 3. A heatmap shows microbial profiles of the propionate-degrading cultures enriched from different sources. 52 dominant OTUs are presented with their relative abundance. The dendrogram between samples (rows) and OTUs (columns) are drawn based on Bray–Curtis dissimilarity. The OTUs are assigned their taxonomic information at the genus level. The gradient color represents relative abundance of observed OTUs in each sample from low to high as light yellow to red, respectively.

3.3. Common and Unique Microorganisms in Propionate-Degrading Cultures Enriched from Different Inoculum Sources

We investigated common and unique microorganisms in the propionate-degrading communities among different inoculum sources (Table 2 and Table S5, respectively). Due to the very distinct resulting taxonomic profile of the enriched culture from the Seafood culture compared to other enriched cultures ($p = 0.018$; Table S3), the analysis was performed without the Seafood sample. Table 2 displays common microbes detected among the enriched propionate-degrading cultures (relative abundance greater than 1% in each sample). *Methanosaeta* (OTU00001) and *Smithella* (OTU00002) appeared as common microbes among all enriched cultures. *Syntrophomonas* is also a common genus in all cultures but with different OTUs (OTU00011, OTU00012, and OTU00022). The *Methanosaeta* (OTU00001) was discovered as a main AM. Another *Methanosaeta* (OTU00003) was also found in all enriched cultures except for the Starch sample. Interestingly, different genera of HMs were discovered in each inoculum source (Table 3). For example, *Methanoregula* was found dominantly in the enriched culture of the Domestic (1.20%) and PigManure (2.77%) samples, *Methanobacterium* was found in the FruitJuice sample (6.07%), *Methanolinea* was found in the Starch sample (4.97%), and *Methanoculleus* was found in the PalmOil sample (1.54%). The genus Syner-01 belonging to the family *Synergistaceae* (OTU00006) appears commonly in the enriched cultures of the Domestic, PalmOil and PigManure samples. Furthermore, OTUs of the family *Anaerolineaceae* (OTU00004, OTU00014, and OTU00046) were revealed dominantly in only the enriched

culture of the PalmOil sample. Some OTUs were found uniquely in each enriched culture but belong to the same genus of common OTUs such as *Syntrophomonas* in the Domestic sample, etc. (Table S5).

Table 2. A list of common propionate-degrading taxa enriched from different anaerobic agro-industrial sludge sources.

OTUs	Taxonomic Lineage	Propionate-Enriched Culture				
		Domestic	FruitJuice	PalmOil	PigManure	Starch
OTU00001	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosaetaceae; Methanosaeta	✓	✓	✓	✓	✓
OTU00002	Bacteria; Proteobacteria; Deltaproteobacteria; Syntrophobacterales; Syntrophaceae; Smithella	✓	✓	✓	✓	✓
OTU00003	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosaeta; Methanosaeta	✓	✓	✓	✓	
OTU00006	Bacteria; Synergistetes; Synergistia; Synergistales; Synergistaceae; Syner-01	✓		✓	✓	
OTU00011	Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae; Syntrophomonas	✓	✓			✓
OTU00012	Bacteria; Firmicutes; Cloastridia; Clostridiales; Syntrophomonadaceae; Syntrophomonas	✓	✓	✓	✓	
OTU00022	Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae; Syntrophomonas	✓	✓	✓	✓	

Table 3. A list of unique hydrogenotrophic methanogens in propionate-degrading cultures enriched from anaerobic agro-industrial sludge sources.

OTU	Hydrogenotrophic Methanogen		Observed Sample
	Family	Genus	
OTU00061	*Methanoregulaceae*	*Methanoregula*	Domestic
OTU00023	*Methanobacteriaceae*	*Methanobacterium*	FruitJuice
OTU00105	*Methanobacteriaceae*	*Methanobacterium*	FruitJuice
OTU00100	*Methanomicrobiaceae*	*Methanoculleus*	PalmOil
OTU00036	*Methanoregulaceae*	*Metanoregula*	PigManure
OTU00097	*Methanoregulaceae*	*Metanoregula*	PigManure
OTU00015	*Methanoregulaceae*	*Methanolinea*	Starch

3.4. Several Uncultured Microbes Found in the Propionate-Degrading Cultures Using the Culture-Independent Amplicon-Based Sequencing Approach

By performing the 16S rRNA gene sequencing, overall microbial communities of the samples have been revealed without the limitation of cultivation. In our study, the majority of the OTUs could be assigned their taxonomy as well-characterized microbes existing in the public databases (Table S1). However, 8.01% of the identified OTUs were classified as the dominant uncultured microbes at the genus level. These taxa are poorly defined in the available database and annotated as uncultured microbes in different taxonomic levels. Several uncultured microbes were detached in the enriched propionate-degrading communities (Table S6). For example, *Desulfobacteraceae* family was found in the enriched culture of the Seafood sample (1.66%). The class of Anaerolineceae was found with the highest number of OTUs in all the sludge samples, and dominant in several sludge sources such as Seafood, Starch, PalmOil, and PigManure (18.62%, 9.94%, 9.54%, and 1.34%, respectively). In addition, the Seafood samples contain high percent abundance of uncultured microbes in both the inoculums (39.29%) and the enriched cultures (22.89%) (Table S7).

4. Discussion

4.1. The Schematic Propionate-Degrading Pathway in the Enriched Cultures for Methane Production

With the limited carbon source of only propionate in the enriched cultures, microbial diversities in the samples were lower than in the inoculum sludges (Table S1). The discovered microbial community profiles and their degradation processes could be affected by the single carbon source feeding. Excluding the Seafood sample, our experiment revealed very small percentages of *Syntrophobacter* (<0.5%), which was previously proven as a propionate-degrading bacterium and found in most of the propionate-degrading communities along with HMs [22,35,44,45]. Interestingly, *Smithella* was found to be the dominant propionate-degrading bacterium [26] in our experiment, instead of the regular *Syntrophobacter*. There might be two main reasons for the presence of *Smithella* in the enriched cultures: (1) the nature of the original sludge containing a higher number of *Smithella* than *Syntrophobacter* (Figure 2; Table S4) and (2) *Syntrophobacter* prefers to grow with propionate and sulfate in the medium [23], which corresponds to our experiment that fed the medium without adding sulfate. The results suggest that the main reaction of the propionate degradation (Figure 4 and Table S8) is through *Smithella,* which can produce acetate and butyrate via a six-carbon intermediate, called the dismutation pathway [23–25]. The total reactions produced more acetate molecules compared to the classical pathway which belongs to *Syntrophobacter* and *Pelotomaculum* [23]. Following this theoretical perspective, we observed a higher abundance of *Methanosaeta*, which produces methane by acetate degradation, in the enriched samples [46]. Furthermore, *Syntrophomonas* was observed in several enriched samples. It was reported as a butyrate utilizer to produce acetate for AMs in the AD system [47]. Therefore, our studies suggest multi-trophic interaction of *Smithella* that can degrade propionate directly to acetate and convert propionate to butyrate, which is a substrate for *Syntrophomonas* (Figure 4). Consequently, *Methanosaeta* utilizes the resulting acetate from both organisms to produce methane and functions as a key AM in the enriched cultures.

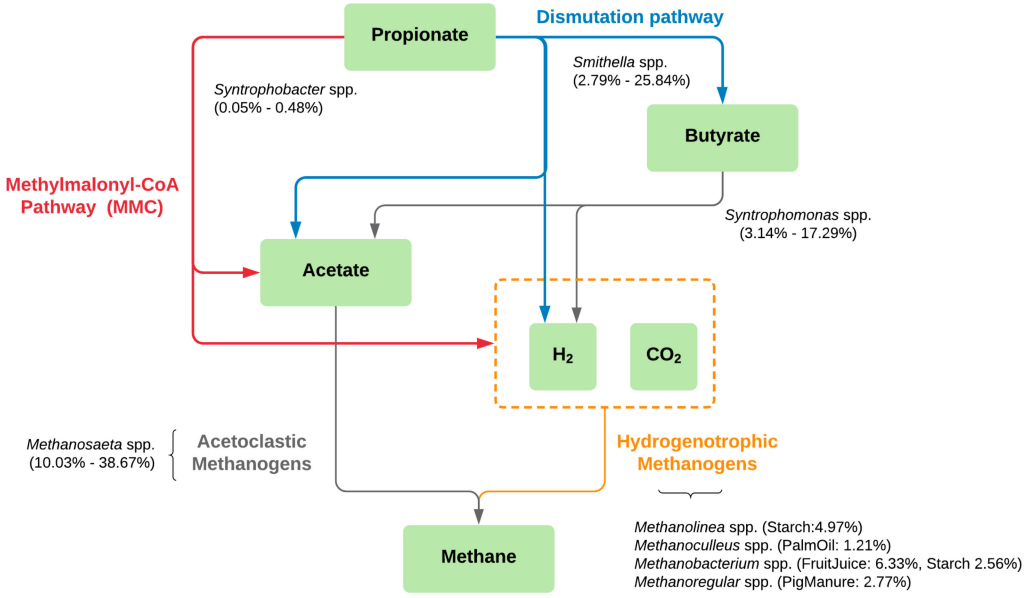


Figure 4. Schematic pathway of methane production based on propionate-degrading cultures enriched from different anaerobic sludge sources excluding the Seafood sample. Colors represent pathways of propionate degradation to methane production; red: methylmalonyl-CoA (MMC) pathway, blue: dismutation pathway, gray: acetoclastic pathway, and yellow: hydrogenotrophic pathway. Microbial taxa found in our study were drawn along the pathways with their percent relative abundance.

4.2. Different Taxa of Hydrogenotrophic Methanogens Found Specifically to Different Sludge Sources

While a single genus of AM was found as dominant taxa in all enriched samples, various genera of HMs were found particular to different sludge sources (Table 3 and Figure 4). In this study, *Methanobacterium*, *Methanoculleus*, and *Methanolinea*, were found in the FruitJuice, PalmOil, and Starch samples, respectively. Different OTUs of *Methanoregula* were found in the Domestic and PigManure samples. All of these HMs were reported in various mesophilic environments [48,49], and some of them, e.g., *Methanolinea* and *Methanoculleus*, were isolated from propionate-enrichment cultures as prevalent methanogen [50,51]. Although relatively smaller amounts of these HMs compared to AMs have been observed, they could also play a role in our systems for methane production by conversion of CO_2/H_2. These small amounts could also result from less H_2 produced from the dismutation pathway compared to the MMC pathway (Table S8). The observed HMs could refer to the syntrophic contribution of propionate degradation with *Smithella* [23]. Several types of HMs resulting from different wastewater treatment sludges suggest possible various pairs of syntrophic propionate oxidation and methane production between *Smithella* and HMs. The information of specific microbial taxa or communities of propionate degradation could be used as a guideline for microbial management, leading to efficient biogas production.

4.3. Unique Microbial Community in the Propionate-Degrading Culture Enriched from Seafood Sludge

The Seafood sludge revealed statistically distinct microbial profiles compared to the other sludges from different wastewater sources (Figure 2 and Table S3). *Thermovirga* and *Anaerolineaceae* uncultured groups affiliating to phylum Synergistetes and Chloroflexi, respectively, were found as prevalent organisms in the enriched propionate-degrading culture. *Thermovirga* were reported as amino acid degrading bacteria and were found dominantly in high salinity environments [52]. This is consistent with the condition of the Seafood sample, that originally contained high salinity. *Anaerolineaceae* were found in the AD system relating to granular formation and maintenance [53]. Both *Thermovirga* and *Anaerolineaceae* have been revealed dominantly with *Methanosaeta* in several AD experiments [54–56], suggesting that these microbes would play an important role in propionate degradation and biogas production pathways. *Syntrophobacter* and *Methanospirillum* were found as syntrophic propionate-oxidizing bacteria and H_2-utilizing methanogen, respectively [22,35]. These microbes have relatively higher abundance in the Seafood sample compared to the other five samples, suggesting an observation of the classical MMC pathway instead of our main discovered *Smithella* pathway (Figure 4). Furthermore, the Seafood sample showed the highest HM:AM ratio compared to other samples (Table S9). This corresponds to the result of a higher percent methane production but less SMA, indicating AM activities of utilizing acetates as substrates, compared to other samples (Table 1). The result suggested that the HMs would play more of a role in this sample as the MMC pathway provides more H_2 than the dismutation pathway (Table S8). The result showed that the Seafood sample has a unique profile and could be further investigated for the enrichment of methanogenic propionate degradation in a saline environment.

4.4. Overall Microbial Profiles of Propionate-Degrading Cultures and Unculturable Microbes Revealed Through Amplicon-Based Sequencing

The utilization of NGS allows the study of microbes taken directly from the samples without cultivation, showing all existing microbes with their abundance in the studied sample. Beforehand, a small number of known microbes has been studied, limited by cultivation [22,26,57]. Since microbes live as a community, this high-resolution technique provides a great opportunity to derive an overall picture of a microbial community and provides more insights to understand the dynamism of the studied consortium. In this study, a set of propionate-degrading communities was revealed according to their original sludge sources. Many OTUs of the class Anaerolineae were empirically revealed as predominant uncultured microbes in the enriched propionate-degrading cultures (Table S6). This microbe has been discovered dominantly in several AD systems [56]. In addition, Mcllroy S.J. et al. [56]

reported a member of Anaerolineae co-located with *Methanosaeta* spp., which was discovered in our study as major archaea. The function of the Anaerolineae and its synergistic relationship to *Methanosaeta* could be worth further investigation. The information from high-throughput sequencing provided a whole microbial community leading to better understanding of the control and management of the AD systems, as the microorganisms work together in the process.

5. Conclusions

The microbiome of the propionate-degrading communities enriched from different inoculum sources was investigated using 16S rRNA gene sequencing analysis. Interestingly, we found *Smithella* as the dominant propionate-degrading bacteria in most of the studied samples, suggesting the dismutation pathway of propionate degradation instead of the classical MMC pathway. The experiment supported a key role of *Smithella* and *Syntrophomonas* that implied a multi-trophic interaction of these two microorganisms to convert propionate to acetate and butyrate, and butyrate to acetate, respectively. A major abundance of *Methanosaeta* was observed as a main methanogen using acetate, while dominant HMs were found specific to different inoculum sources. The Seafood sludge sample shows a distinctive microbial profile containing *Thermovirga*, *Anaerolinaceae,* and *Methanosaeta* as dominant taxa, as well as *Syntrophobacter* and *Methanospirillum* which are mostly reported as regular syntrophic propionate-degrading culture through the MMC pathway. The highest HM:AM ratio was found in the Seafood sample, which corresponds to the MMC pathway producing more hydrogen that is utilized by HMs than the *Smithella* pathway. On the other hand, the relative abundances of AMs in the samples with the dismutation pathway were higher than in the Seafood sample, as more acetates are produced from that pathway. Furthermore, several uncultured bacteria of the class Anaerolinea were revealed in the enriched cultures. Our study shows that digesters with comparable performance and methane production could contain different communities of propionate-degrading microbes corresponding to their original sludge sources. The result suggests that inside information of specific propionate-degrading communities could be further applied to microbial monitoring and manipulation of wastewater treatment systems to increase biogas production.

Supplementary Materials: The following are available online at http://www.mdpi.com/2076-2607/8/2/277/s1, Table S1: Alpha-diversity estimation of microbiome in inoculums and their corresponding propionate-degrading cultures, Table S2: Statistical comparison of microbial profile between stages of inoculum and enriched culture using analysis of similarity (ANOSIM), Table S3: Statistical comparison of microbial profiles among different sources of anaerobic sludge using analysis of similarity (ANOSIM), Table S4: Relative abundance of 87 dominant OTUs (relative abundance greater than 1% in at least one sample) in our studied samples, Table S5: A list of unique propionate-degrading microbes enriched from different anaerobic sludge sources, Table S6: A list of SILVA-annotated uncultured microbes found dominantly (greater than 1% relative abundance) in the propionate-degrading cultures, Table S7: Percent relative abundance of uncultured microbes found in the inoculums and propionate-degrading cultures using the culture-independent amplicon-based sequencing approach, Table S8: Reactions of syntrophic metabolism of obligate proton-reducing acetogens and methanogens [19,25]. Table S9: Relative abundance of dominant methanogens found in the inoculum and propionate-degrading cultures and percentage HM:AM ratio.

Author Contributions: B.S., P.P. (Peerada Prommeenate), S.C., and K.K. contributed to the conception and design of the study; P.P. (Pantakan Puengrang) and K.K. carried out the microbiome and statistical analyses; B.S. and N.B. prepared samples and extracted DNA; P.P. (Pantakan Puengrang), B.S., P.P. (Peerada Prommeenate), N.B., S.C., M.T., and K.K. interpreted the results; K.K. supervised bioinformatics analysis; P.P. (Pantakan Puengrang) and K.K. wrote the first draft of the manuscript. All authors read, revised, and approved the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Thailand Research Fund (MRG6080183), King Mongkut's University of Technology, Thonburi, through the "KMUTT 55th Anniversary Commemorative Fund", the "KMUTT Research Fund", and the Joint Research Program Under the Cooperation Between Thailand and Japan (NRCT–JSPS) 2017.

Acknowledgments: We acknowledge the financial support from the Thailand Research Fund (MRG6080183), King Mongkut's University of Technology, Thonburi, through the "KMUTT 55th Anniversary Commemorative Fund", the "KMUTT Research Fund", the "Petchra Pra Jom Klao Master's Degree Research Scholarship from King Mongkut's University of Technology, Thonburi", and the Joint Research Program Under the Cooperation Between Thailand and Japan (NRCT–JSPS) 2017. We thank Kiattiyot Leahman for his technical support.

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

References

1. Bajpai, P. Basics of Anaerobic Digestion Process. In *Anaerobic Technology in Pulp and Paper Industry*; Springer: Berlin/Heidelberg, Germany, 2017.
2. Abdelgadir, A.; Chen, X.; Liu, J.; Xie, X.; Zhang, J.; Zhang, K.; Wang, H.; Liu, N. Characteristics, process parameters, and inner components of anaerobic bioreactors. *Biomed. Res. Int.* **2014**, *2014*, 841573. [CrossRef] [PubMed]
3. Amha, Y.M.; Anwar, M.Z.; Brower, A.; Jacobsen, C.S.; Stadler, L.B.; Webster, T.M.; Smith, A.L. Inhibition of anaerobic digestion processes: Applications of molecular tools. *Bioresour. Technol.* **2018**, *247*, 999–1014. [CrossRef] [PubMed]
4. Grzelak, J.; Ślęzak, R.; Krzystek, L.; Ledakowicz, S. Effect of pH on the production of volatile fatty acids in dark fermentation process of organic waste. *Ecol. Chem. Eng. S* **2018**, *25*, 295–306. [CrossRef]
5. Mir, M.A.; Hussain, A.; Verma, C. Design considerations and operational performance of anaerobic digester: A review. *Cogent Eng.* **2016**, *3*, 1181696. [CrossRef]
6. Ward, A.J.; Hobbs, P.J.; Holliman, P.J.; Jones, D.L. Optimisation of the anaerobic digestion of agricultural resources. *Bioresour. Technol.* **2008**, *99*, 7928–7940. [CrossRef]
7. Wrońska, I.; Cybulska, K. Quantity and quality of biogas produced from the poultry sludge optimized by filamentous fungi. *Ecol. Chem. Eng. S* **2018**, *25*, 395–404. [CrossRef]
8. Wacławek, S.; Grübel, K.; Silvestri, D.; Padil, V.V.; Wacławek, M.; Černík, M.; Varma, R.S. Disintegration of wastewater activated sludge (WAS) for improved biogas production. *Energies* **2019**, *12*, 21. [CrossRef]
9. Meyer, T.; Edwards, E.A. Anaerobic digestion of pulp and paper mill wastewater and sludge. *Water Res.* **2014**, *65*, 321–349. [CrossRef]
10. Mao, C.; Feng, Y.; Wang, X.; Ren, G. Review on research achievements of biogas from anaerobic digestion. *Renew. Sustain. Energy Rev.* **2015**, *45*, 540–555. [CrossRef]
11. Hu, D.; Su, H.; Chen, Z.; Cui, Y.; Ran, C.; Xu, J.; Xiao, T.; Li, X.; Wang, H.; Tian, Y. Performance evaluation and microbial community dynamics in a novel AnMBR for treating antibiotic solvent wastewater. *Bioresour. Technol.* **2017**, *243*, 218–227. [CrossRef]
12. Fukuzaki, S.; Nishio, N.; Shobayashi, M.; Nagai, S. Inhibition of the fermentation of propionate to methane by hydrogen, acetate, and propionate. *Appl. Env. Microbiol.* **1990**, *56*, 719–723. [CrossRef]
13. Lins, P.; Malin, C.; Wagner, A.O.; Illmer, P. Reduction of accumulated volatile fatty acids by an acetate-degrading enrichment culture. *FEMS Microbiol. Ecol.* **2010**, *71*, 469–478. [CrossRef]
14. Amani, T.; Nosrati, M.; Mousavi, S.; Kermanshahi, R. Study of syntrophic anaerobic digestion of volatile fatty acids using enriched cultures at mesophilic conditions. *Int. J. Env. Sci. Technol.* **2011**, *8*, 83–96. [CrossRef]
15. Tale, V.P.; Maki, J.S.; Struble, C.A.; Zitomer, D.H. Methanogen community structure-activity relationship and bioaugmentation of overloaded anaerobic digesters. *Water Res.* **2011**, *45*, 5249–5256. [CrossRef] [PubMed]
16. Li, Y.; Zhang, Y.; Sun, Y.; Wu, S.; Kong, X.; Yuan, Z.; Dong, R. The performance efficiency of bioaugmentation to prevent anaerobic digestion failure from ammonia and propionate inhibition. *Bioresour. Technol.* **2017**, *231*, 94–100. [CrossRef] [PubMed]
17. Houwen, F.P.; Plokker, J.; Stams, A.J.; Zehnder, A.J. Enzymatic evidence for involvement of the methylmalonyl-CoA pathway in propionate oxidation by Syntrophobacter wolinii. *Arch. Microbiol.* **1990**, *155*, 52–55. [CrossRef]
18. Kosaka, T.; Uchiyama, T.; Ishii, S.-I.; Enoki, M.; Imachi, H.; Kamagata, Y.; Ohashi, A.; Harada, H.; Ikenaga, H.; Watanabe, K. Reconstruction and regulation of the central catabolic pathway in the thermophilic propionate-oxidizing syntroph Pelotomaculum thermopropionicum. *J. Bacteriol.* **2006**, *188*, 202–210. [CrossRef]
19. Ferry, J.G. Fundamentals of methanogenic pathways that are key to the biomethanation of complex biomass. *Curr. Opin. Biotechnol.* **2011**, *22*, 351–357. [CrossRef]
20. Harmsen, H.J.; Van Kuijk, B.L.; Plugge, C.M.; Akkermans, A.D.; De Vos, W.M.; Stams, A.J. Syntrophobacter fumaroxidans sp. nov., a syntrophic propionate-degrading sulfate-reducing bacterium. *Int. J. Syst. Evol. Microbiol.* **1998**, *48*, 1383–1387. [CrossRef]

21. Wallrabenstein, C.; Hauschild, E.; Schink, B. Syntrophobacter pfennigii sp. nov., new syntrophically propionate-oxidizing anaerobe growing in pure culture with propionate and sulfate. *Arch. Microbiol.* **1995**, *164*, 346–352. [CrossRef]
22. Boone, D.R.; Bryant, M.P. Propionate-degrading bacterium, Syntrophobacter wolinii sp. nov. gen. nov., from methanogenic ecosystems. *Appl. Env. Microbiol.* **1980**, *40*, 626–632. [CrossRef]
23. Li, J.; Ban, Q.; Zhang, L.; Jha, A.K. Syntrophic propionate degradation in anaerobic digestion: A review. *Int. J. Agric. Biol.* **2012**, *14*, 843–850.
24. De Bok, F.A.; Stams, A.J.; Dijkema, C.; Boone, D.R. Pathway of propionate oxidation by a syntrophic culture of Smithella propionica and Methanospirillum hungatei. *Appl. Env. Microbiol.* **2001**, *67*, 1800–1804. [CrossRef]
25. Dolfing, J. Syntrophic propionate oxidation via butyrate: A novel window of opportunity under methanogenic conditions. *Appl. Env. Microbiol.* **2013**, *79*, 4515–4516. [CrossRef]
26. Liu, Y.; Balkwill, D.L.; Aldrich, H.C.; Drake, G.R.; Boone, D.R. Characterization of the anaerobic propionate-degrading syntrophs Smithella propionica gen. nov., sp. nov. and Syntrophobacter wolinii. *Int. J. Syst. Bacteriol.* **1999**, *49*, 545–556. [CrossRef]
27. Wawrik, B.; Marks, C.R.; Davidova, I.A.; McInerney, M.J.; Pruitt, S.; Duncan, K.E.; Suflita, J.M.; Callaghan, A.V. Methanogenic paraffin degradation proceeds via alkane addition to fumarate by 'Smithella'spp. mediated by a syntrophic coupling with hydrogenotrophic methanogens. *Environ. Microbiol.* **2016**, *18*, 2604–2619. [CrossRef]
28. Glenn, T.C. Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.* **2011**, *11*, 759–769. [CrossRef]
29. Handelsman, J. Metagenomics: Application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 669–685. [CrossRef]
30. Escobar-Zepeda, A.; Vera-Ponce de Leon, A.; Sanchez-Flores, A. The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics. *Front. Genet.* **2015**, *6*, 348. [CrossRef]
31. Wirth, R.; Kovacs, E.; Maroti, G.; Bagi, Z.; Rakhely, G.; Kovacs, K.L. Characterization of a biogas-producing microbial community by short-read next generation DNA sequencing. *Biotechnol. Biofuels* **2012**, *5*, 41. [CrossRef]
32. Schluter, A.; Bekel, T.; Diaz, N.N.; Dondrup, M.; Eichenlaub, R.; Gartemann, K.H.; 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*, 77–90. [CrossRef] [PubMed]
33. Cai, M.; Wilkins, D.; Chen, J.; Ng, S.-K.; Lu, H.; Jia, Y.; Lee, P.K. Metagenomic reconstruction of key anaerobic digestion pathways in municipal sludge and industrial wastewater biogas-producing systems. *Front. Microbiol.* **2016**, *7*, 778. [CrossRef] [PubMed]
34. Guo, J.; Peng, Y.; Ni, B.J.; Han, X.; Fan, L.; Yuan, Z. Dissecting microbial community structure and methane-producing pathways of a full-scale anaerobic reactor digesting activated sludge from wastewater treatment by metagenomic sequencing. *Microb. Cell Fact.* **2015**, *14*, 33. [CrossRef] [PubMed]
35. Li, Y.; Sun, Y.; Yang, G.; Hu, K.; Lv, P.; Li, L. Vertical distribution of microbial community and metabolic pathway in a methanogenic propionate degradation bioreactor. *Bioresour. Technol.* **2017**, *245*, 1022–1029. [CrossRef]
36. Li, Y.; Sun, Y.; Li, L.; Yuan, Z. Acclimation of acid-tolerant methanogenic propionate-utilizing culture and microbial community dissecting. *Bioresour. Technol.* **2018**, *250*, 117–123. [CrossRef]
37. Scholten, J.C.; Conrad, R. Energetics of syntrophic propionate oxidation in defined batch and chemostat cocultures. *Appl. Env. Microbiol.* **2000**, *66*, 2934–2942. [CrossRef]
38. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Env. Microbiol.* **2009**, *75*, 7537–7541. [CrossRef]
39. Edgar, R.C.; Haas, B.J.; Clemente, J.C.; Quince, C.; Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **2011**, *27*, 2194–2200. [CrossRef]
40. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glockner, F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2013**, *41*, D590–D596. [CrossRef]
41. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: Berlin/Heidelberg, Germany, 2016.

42. Kolde, R. Pheatmap: Pretty Heatmaps. R package version 1.0.12. 2019. Available online: https://rdrr.io/cran/pheatmap/ (accessed on 18 December 2019).
43. Clarke, K.; Green, R. Statistical design and analysis for a'biological effects' study. *Mar. Ecol. Prog. Ser.* **1988**, 213–226. [CrossRef]
44. Koch, M.; Dolfing, J.; Wuhrmann, K.; Zehnder, A.J. Pathways of propionate degradation by enriched methanogenic cultures. *Appl. Env. Microbiol.* **1983**, *45*, 1411–1414. [CrossRef]
45. Shigematsu, T.; Era, S.; Mizuno, Y.; Ninomiya, K.; Kamegawa, Y.; Morimura, S.; Kida, K. Microbial community of a mesophilic propionate-degrading methanogenic consortium in chemostat cultivation analyzed based on 16S rRNA and acetate kinase genes. *Appl. Microbiol. Biotechnol.* **2006**, *72*, 401–415. [CrossRef] [PubMed]
46. Chaiprasert, P.; Hudayah, N.; Auphimai, C. Efficacies of Various Anaerobic Starter Seeds for Biogas Production from Different Types of Wastewater. *Biomed. Res. Int.* **2017**, *2017*, 2782850. [CrossRef] [PubMed]
47. McInerney, M.J.; Bryant, M.P.; Hespell, R.B.; Costerton, J.W. Syntrophomonas wolfei gen. nov. sp. nov., an Anaerobic, Syntrophic, Fatty Acid-Oxidizing Bacterium. *Appl. Env. Microbiol.* **1981**, *41*, 1029–1039. [CrossRef]
48. Kitamura, K.; Fujita, T.; Akada, S.; Tonouchi, A. Methanobacterium kanagiense sp. nov., a hydrogenotrophic methanogen, isolated from rice-field soil. *Int. J. Syst. Evol. Microbiol.* **2011**, *61*, 1246–1252. [CrossRef] [PubMed]
49. Yashiro, Y.; Sakai, S.; Ehara, M.; Miyazaki, M.; Yamaguchi, T.; Imachi, H. Methanoregula formicica sp. nov., a methane-producing archaeon isolated from methanogenic sludge. *Int. J. Syst. Evol. Microbiol.* **2011**, *61*, 53–59. [CrossRef] [PubMed]
50. Kougias, P.G.; Campanaro, S.; Treu, L.; Zhu, X.; Angelidaki, I. A novel archaeal species belonging to Methanoculleus genus identified via de-novo assembly and metagenomic binning process in biogas reactors. *Anaerobe* **2017**, *46*, 23–32. [CrossRef]
51. Imachi, H.; Sakai, S.; Sekiguchi, Y.; Hanada, S.; Kamagata, Y.; Ohashi, A.; Harada, H. Methanolinea tarda gen. nov., sp. nov., a methane-producing archaeon isolated from a methanogenic digester sludge. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 294–301. [CrossRef]
52. Dahle, H.; Birkeland, N.K. Thermovirga lienii gen. nov., sp. nov., a novel moderately thermophilic, anaerobic, amino-acid-degrading bacterium isolated from a North Sea oil well. *Int. J. Syst. Evol. Microbiol.* **2006**, *56*, 1539–1545. [CrossRef]
53. Yamada, T.; Sekiguchi, Y.; Imachi, H.; Kamagata, Y.; Ohashi, A.; Harada, H. Diversity, localization, and physiological properties of filamentous microbes belonging to Chloroflexi subphylum I in mesophilic and thermophilic methanogenic sludge granules. *Appl. Env. Microbiol.* **2005**, *71*, 7493–7503. [CrossRef]
54. Wang, S.; Hou, X.; Su, H. Exploration of the relationship between biogas production and microbial community under high salinity conditions. *Sci. Rep.* **2017**, *7*, 1149. [CrossRef] [PubMed]
55. Liang, B.; Wang, L.Y.; Mbadinga, S.M.; Liu, J.F.; Yang, S.Z.; Gu, J.D.; Mu, B.Z. Anaerolineaceae and Methanosaeta turned to be the dominant microorganisms in alkanes-dependent methanogenic culture after long-term of incubation. *AMB Express* **2015**, *5*, 117. [CrossRef] [PubMed]
56. McIlroy, S.J.; Kirkegaard, R.H.; Dueholm, M.S.; Fernando, E.; Karst, S.M.; Albertsen, M.; Nielsen, P.H. Culture-independent analyses reveal novel anaerolineaceae as abundant primary fermenters in anaerobic digesters treating waste activated sludge. *Front. Microbiol.* **2017**, *8*, 1134. [CrossRef] [PubMed]
57. Imachi, H.; Sekiguchi, Y.; Kamagata, Y.; Ohashi, A.; Harada, H. Cultivation and in situ detection of a thermophilic bacterium capable of oxidizing propionate in syntrophic association with hydrogenotrophic methanogens in a thermophilic methanogenic granular sludge. *Appl. Env. Microbiol.* **2000**, *66*, 3608–3615. [CrossRef] [PubMed]

Evaluation of Shotgun Metagenomic Classification Performance via Microbial Richness and Diversity

Pichahpuk Uthaipaisanwong Systems Biology and Bioinformatics Research Laboratory, King Mongkut's University of Technology Thonburi, Thailand
pichahpuk@gmail.com
Tanawut Srisuk Systems Biology and Bioinformatics Research Laboratory, King Mongkut's University of Technology Thonburi, Thailand
tanawut.sri@mail.kmutt.ac.th
Kiattiyot Laeman Systems Biology and Bioinformatics Research Laboratory, King Mongkut's University of Technology Thonburi, Thailand
kiattiyot.la@mail.kmutt.ac.th
Supapon Cheevadhanarak Systems Biology and Bioinformatics Research Laboratory, King Mongkut's University of Technology Thonburi, Thailand
supapon.che@kmutt.ac.th
Kanthida Kusonmano† Systems Biology and Bioinformatics Research Laboratory, King Mongkut's University of Technology Thonburi, Thailand
kanthida.kus@mail.kmutt.ac.th

† Corresponding author

ABSTRACT
The advent of next-generation sequence technology has opened a new door for studying microbial community called metagenomics. The whole genome shotgun metagenomic approach provides a potential to analyze all genomic content in a particular environment and reveal microbiota components and their functions. The technique has been widely applied in several area such as soil, water, air, plant, animal and human. So far, the results of metagenomic studies strongly depend on the chosen computational analysis methods and parameters, which yield different microbial and functional profiles, and consequently affect the downstream interpretation. Despite the performance of available tools for shotgun metagenomic data have been compared, these reports mainly focused on run time, database size and/or taxonomic classification accuracy based on a fraction of reads classified. Nevertheless, the investigation in term of detected microbial richness and diversity has not explicitly been reported, which are the main results of a microbiome study. Here, widely used metagenomic tools including Centrifuge, Kraken, MEGAN and MetaPhlAn2 are examined for their performance of taxonomic classification. Not only considering the sequencing reads whether they are accurately classified but the resulting microbial richness and diversity are also investigated. To evaluate the tools, we generated mock communities combining mock and shuffled input metagenomic reads. We report performance evaluation of commonly used metagenomic tools with their different resulting microbial profiles. The study could provide a guideline for choosing a metagenomic tool for more accurate reported taxonomic profiles.

KEYWORDS
Metagenomics, Taxonomic classification, Whole metagenome shotgun, Microbial richness, Microbial diversity

MMR-O-06

Hypervariable regions assessment for 16S rRNA gene-based metagenomic analysis in activated sludge and anaerobic digestion

Chalida Rangsiwutisak[1], Supapon Cheevadhanarak[1,2], and Kanthida Kusonmano[1,2,*]

[1]Systems Biology and Bioinformatics Research Laboratory, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bang Khun Thian, Bangkok, 10150, Thailand
[2]Bioinformatics and Systems Biology Program, School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bang Khun Thian, Bangkok, 10150, Thailand
*Email: kanthida.kus@mail.kmutt.ac.th

ABSTRACT

Metagenomics has become popular to study microbial communities in wastewater treatment for manipulating pollution control and biogas production. The technique allows broad study of microbial communities without culturing by directly sequencing genetic materials from an environment sample using next-generation sequencing (NGS) technology. The 16s rRNA gene-based metagenomic analysis is a commonly used method for identifying bacteria and archaea in studied samples by sequencing targeted hypervariable regions on their 16s rRNA genes. A choice of a selected hypervariable region and its primers has been shown as an important factor for detecting microbes. Recent studies show that different hypervariable regions could capture different community profiles depending on types of samples and microbial compositions. Nevertheless, there is no specific report for the study of microbial communities in wastewater treatment. Here we performed bioinformatics analysis to evaluate sensitivity of commonly used primers in activated sludge and anaerobic digestion studies. The results show that overall of the studied primers could capture 85-99% of all bacteria and archaea based on the studied 16s rRNA gene database of activated sludge and anaerobic digesters. From our study, V4 and V4-V5 primers provide better resolution to detect the microorganisms than V3-V4 primers especially those in the phyla Euryarchaeota, Thaumarchaeota and Woesearchaeota. Moreover, target size of each hypervariable region are reported, which are useful information for filtering amplified targeted regions in metagenomic analysis. This study could be introduced as a guideline for selecting optimal hypervariable regions for 16s rRNA gene-based metagenomics in wastewater treatment systems.

KEYWORDS: 16s rRNA gene, Activated sludge, Anaerobic digestion, Hypervariable region, Metagenomics

INTRODUCTION

Microbes live as a community and survive in various sources of wastewater such as food waste (Hagen et al., 2017; He et al., 2017), manure slurry (Lin et al., 2016), mining (Ma et al., 2015), and residence (He et al., 2017). In wastewater treatment systems, microbes are key players for pollution control by degrading from complex organic polymers into small molecules. In anaerobic digestion systems, the degradation processes result formation of biogas (i.e. methane and carbon dioxide), which has been considered to contribute as sustainable energy. The systems have been studied not only for treating water but also produce biogas as alternative fuel source (Doloman, Soboh, Walters, Sims, & Miller, 2017). To control and manage the wastewater treatment system, microbial management is essential and understanding the microbial communities is a key step. Nevertheless, a clear picture of the complex communities and their behaviors has not been characterized. The studies to understand the communities are still going on (Doloman et al., 2017; Kouzuma et al., 2017; Tian, Zhang, & Yang, 2018).

With the availability of the next generation sequencing (NGS) technology, living cells can be studied in molecular levels in a high-throughput manner. Recently, metagenomic technique has been introduced to study microbial communities by sequencing genetic materials from a particular environment. Unlike the conventional microbial techniques that are culture-dependent techniques, metagenomics extracts DNAs directly from a sample of the study without cultivation. The technique provides overall pictures of microorganism and has been more applied to study microorganisms including in wastewater treatment systems (Doloman et al., 2017; He et al., 2017; Lin et al., 2016; Tian et al., 2018).

The 16s rRNA gene-based metagenomic analysis is one of commonly used methods for identifying microbial communities in studied samples by sequencing targeted hypervariable regions on their 16s rRNA genes. These gene encodes small subunit of ribosomal with approximate lengths of 1,500 bp. (Jo, Kennedy, & Kong, 2016), containing conserved and nine hypervariable regions (V1-V9) regions which are specific to each organisms (B. Yang, Wang, & Qian, 2016). From this reasons, selected hypervariable regions of 16s rRNA genes were used extensively in varieties of taxonomic investigation based on Illumina sequencing platform (Guo, Ju, Cai, & Zhang, 2013). With the popular Illumina sequencing, high-throughput sequences could be produced at a time but the derived sequences are relatively short (about 100-300 base pairs). Thus, one or consecutive two hypervariable regions is selected to study microorganism to represent the studied microbes. Sets of commonly used primers from Illumina and some modified primers are specified to various hypervariable regions of 16s rRNA genes such as V3-V4 (Doloman et al., 2017; S. Yang et al., 2017), V4 (Kouzuma et al., 2017; Lin et al., 2016) and V4-V5 (Mei, Narihiro, Nobu, & Liu, 2016). Those targeted hypervariable regions has been reported to provide sufficient results for investigating microorganisms in metagenomic studies. After

sequencing, the targeted sequences are clustered into OTUs (Operational Taxonomic Units) and annotated their taxonomy by searching for similarity to the existing sequences in available databases (Doloman et al., 2017) such as Microbial Database for Activated Sludge (MiDAS) (McIlroy et al., 2017), Silva (Quast et al., 2012), Greengenes (DeSantis et al., 2006), and Ribosomal Database Project (RDP) (Cole et al., 2013).

Even though metagenomic studies have been increasing, a choice of a selected hypervariable region and its primers is still remained as a question since it has been shown as an important factor for deriving community profiles. Several recent studies show that different hypervariable regions could capture different community profiles depending on types of samples and microbial compositions (Graspeuntner, Loeper, Künzel, Baines, & Rupp, 2018; B. Yang et al., 2016). For example, hypervariable region V3-V4 has been shown to capture most of bacterial and archaeal profiles in activated sludge studies (Cai, Ye, Tong, Lok, & Zhang, 2013; Guo et al., 2013) and vaginal microbiome study (Graspeuntner et al., 2018). Another report showed that hypervariable region V4-V6 is suitable for microbial taxonomy identification in general condition by computational analysis (B. Yang et al., 2016). Moreover, one study showed that the reliable of hypervariable region V4 could represent general taxonomy of microbes better than using whole sequence of 16s rRNA gene (B. Yang et al., 2016). Nevertheless, there is no systematic investigation for detecting microbial profiles using different targeted hypervariable regions in wastewater treatment, particularly activated sludge and anaerobic digestion. There is a need of a guideline for selecting a hypervariable region to perform the metagenomic analysis.

Here we performed bioinformatics analysis to evaluate sensitivity of selected universal primers in wastewater treatment systems using Microbial Database for Activated Sludge (MiDAS) (McIlroy et al., 2017; McIlroy et al., 2015), an integrated collection of microbial sequences in activated sludge and anaerobic digesters. Comparative genomics analysis has been performed in order to investigate sensitivities of the selected universal primers to capture the desired targeted regions and annotate the studied microbes. Moreover, we also report the target lengths of each hypervariable region among microorganisms by commonly used primers in metagenomic studies, which are useful information for filtering desired sizes of the amplified targeted regions. We report the results of hypervariable regions assessment, which could be suggested as a guideline for selecting a hypervariable region for 16s rRNA gene-based metagenomics in wastewater treatment systems.

MATERIALS AND METHODS

1. The investigated 16s rRNA genes and metagenomic primer sequences

In order to investigate 16s rRNA gene of microorganisms in activated sludge and anaerobic digesters, the data were retrieved from MiDAS (http://www.midasfieldguide.org/en/download/).

MiDAS is a huge collection of curated taxonomy of 16s rRNA genes specifically to activated sludge and anaerobic digesters. The database integrates 16s rRNA genes from SILVA database, a comprehensive ribosomal RNA sequences, and Danish wastewater treatment plants. It comprises 21,483 archaea and 453,045 bacteria. Four commonly used primers targeting hypervariable regions were selected based on the most reliable regions for representing 16s rRNA gene (B. Yang et al., 2016). These primers are universal primers in metagenomic studies based on Illumina platform and used in various environments including wastewater. This study includes two pairs of V3-V4, one pair of V4, and one pair of V4-V5. Information of the primers are shown in Table 1.

Table 1 The information of commonly used primers for 16s rRNA genes-based metagenomic analysis.

Hypervariable regions	Primer name	Primer sequences	Source
V4	515F	5'-GTGCCAGCMGCCGCGGTAA-3'	(Kouzuma et al., 2017; Lin et al., 2016)
	806R	5'-GGACTACHVGGGTWTCTAAT-3'	
V3-V4 (1)	341F	5'-CTAYGGGRBGCASCAG-3'	(S. Yang et al., 2017)
	806R	5'-GGACTACNNGGGTATCTAAT-3'	
V3-V4 (2)	338F	5'-ACTCCTACGGGAGGCAGC-3'	(Doloman et al., 2017)
	785R	5'-TACNVGGGTATCTAATCC-3'	
V4-V5	515F	5'-GTGCCAGCMGCCGCGGTAA-3'	(Mei et al., 2016)
	909R	5'-CCCCGYCAATTCMTTTRAGT-3'	

2. Hypervariable regions assessment for 16S rRNA gene-based metagenomic analysis

In this study, we investigated sensitivities of the selected primers targeting different hypervariable regions of 16s rRNA genes of microorganisms in wastewater treatment systems. The studied primers were mapped to the reference 16s rRNA genes from the MiDAS database by bioinformatics approaches. The primer sequences were aligned to the 16s rRNA genes database using Clustal Omega version 1.2.4 (Sievers et al., 2011) with default parameters. This is to determine whether the primers could amplify the desired target regions. From the alignment results, the targeted hypervariable regions, or the regions between the forward and reverse primers, were extracted by our in-house python scripts. In our analysis, the reported candidate target lengths were determined using the criteria that a particular position contains equal or greater than 100 detected sequences.

RESULTS AND DISCUSSION

1. Using different hypervariable regions can overall detect microbes in activated sludge and anaerobic digestion systems similarly

We first investigated how the metagenomic primers for different hypervariable regions (Table 1) can capture the targeted sequences of the microbial communities in wastewater treatment systems. By our computational analysis, the results show that 85 – 99% of 16s rRNA genes can be captured within the candidate target sizes based on the analyzed primers (Figure 1). Larger proportions of bacteria sequences can be detected by the studied primers compared to archaea as the greater proportion of the bacteria domain in the database. From the analytic results, the primers of V4 and V4-V5 hypervariable regions provide better resolutions to detect archaea and bacteria compared to V3-V4. Even through the primer positions of two analyzing V3-V4 regions are very similar, numbers of captured sequences are different. The results suggest that a selection of hypervariable regions could slightly reflect the numbers of detected microbes, and different primers of the same hypervariable region could detect microbial communities differently (Nikolaki & Tsiamis, 2013; Youssef et al., 2009).

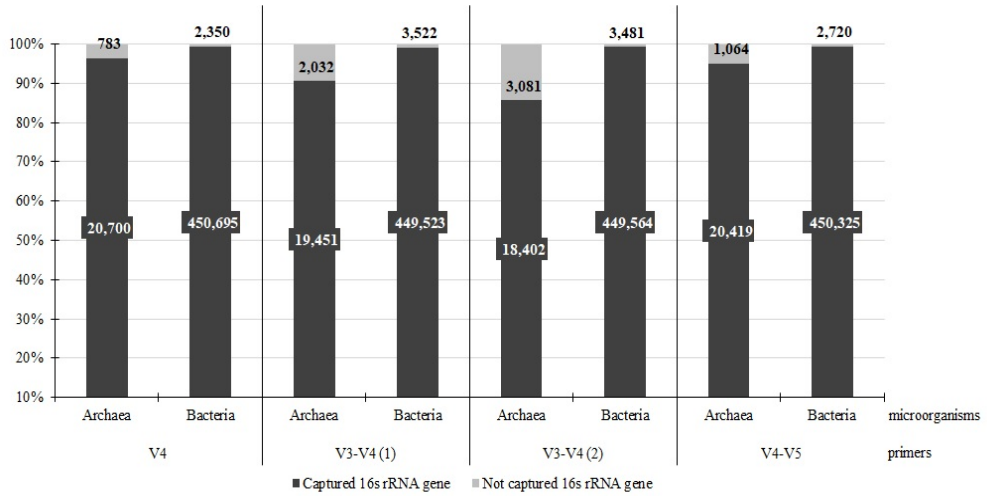


Figure 1 Proportions of captured 16s rRNA genes using different hypervariable region primers. The analyzed 16s rRNA genes are based on MiDAS database containing microbes related to wastewater treatment systems with total numbers of 21,483 archaea and 453,045 bacteria.

2. Target sizes of the studied hypervariable regions are varying among archaea and bacteria

Table 2 and Figure 2 show ranges of candidate target sizes detected by the studied primers. The target sizes captured by V4 and V4-V5 primers are slightly shifted between archaea and bacteria. However, the candidate target sizes by V3-V4 primers are not in an overlapped range. The results indicate different target sizes between archaea and bacteria and give a notification for filtering desired targeted sequences in metagenomic analysis. For example, Mothur (Schloss et al., 2009) requires users to specify the target length of interest. Having such information would provide more accurate results of the analysis. Furthermore, we found that both primers of V3-V4 provide the same candidate target sizes. Comparing

to *E. coli*, the identified candidate target sizes from V4 and V4-V5 primers were slightly smaller than *E. coli* depending on the binding positions of primers (Table 2). On the other hand, the candidate target sizes from V3-V4 primers cover V3-V4 region of *E. coli* (Table 2) (B. Yang et al., 2016). From the analytic results, the prediction show the reliable of candidate target sizes.

Table 2 Candidate target sizes of the studied hypervariable regions of archaea and bacteria and *E. coli*'s target size.

Hypervariable regions	Domain		*E. coli* (B. Yang et al., 2016)
	Archaea	Bacteria	
V4	252 - 256 bp.	243 - 266 bp.	259 bp.
V3-V4 (1)	382 - 393 bp.	397 - 443 bp.	440 bp.
V3-V4 (2)	382 - 393 bp.	397 - 443 bp.	440 bp.
V4-V5	377 – 383 bp.	363 – 393 bp.	398 bp.

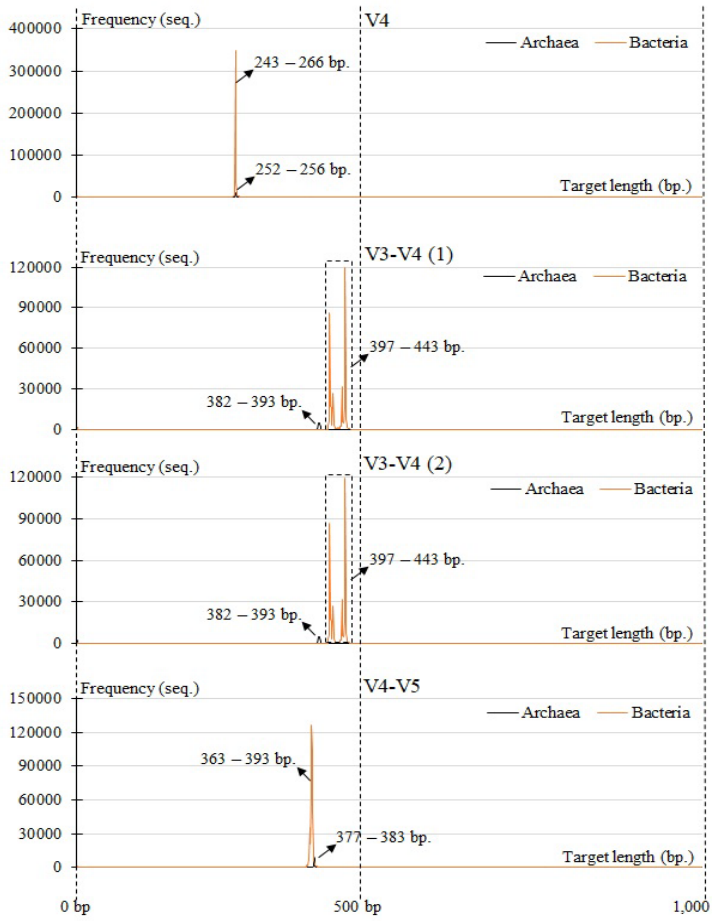


Figure 2 Distribution of target sizes of different hypervariable region primers, V4 (A), V3-V4 (B-C), and V4-V5 (D). X-axis represents a target length and Y-axis represents a number of detected 16s rRNA genes using the primers. The black and orange lines represent the domain of archaea and bacteria, respectively.

3. Some microbial organisms in activated sludge and anaerobic digestion studies could be detected with different resolution among different hypervariable regions

We assessed sensitivities for capturing microbes using different hypervariable region primers for 16s rRNA genes-based metagenomic analysis. All captured 16s rRNA genes with candidate target length were assigned their taxonomy according to MiDAS database. Taxonomic annotations of the captured archaea and bacteria using different hypervariable regions are shown in Figure 3 and Figure 4, respectively. The top detected archaea (Figure 3A, 3C) and bacteria (Figure 4A, 4C) are very similar among different utilized primers in both phylum and genus levels. In addition, the abundances of each detected organisms are slightly different. By considering the abundances of the detected phylum in archaea, the phylum of Thaumarchaeota is the one that can be captured less using both V3-V4 primers. Interestingly, the abundances of Thaumarchaeota and Miscellaneous Crenarchaeotic Group are different between V3-V4 (1) and V3-V4 (2) even though the same hypervariable regions are utilized. This suggests that different primers used in a particular variable region could also provide different resolution of microbes detected.

Figure 3B and Figure 3D show the numbers of archaea that could not be captured by the studied primers in phylum and genus levels, respectively. From our analysis, V3-V4 primers could not capture many of three archaeal phyla including Euryarchaeota (6.6% and 9.2% of total Euryarchaeota for V3-V4(1) and V3-V4(2), respectively), Thaumarchaeota (14.2% and 23.2% of total Thaumarchaeota), and Woesearchaeota (DHVEG-6) (39.7% and 42.8% of total Woesearchaeota) more than other primers (Figure3B). This should be noted especially if the phylum is in the studied samples or under a specific interest. For instance, Euryarchaeota include methanogens which play an important role for producing methane. The organism has been focused in anaerobic wastewater treatment systems for enhancing biogas production (Doloman et al., 2017). In bacteria domain, several Proteobacteria (0.54% - 0.74% of total Proteobacteria for all studied primers), Firmicutes (0.46% - 0.58% of total Firmicutes), Actinobacteria (0.32% - 0.52% of total Actinobacteria) and Planctomycetes (0.86% - 7.79% of total Planctomycetes) could not be captured. Interestingly, Planctomycetes phylum could not be captured by V3-V4 primers for about 4-10 times comparing to other primers (Figure 4B). At the genus level, the less captured bacteria include unknown, uncultured, and *Bacillus* (Figure 4D). Many of unknown and uncultured genus in archaea and bacteria were not detected by the selected primers (Figure 3D and Figure 4D). This suggests current commonly used primers could still not capture some non-informative microorganisms pointing out a limitation to study novel microorganisms in metagenomics.

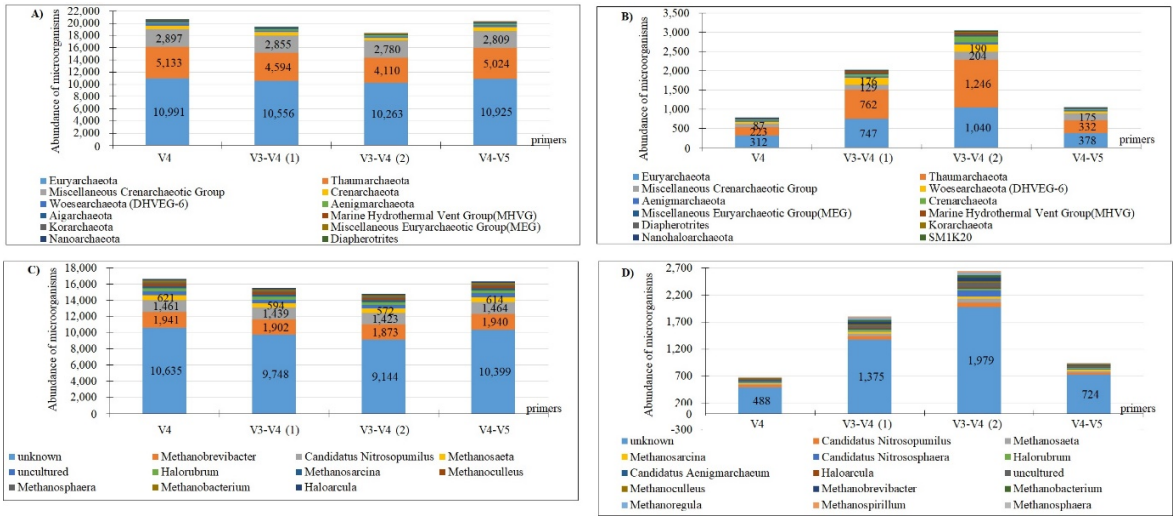


Figure 3 Abundance of top 10 archaeal microorganisms that could be captured (left) and could not be captured (right) using different hypervariable primers at phylum (A-B) and genus (C-D) levels, respectively.

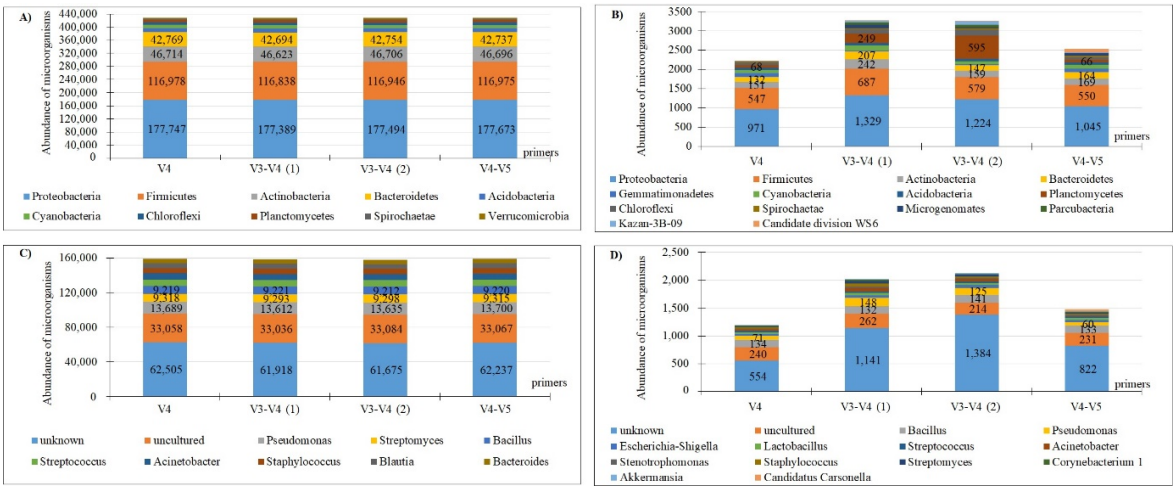


Figure 4 Abundance of Top10 bacterial microorganisms that could be captured (left) and could not be captured (right) using different hypervariable primers at phylum (A-B) and genus (C-D) levels, respectively.

CONCLUSION

From our analysis, different hypervariable region primers for 16s rRNA genes-based metagenomics could capture the majority of the microorganisms in wastewater treatment systems based on MiDAS database. V4 and V4-V5 primers provide better resolution to detect the microorganisms than V3-V4 primers. Some microbes could be captured less using V3-V4 primers than V4 and V4-V5. Moreover, candidate target sizes of each analytic primers are reported showing slightly different ranges of target

size of archaea and bacteria domains. This study could serve as a guideline for selecting hypervariable primers for 16s rRNA genes-based metagenomics in wastewater treatment systems. In addition, the identified candidate target size could be provided as background information to filter the desired target size for more accurate metagenomic analysis.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support provided by Thailand Research Fund (MRG6080183), King Mongkut's University of Technology Thonburi through the "KMUTT 55th Anniversary Commemorative Fund" and "KMUTT Research Fund". We thank Mr. Kiattiyot Leahman for technical supports of computational analysis and Dr. Pichahpuk Uthaipaisanwong for fruitful scientific discussion.

REFERENCES

Cai, L., Ye, L., Tong, A. H. Y., Lok, S., & Zhang, T. (2013). Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. *PloS one, 8*(1), e53649. doi:https://doi.org/10.1371/journal.pone.0053649

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., . . . Tiedje, J. M. (2013). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic acids research, 42*(D1), D633-D642. doi:https://doi.org/10.1093/nar/gkt1244

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., . . . Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology, 72*(7), 5069-5072. doi:http://doi.org/10.1128/AEM.03006-05

Doloman, A., Soboh, Y., Walters, A. J., Sims, R. C., & Miller, C. D. (2017). Qualitative Analysis of Microbial Dynamics during Anaerobic Digestion of Microalgal Biomass in a UASB Reactor. *International journal of microbiology, 2017*. doi:https://doi.org/10.1155/2017/5291283

Graspeuntner, S., Loeper, N., Künzel, S., Baines, J. F., & Rupp, J. (2018). Selection of validated hypervariable regions is crucial in 16S-based microbiota studies of the female genital tract. *Scientific reports, 8*(1), 9678. doi:https://doi.org/10.1038/s41598-018-27757-8

Guo, F., Ju, F., Cai, L., & Zhang, T. (2013). Taxonomic precision of different hypervariable regions of 16S rRNA gene and annotation methods for functional bacterial groups in biological wastewater treatment. *PloS one, 8*(10), e76185. doi:https://doi.org/10.1371/journal.pone.0076185

Hagen, L. H., Frank, J. A., Zamanzadeh, M., Eijsink, V. G., Pope, P. B., Horn, S. J., & Arntzen, M. Ø. (2017). Quantitative metaproteomics highlight the metabolic contributions of uncultured phylotypes in a thermophilic anaerobic digester. *Applied and environmental microbiology, 83*(2), e01955-01916. doi:https://doi.org/10.1128/AEM.01955-16

He, Q., Li, L., Zhao, X., Qu, L., Wu, D., & Peng, X. (2017). Investigation of foaming causes in three mesophilic food waste digesters: reactor performance and microbial analysis. *Scientific reports, 7*(1), 13701. doi:https://doi.org/10.1038/s41598-017-14258-3

Jo, J.-H., Kennedy, E. A., & Kong, H. H. (2016). Bacterial 16S ribosomal RNA gene sequencing in cutaneous research. *J Invest Dermatol, 136*(3), e23-e27. doi:https://doi.org/10.1016/j.jid.2016.01.005

Kouzuma, A., Tsutsumi, M., Ishii, S. i., Ueno, Y., Abe, T., & Watanabe, K. (2017). Non-autotrophic methanogens dominate in anaerobic digesters. *Scientific reports, 7*(1), 1510. doi:https://doi.org/10.1038/s41598-017-01752-x

Lin, Q., He, G., Rui, J., Fang, X., Tao, Y., Li, J., & Li, X. (2016). Microorganism-regulated mechanisms of temperature effects on the performance of anaerobic digestion. *Microbial cell factories, 15*(1), 96. doi:https://doi.org/10.1186/s12934-016-0491-x

Ma, Q., Qu, Y.-Y., Zhang, X.-W., Shen, W.-L., Liu, Z.-Y., Wang, J.-W., . . . Zhou, J.-T. (2015). Identification of the microbial community composition and structure of coal-mine wastewater treatment plants. *Microbiological research, 175*, 1-5. doi:https://doi.org/10.1016/j.micres.2014.12.013

McIlroy, S. J., Kirkegaard, R. H., McIlroy, B., Nierychlo, M., Kristensen, J. M., Karst, S. M., . . . Nielsen, P. H. (2017). MiDAS 2.0: an ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups. *Database, 2017*(1). doi:https://doi.org/10.1093/database/bax016

McIlroy, S. J., Saunders, A. M., Albertsen, M., Nierychlo, M., McIlroy, B., Hansen, A. A., . . . Nielsen, P. H. (2015). MiDAS: the field guide to the microbes of activated sludge. *Database, 2015*. doi:https://doi.org/10.1093/database/bav062

Mei, R., Narihiro, T., Nobu, M., & Liu, W. T. (2016). Effects of heat shocks on microbial community structure and microbial activity of a methanogenic enrichment degrading benzoate. *Letters in applied microbiology, 63*(5), 356-362. doi:https://doi.org/10.1111/lam.12629

Nikolaki, S., & Tsiamis, G. (2013). Microbial diversity in the era of omic technologies. *BioMed research international, 2013*. doi:http://dx.doi.org/10.1155/2013/958719

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., . . . Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research, 41*(D1), D590-D596. doi:https://doi.org/10.1093/nar/gks1219

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., . . . Robinson, C. J. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology, 75*(23), 7537-7541. doi:http://doi.org/10.1128/AEM.01541-09

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Söding, J. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology, 7*(1), 539. doi:https://doi.org/10.1038/msb.2011.75

Tian, Z., Zhang, Y., & Yang, M. (2018). Chronic impacts of oxytetracycline on mesophilic anaerobic digestion of excess sludge: inhibition of hydrolytic acidification and enrichment of antibiotic resistome. *Environmental Pollution, 238*, 1017-1026. doi:https://doi.org/10.1016/j.envpol.2018.02.023

Yang, B., Wang, Y., & Qian, P.-Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC bioinformatics, 17*(1), 135. doi:https://doi.org/10.1186/s12859-016-0992-y

Yang, S., Phan, H. V., Bustamante, H., Guo, W., Ngo, H. H., & Nghiem, L. D. (2017). Effects of shearing on biogas production and microbial community structure during anaerobic digestion with recuperative thickening. *Bioresource technology, 234*, 439-447. doi:https://doi.org/10.1016/j.biortech.2017.03.051

Youssef, N., Sheik, C. S., Krumholz, L. R., Najar, F. Z., Roe, B. A., & Elshahed, M. S. (2009). Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. *Applied and environmental microbiology, 75*(16), 5227-5236. doi:https://doi.org/10.1128/AEM.00592-09

The 30th Annual Meeting of the Thai Society for Biotechnology and International Conference
November 22-23, 2018 Ambassador Hotel, Bangkok, Thailand

MBB-P-10

SBI Metagenomics: An integrated database and visualization platform for metagenomic studies

Kiattiyot Laeman[1], Pichahpuk Uthaipaisonwong[1], Chalida Rangsiwutisak[1], Supapon Cheevadhanarak[1,2], and Kanthida Kusonmano[1,2,*]

[1]Systems Biology and Bioinformatics Research Laboratory, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bang Khun Thian, Bangkok, 10150, Thailand
[2]Bioinformatics and Systems Biology Program, School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bang Khun Thian, Bangkok, 10150, Thailand
*Email: kanthida.kus@mail.kmutt.ac.th

ABSTRACT

In the last decades, microbiome research has been growing rapidly as a consequence of dramatic improvement of next-generation sequencing technologies (NGS). The study, called metagenomics, allows us to study microbial communities by sequencing all genetic materials from a particular environment without culturing microorganisms in a laboratory. As a huge volume of such data has been produced and still increasing, several databases have been developed to collect and manage metagenomic samples including raw sequences, metadata and some processed results. These well-known databases include Sequence Read Archive (SRA), The European Nucleotide Archive (ENA), DNA Data Bank of Japan (DDBJ), and EBI metagenomics (or MGnify). Even though, these databases are claimed to be linked and exchanging the data, they still have different formats of metadata and data could be found in each database using different keywords of the sample properties. An integrated database that could synchronize all metagenomic data would facilitate scientists to search for datasets of interest without going through different databases. In this work, we construct a platform for compressive non-redundant metagenomic database and web-application called SBI Metagenomics (http://metagenomics.sbi.kmutt.ac.th). The platform integrates metagenomic data from the mentioned available repositories and provides a powerful access to the data via user-friendly web application for searching, visualizing and comparing metagenomic data associated with interested metadata. The platform would be useful for researchers who are interested in a particular filed of metagenomics and would like to revise existing data or compare to their own data to the available data.

KEYWORDS: Metagenomics; Microbiome Database, Web Applications; Data Visualization;

SBI Metagenomics:
An integrated database and visualization platform for metagenomic studies

Kiattiyot Laeman[1], Pichahpuk Uthaipaisanwong[1], Chalida Rangsiwutisak[1], Supapon Cheevadhanarak[1,2] and Kanthida Kusonmano[1,2*]

[1]Systems Biology and Bioinformatics Research Laboratory, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bang Khun Thian, Bangkok, 10150, Thailand

[2]Bioinformatics and Systems Biology Program, School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bang Khun Thian, Bangkok, 10150, Thailand

*Email: kanthida.kus@mail.kmutt.ac.th

www.sbi.kmutt.ac.th

Abstract

In the last decades, microbiome research has been growing rapidly as a consequence of dramatic improvement of next-generation sequencing technologies (NGS). The study, called metagenomics, allows us to study microbial communities by sequencing all genetic materials from a particular environment without culturing microorganisms in a laboratory. As a huge volume of such data has been produced and still increasing, several databases have been developed to collect and manage metagenomic samples including raw sequences, metadata and some processed results. These well-known databases include Sequence Read Archive (SRA), The European Nucleotide Archive (ENA), DNA Data Bank of Japan (DDBJ), and EBI metagenomics (or MGnify). Even though, these databases are claimed to be linked and exchanging the data, they still have different formats of metadata and data could be found in each database using different keywords of the sample properties. An integrated database that could synchronize all metagenomic data would facilitate scientists to search for datasets of interest without going through different databases. In this work, we construct a platform for compressive non-redundant metagenomic database and web-application called SBI Metagenomics (http://metagenomics.sbi.kmutt.ac.th). The platform integrates metagenomic data from the mentioned available repositories and provides a powerful access to the data via user-friendly web application for searching, visualizing and comparing metagenomic data associated with interested metadata. The platform would be useful for researchers who are interested in a particular filed of metagenomics and would like to revise existing data or compare to their own data to the available data.

Objective

To develop integrated metagenomic database and web application for searching, visualizing and comparing microbiome and asscociated metadata.

Conclusion

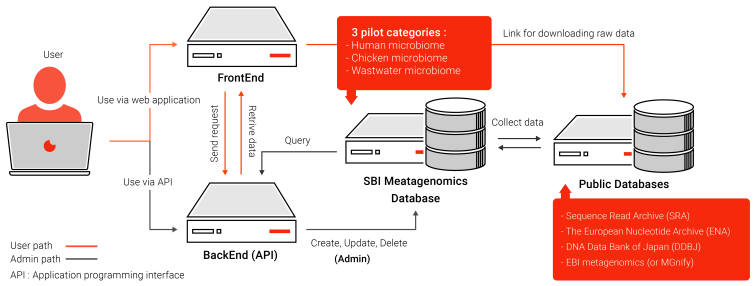
SBImetagenomics is an integrated platform for a comprehensive microbiome study. Currently, it provides 3 pilot microbiome categories, which are those of human, chicken and wastewater. It is a user-friendly platform that users can easily search, visualize and compare microbiome data with their asscociated metadata. More microbiome data will be added in the near future.

Acknowledgement

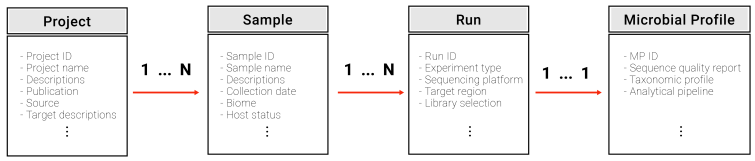
The authors acknowledge the financial support provided by Thailand Research Fund (MRG6080183), King Mongkut's University of Technology Thonburi through the "KMUTT 55th Anniversary Commemorative Fund" and "KMUTT Research Fund". We also thank Systems Biology and Bioinformatics Research Laboratory members for high performance computing support and fruitful disscussion.

Method

System architecture of SBImetagenomics



Data Model SBImetagenomics data contains 4 relational datatypes



***Note:** **1 ... 1** : One to one relationship e.g. one **Run** provides one **Microbial profile**
1 ... N : One to many relationship e.g. one **Project** has several **Samples**

Result

Sitemap of SBImetagenomics web application