# BIOLOGICAL DEGRADATION OF SOLUBLE MICROBIAL PRODUCTS IN A COMBINED SYSTEM OF ANAEROBIC PACKED-BED REACTORS AND A DOWN-FLOW HANGING SPONGE REACTOR

BY

NA KYUNG KIM

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Environmental Engineering in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Wen-Tso Liu, Chair
Professor Benito Mariñas
Assistant Professor Jeremy Guest
Assistant Professor Peiying Hong, King Abdullah University of Science and Technology

## ABSTRACT

Anaerobic biological processes are a reliable alternative to the conventional activated sludge process for the treatment of high-strength industrial wastewater, offering various advantages. Such advantages include, for example, less sludge generation, less operational cost, greater energy recovery, and a smaller footprint. An anaerobic up-flow packed-bed reactor maximizes the advantages by retaining a high concentration of biomass in the system, providing sufficient sludge retention time to slow growing anaerobic microorganisms. The inherent configuration of the reactor, however, is prone to increasing soluble microbial products (SMP). SMP are soluble organic cellular components that are released from biomass metabolisms in mixed culture biotechnology, which often result in a hindrance to efficient performance, lower effluent quality, and toxicity and a precursor of disinfectant by-products in discharged water. Despite several attempts to reduce SMP through coagulation and adsorption, a long-term treatment of SMP has not been achieved.

In this study, a combined process of anaerobic packed-bed reactors and a down-flow hanging sponge (DHS) reactor is proposed. As a matter of post-treatment, the DHS reactor further degraded SMP produced from the anaerobic methanogenic reactors, using selectively enriched microbial consortia-utilizing SMP. As such, the primary research aims of this project are as follows: (1) to understand the microbial community structure and ecology treating high-strength organic wastewater in the anaerobic packed-bed reactors; (2) to investigate biological SMP degradation in the DHS reactor; and (3) to explore phylogenetic characteristics and the metabolic functionality of the enriched microbial community involved in SMP degradation.

This study discussed the diversity and dynamics of microbial communities in anaerobic packed-bed reactors in the process of optimizing operational parameters. The communities were influenced by an increasing organic loading rate, which indicated a strong association with the abundance of *Bacteroidetes* and *Chloroflexi* among the dominant populations. These populations may take charge of initiating the degradation of organic compounds in the system. Next, the biological degradation of SMP, with respect to the selective enrichment of the microbial community in the DHS reactor, was demonstrated. SMP produced from the anaerobic reactors originated primarily from biomass metabolisms, exhibiting a bimodal MW distribution with 14-20 kDa and <4 kDa. The sub-fractions of SMP indicated different degradation fates in the DHS reactor with an overall stable removal (>70%) of the total SMP. Spatial and temporal variability

of the DHS microbial communities was significantly influenced by operational parameters. In particular, *Saprospiraceae* was the most correlated population in the community for increasing SMP loading, which indicated positive co-occurrences with neighboring bacterial populations. Different microbial diversity, along with the vertical depth of the reactor, suggested that stratified microbial communities might participate in the SMP degradation. Lastly, the genetic functional potential and expression of the DHS microbial community, with regard to SMP degradation, were explored. Despite the disparate microbial communities with the increase of SMP loading, a functional convergence for the SMP degradation was observed. The gene expression of the dominant draft genomes, based on carbohydrate-active enzymes, indicated that *Bacteroidetes*-related draft genomes actively represented cell associated enzyme-related genes, which were specific to the polysaccharide components of peptidoglycan. This finding led to speculation that the majority of SMP herein may be composed of detrital cell structural components released from peptidoglycan.

Ultimately, the findings from this study suggest a possible application of the biological SMP degradation, using a DHS reactor, to improve treatment performance and efficiency in bioprocesses. It also broadens current understanding of SMP, which are produced from mixed culture biotechnology, and their microbial utilization.

## ACKNOWLEDGEMENTS

The study presented here would not have been possible without the support of many people. First of all, I would like to thank my advisor, Dr. Wen-Tso Liu, for his academic guidance and support. His critical and enlightening advices on my research and presentations were nourishment that made this study improved. I also thank him for his continuous understanding and patient in the moments when I was a demanding student, exploring the new technologies and unknown fields.

I would like to express my gratitude to my committee members, Dr. Benito Mariñas, Dr. Peiying Hong, and Dr. Jeremy Guest, for their dedication to serving on my committee and for their insightful advices and encouragement that helped me improve this dissertation. Especially, I really appreciate Dr. Peiying Hong not only for her scientific and critical comments and suggestion, but also for the study opportunity and her help for me to adjust to the new cultural environment while I was at KAUST as a visiting student. I am grateful to Dr. Jenny D. Zadeh, Dr. Chris J. Fields, and Dr. Kathleen M. Keating at HPCBio for having valuable discussion with me and for helping me learn and fulfill the metagenomic and metatranscriptomic analyses and statistical differential analyses. I deeply appreciate Dr. Vernon L. Snoeyink for his devotion to help students grow and for sharing his great wisdom and pleasure.

Many thanks to past and present graduate students and post-docs in our department for giving me the privilege of working with over the years and creating enjoyable research environment. In particularly, I thank Dr. Ching-Sung Tsai and Weiwei Wu for helping me in learning molecular biology experiments and laboratory works, for having valuable discussions, and for being very sincere friends. I would like to thank Dr. Takashi Narihiro, Dr. Ye Lin, Dr. Seungdae Oh, Apeksha Tare, Dr. Chiachi Hwang, and Dr. Fangqiong Ling for helping me in learning a sequencing data analysis and for sharing their experiences and knowledge that represented a substantial part of my study.

Most importantly, I am deeply indebted to my parents, Bongju Kim and Munyong Kim, and my brother, Jintai Kim, for their constant support, understanding, and dedication that made me sustain life and this dissertation possible.

## TABLE OF CONTENTS

# CHAPTER 1: INTRODUCTION

## 1.1 Background

### 1.1.1 Biological anaerobic treatment of high strength organic wastewater

Anaerobic biological treatment for wastewater treatment is a promising alternative method to conventional treatments that use an aerobic activated sludge (AS) process. It holds potential due to its high capacity to degrade concentrated and recalcitrant substrates,[1-2] as well as low requirement of operational cost, small land requirements, and low excess sludge production.[1, 3-4] Additionally, it minimizes the energy requirement by producing biogas as a renewable energy, which can compensate for the electricity required for the operation.[5-8] The high-rate anaerobic treatment makes the process to be more appealing for high-strength organic wastewater treatment, such as food, soft drinks, and distillery industrial wastewater.[1, 9-14] The up-flow anaerobic sludge blanket (UASB) reactor was one of the robust anaerobic configurations in cases of organic overloads, providing favorable conditions for slow-growing anaerobic microorganisms to be well retained with a long sludge retention time (SRT).[15-16] Further by maximizing the density of biomass in the system with immobilized supporting media, an anaerobic packed-bed reactor was reported to provide greater efficiency, stability, and resilience than a UASB reactor.[17-18] Despite the advantages of the anaerobic treatment and the development of its configuration, the effluent of the anaerobic processes still contain residual organic matters and nutrients, which are not suitable to be discharged into the natural water body,[19-21] suggesting the need for a post-treatment system to further polish the effluent.

### 1.1.2 Application of a down-flow hanging sponge (DHS) reactor as a post treatment for anaerobic process

A down-flow hanging sponge (DHS) reactor was recently developed as a post-treatment for UASB processes.[22] The configuration of the DHS reactor is the same as a trickling filter reactor: wastewater is sprinkled over the tops of the filters, which trickles down to where biofilms are attached. Since air diffuses naturally through the highly porous polyurethane sponge filters, which are used as biofilm supporting media in the DHS reactor, high levels of dissolved oxygen throughout the reactor can be maintained without aeration.[23-24] The practical application of the DHS reactor has been studied intensively, exhibiting various advantages in terms of cost

and treatment efficiency. There is no need for external controls, such as pH and temperature, and it achieves a large capacity for biomass growth and a long sludge retention time (SRT).[22, 25-27] The combined system of an up-flow anaerobic reactors and a DHS reactor is a promising technology for the treatment of high strength organic wastewater.

### 1.1.3 Negative impacts of soluble microbial products on water and wastewater treatment systems

The residual organic matters in the effluent of the anaerobic process, mentioned above, were derived from the metabolism of biomass, known as soluble microbial products (SMP). Anaerobic reactors operated under a long SRT condition, such as an up-flow packed bed reactor, is prone to generate a high content of SMP. With an application of membrane separation in the anaerobic process, although the system enables the achievement of high-sludge concentration by decoupling a hydraulic retention time (HRT) from a SRT,[5, 28-29] SMP are also severely accumulated in the system, resulting in fouling on the membrane and deteriorating the quality of the effluent.[5, 30-32] Besides causing fouling in the membrane-based processes, SMP comprise a large portion of the remaining soluble chemical oxygen demand (SCOD) in effluents from conventional biological wastewater treatment processes.[33-34] SMP in discharge water from wastewater treatment systems alone cause toxicity as well as environmental hazards by acting as precursors of disinfection by-products.[35-36] Their accumulation in the system hinders efficient respiration, flocculation, and the settling ability of AS by deforming the physical properties of the AS.[37-38] In a nitrification process, SMP are one of the main causes inhibiting a nitrification efficiency.[39] Consequently, understanding the property of SMP and finding methods to control SMP production as well as their removal remain important for improving the performance of the anaerobic processes and the effluent quality. Ultimately, this aids in the achievement of gradually stricter discharge standards.[40]

### 1.1.4 Definition of SMP and their characteristics

SMP are soluble organic cellular components that are released from cell metabolism and lysis in bioprocesses.[41] Since SMP can be generated from any microbial activity, they are ubiquitous in bioprocesses and contain various complex mixtures of polysaccharides, proteins, lipids, humic and fulvic acids, extracellular enzymes, amino acids, DNA, and other cell structure

debris.[41-42] According to the unified theory of SMP proposed by Laspidou and Rittmann,[41] SMP are classified into two sub-groups: utilization-associated products (UAP) that are produced directly from substrate utilization, i.e., released metabolic intermediates, and biomass-associated products (BAP) that are formed from cell lysis and decay. Regarding the characterization of SMP, various extant studies measured the molecular weight (MW) distribution of SMP from various origins.[33-34, 43-44] Despite a wide range of MW distribution from different kinds and amounts of SMP, researchers commonly found that a majority showed a bimodal distribution with small MW, less than 1 kDa, and large MW, greater than 10-100 kDa. Moreover, a minor portion of SMP exhibited MWs between the two clusters.[33-34] In addition, Ni et al.[45] reported that the bimodal distribution could be related to the sub-fractions of the SMP. UAP were found to have low MWs, and they were readily utilized by the AS as substrates; whereas, BAP tended to have high MWs and accumulated in the reactor.

### 1.1.5 Parameters affecting SMP production

The production and accumulation of SMP in anaerobic processes are affected by various operational factors and the biodegradability of their sub-fractions. It has been reported that any kind of stress conditions on microbial activity led to increased SMP production; nutrient deficiency and toxic compounds considerably increased SMP concentrations in anaerobic chemostats.[46] Temporal organic shock loading, reduction of HRTs, and low pH also enhanced SMP formation. In particular, this includes accelerated cell lysis from shortened HRTs which increased the release of BAP.[47] Biomass treating highly saline substrates (over 30 g NaCl $l^{-1}$) produced more high-MW SMP that were difficult to degrade compared with those treating low-salinity substrates.[48] Decreasing temperature and a higher initial biomass concentration were reported as other factors enhancing SMP production in an anaerobic baffled reactor.[49] In addition, the long SRTs and short HRTs inherent in membrane bioreactors (MBR) are the most significant factors in the production and accumulation of SMP. Although there were some controversial results in the effects of SRT and HRT on the SMP production,[49-50] many previous studies have reported that SRTs longer than 10 days led to the accumulation of SMP, especially by increasing the BAP concentrations; whereas, the portion of UAP among the SMP decreased.[29, 34, 45, 51-55] Regarding these simultaneous effects, Huang et al.[51-52] concluded that

decreased HRTs and long SRTs in the submerged MBR, accelerated membrane fouling, causing a high SMP production.

### 1.1.6 Biological degradation as an alternative strategy to reduce SMP

Among SMP, the large BAP compounds tend to be accumulated as a semi-labile and refractory matter in a biological process.[43] Structural groups, such as carboxylic or phenolic groups, make cross-links with polyvalent cations, like $Ca^{2+}$ and $Mg^{2+}$, through acid-metal complexation, resulting in the formation of foulants in MBRs.[56-59] Various strategies were attempted in an effort to directly reduce SMP and to prevent formation of the foulants caused by SMP in bioprocesses, which included the adsorption and coagulation of SMP. It also included a chemical cleaning of polyvalent cations, using ion exchanges, salt cleaning, and metal chelating agents.[5-6, 60-65] Regardless, these chemical and physical attempts to remove SMP may not be a suitable long-term solution, providing just weeks to months of limited applications. Instead, the biological removal of SMP was considered recently as an alternative strategy to control SMP. However, the very low biodegradability of SMP was reported for both aerobic and anaerobic treatments. Between the two sub-fractions of SMP, BAP exhibited a slow biodegradation rate of 0.1 g COD/g VSS-d, whereas that of UAP reached to 1.8 g COD/g VSS-d, indicating that BAP tend to accumulate in the system while UAP might be readily degradable.[54, 66-68] Despite the tendency of the slow degradation of SMP, Backer et al.[49] suggested that SMP generated from anaerobic chemostats could be effectively removed in the following aerobically conditioned reactors. In their study, it was observed, specifically, that large MW SMP (> 10 kDa, >100 kDa, and >300 kDa), which were considered to be BAP, showed almost complete degradation (up to 96%) under aerobic conditions with enriched sludge for SMP uptakes.

### 1.1.7 Microorganisms involved in the degradation of SMP

Despite the low biodegradability of SMP, previous research reported that biological degradation of SMP was possible even with general activated sludge which was not acclimatized to specifically utilize the SMP as a substrate. This implies that a more effective degradation of SMP can be achieved with microbial consortia, preferentially, by utilizing microbial products. To identify SMP-degrading microorganisms and their community-level groups and to understand how they are involved in the degradation remain to be characterized. Few previous studies are

limited to identifying phylogenetic groups of heterotrophic bacteria, utilizing SMP that were produced by nitrifying bacteria at the phylum and class levels.[69-71] In particular, Okabe et al.[69] observed that *Chloroflexi* played an important role to utilize microbial products together with a *Cytophaga-Flavobacterium* cluster, *α-Proteobacteria,* and *γ-Proteobacteira* among the heterotrophs. These were coexisting with nitrifying autotrophs without an external carbon supply. As such, the *Chloroflexi* tended to uptake microbial products derived from biomass decay (BAP), rapidly degrading glucose and N-acetyl glucosamine (the main component of cell peptidoglycan layers), whereas the *Cytophaga-Flavobacterium* cluster gradually ingested both metabolic intermediates (UAP) and structural cell components (BAP) from the nitrifying bacteria. Okamura et al.[57, 72] isolated *Phialemonium curvatum* from AS for removing the uronic acids, which formed a matrix-like layer on the membranes, and evaluated the efficiency of preventing membrane fouling. Further, the decrease of SMP in the system was speculated to have a correlation with the abundance of *Klebsiella* in a biological activated carbon reactor[73] and *Chloroflexi* in a membrane bioreactor (MBR)[74]. Therefore, it is speculated that an abundance of SMP-degrading microbes might exist, a significant amount of which may not be usually found in the conventional bioprocesses of wastewater treatment. Information about them, such as their phylogenetic relationship and metabolic properties, remain to be revealed.

### 1.1.8 Application of high throughput sequencing to explore SMP degrading microorganisms

The gap of knowledge related to the microbial community structures and their metabolic functions involved in production and degradation of SMP may be addressed by using 16S ribosomal RNA (rRNA) gene, metagenomic, and metatranscriptomic sequencings, which are based on high throughput Next Generation Sequencing (NGS) (Figure 1.1).[75-78] First, these high throughput sequencing techniques, which are culture-independent, allow us to characterize unknown microbiomes that might rely on complex symbioses, representing in situ conditions of biological samples.[79-80] An amplicon sequencing of the 16S rRNA genes, which are highly conserved and used to differentiate among organisms of other species, enables us to reveal the phylogenetic microbial community composition of the complex microbiomes.[81-82] Metagenomics provides complementary characteristics of the community composition and information about the metabolic potential of entire communities and individual genomes in the biological niche.

This is accomplished by analyzing the entire genetic materials of the samples.[83-86] Further, metatranscriptomics enables a determination of the functional profile of the active populations in the microbial community.[85-87] Given the merits of these techniques, a multi-pronged approach for profiling genomic and expressional diversity and dynamics have been applied to various environmental and engineered microbiomes from marine water,[88-89] oil spill,[90] soil,[86, 91-93] and human[94-95] and animal[96-97] guts. Recently, the application of these integrated sequencing methods expanded to analyses of microbial communities found in biological wastewater treatment processes.[78, 98-99] The in-depth resolution of the genetic information, using these integrated sequencing approaches, would be helpful to enlighten characteristics of the microbial communities involved in the degradation of SMP in this study.

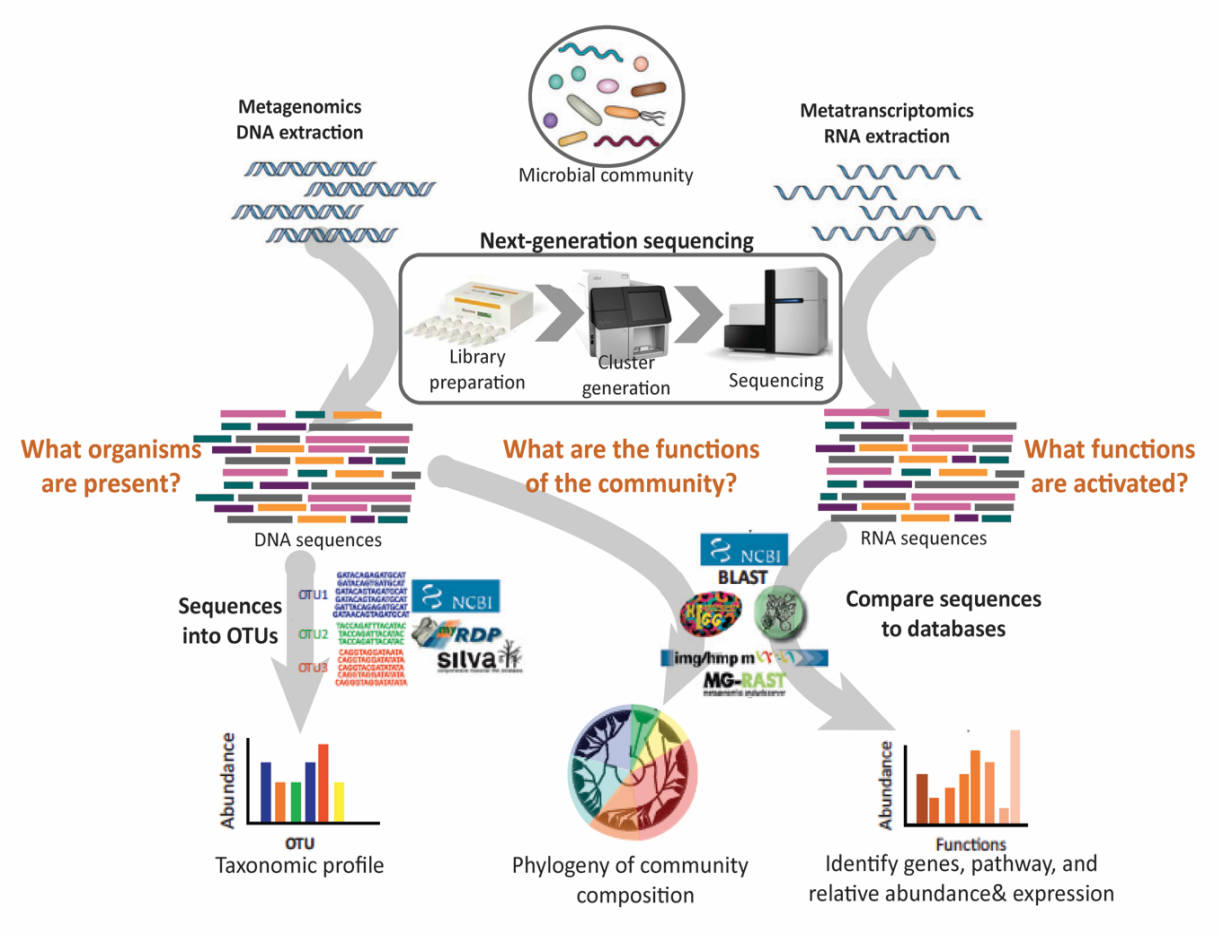



**Figure 1.1** Schematic representation of metagenomic and metatranscriptomic approaches to analyze an uncultured microbial community.

## 1.2 Objectives

Many previous studies concluded that despite various advantages of up-flow anaerobic packed-bed reactors to treat high strength wastewater, their inherent configurations, resulting in long SRT, are prone to the low efficiency of COD removal, which is mostly caused by accumulation of SMP. However, the SMP from anaerobic processes can be biologically degraded in a subsequent aerobic process. A more effective degradability of the SMP is expected when the microbial consortia have been acclimated to utilizing the SMP. Therefore, in this study, a combined system of anaerobic packed-bed reactors and a DHS reactor was applied in an effort to provide a long SRT for the anaerobic process. Also, it aimed to produce high quality effluent by reducing SMP in the effluent from the anaerobic reactors using the DHS reactor. This research expects to enhance the degradation of the SMP that are produced from the anaerobic reactors by enriching microbial consortia specifically utilizing the SMP. Significantly, the purpose of this study is to understand the anaerobic microbial communities treating high strength organic wastewater, characteristics of the SMP produced by them, the biological SMP degradation by the enriched microbial consortia in the DHS reactor, and their metabolic characteristics. To address this purpose, the specific objectives of the chapters are as follows:

1. To understand the microbial community structure and ecology treating high-strength organic wastewater in the anaerobic packed-bed reactors by investigating their temporal changes during the operation through 16S rRNA gene pyrosequencing.
2. To investigate the biological degradation of SMP produced from the anaerobic packed-bed reactors using selectively enriched microbial consortia in the DHS reactor. The spatial and temporal variability of the microbial community composition and structure was characterized using 16S rRNA gene pyrosequencing. The relationships between the microbial populations and the operational factors were identified and evaluated by applying network and redundancy analyses.
3. To explore metabolic potential and expression of the microbial consortia involved in SMP degradation and to disclose the active roles of the key microbial populations by analyzing overrepresented metabolic genes.

### 1.3 Experimental approach

A combined system of two up-flow anaerobic packed-bed reactors (an anaerobic packed-bed (AP) reactor and a hybrid packed-bed (HP) reactor) and a DHS reactor was configured to treat synthetic soft-drink-production wastewater containing polyethylene glycol (PEG), fructose, and glucose as the primary constituents (SCOD 3000 mg/L) (Figure 1.2). The anaerobic packed-bed reactors (7.6 L working volume) were filled with ceramic supporting media and operated at 35 $^{o}$C without regular discharge of biomass to provide a sufficient SRT for the mesophilic anaerobic microbiomes to proliferate. To optimize the operational factors, the organic loading rate (OLR) increased from 0.5 g SCOD/L/day to 2.0 g SCOD/L/day by decreasing the HRT, stepwise, for over 800 days. The DHS reactor (10L working volume), which was filled with polyurethane sponge media (porosity 0.985 vol./vol.), was fed with a combined effluent discharged from the anaerobic packed-bed reactors as the sole substrate. The OLR and HRT in the DHS reactor were adjusted as the HRT and the effluent organic concentration in the anaerobic reactors changed and divided into five phases. The reactor was maintained at room temperature without external adjustment, such as aeration and pH control. To determine the SMP contained in the effluent from the anaerobic reactors and degraded in the DHS reactor, SCOD removal and reduction propensity of SMP sub-fractions by analyzing the molecular weight (MW) distribution were investigated. Biomass samples for microbial community analysis were periodically collected from the anaerobic packed-bed reactors and the DHS reactor over the different operational phases. Separate biomass samplings from the DHS reactor at low and high OLR conditions were conducted for metabolic characterization involved in the SMP degradation.

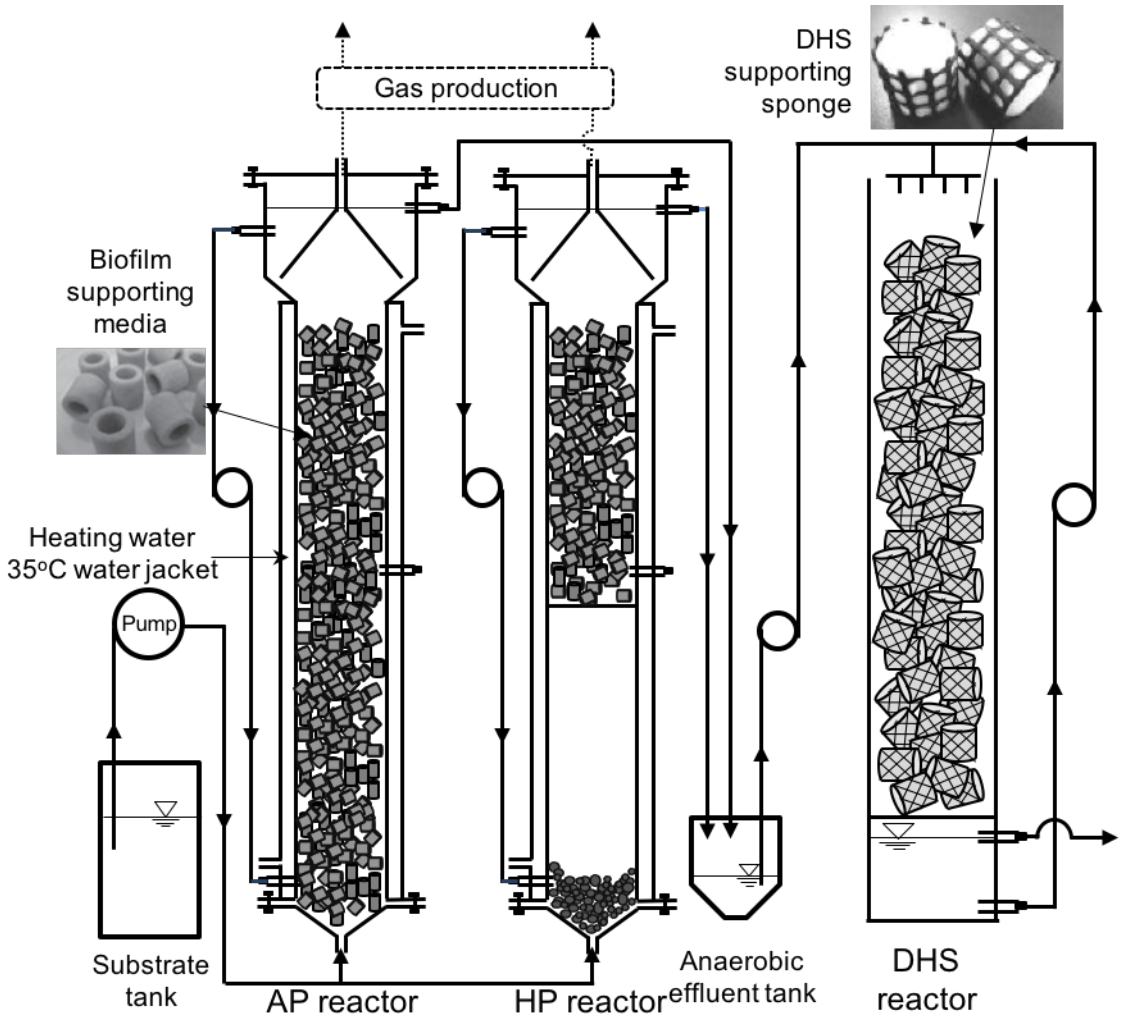



**Figure 1.2** Schematic diagram of the combined system of the anaerobic packed-bed reactors (AP and HP) and the DHS reactor used in this study.

### 1.4 Dissertation organization

In Chapter 2, titled "Microbial community analysis of anaerobic reactors treating soft drink wastewater," the methanogenic microbial communities in the AP and HP reactors, achieving >95% SCOD removal efficiency, were studied using 16S rRNA gene pyrosequencing. The diversity and dynamics of the microbial communities were correlated with respect to the optimized operational parameters. The results indicated that both AP and HP communities were predominated by *Bacteroidetes*, *Chloroflexi*, *Firmicutes*, and candidate phylum KSB3, which may degrade organic compounds in wastewater treatment processes. The community compositions were influenced by the increasing OLR, indicating a strong association with an abundance of *Bacteroidetes* and *Chloroflexi* among the dominant populations.

In Chapter 3, titled "Enrichment and characterization of microbial consortia degrading soluble microbial products discharged from anaerobic methanogenic bioreactors," biological degradation of SMP produced from the AP and HP reactors, using selectively enriched microbial community in the DHS reactor, were demonstrated. As the operational conditions were changed in the five phases for >800 days, a stable SMP removal between 68.9 to 87.5% was achieved. The size-exclusive chromatogram demonstrated that the SMP produced from the AP and HP reactors exhibited a bimodal MW distribution with 14-20 kDa and <4 kDa. The sub-fractions of SMP indicated different degradation fates in the DHS reactor. The enriched microbial communities were characterized using 16S rRNA gene pyrosequencing, and their spatial and temporal variability were correlated with operational parameters. The results indicated that a great shift in the dominant microbial populations was observed as increasing SMP loading. *Saprospiraceae* was the most correlated population to the loading increase, indicating positive co-occurrences with neighboring bacterial populations. Different microbial diversity at the different vertical depth of the reactor was observed, suggesting that stratified microbial communities might participate in the SMP degradation.

In Chapter 4, titled "Phylogenetic and functional characterization of the microbial community degrading soluble microbial products in a DHS reactor using a metagenomic and metatranscriptomic approaches," the genetic functional potential and expression of the microbial community in the DHS reactor, which were expected to be related to the mechanism of SMP degradation, were studied using metagenomic and metatranscriptomic sequencing analyses. The functional annotation based on SEED Subsystems exhibited that although the microbial community compositions became disparate as SMP loading, a functional convergence was observed for the SMP degradation, including amino acids and derivatives, carbohydrates, and protein metabolisms. The gene expression of the dominant draft genomes base on carbohydrate-active enzymes (CAZy) indicated that *Bacteroidetes*-related draft genomes actively represented cell associated enzyme-related genes, which were specific to polysaccharide components of peptidoglycan. This finding implies that the microbial communities, degrading SMP in the DHS reactor, were selectively enriched for the utilization of detrital cell structural components, which were released from peptidoglycan.

Chapter 5 summarizes the main findings and contributions of this research, and it proposes future works. The research evidenced in Chapter 2 and 3 was published. Moreover, the recent work demonstrated in Chapter 4 will be submitted in the near future.

### 1.5 References


1. Housley, J.; Zoutberg, G., Application of the 'biothane'wastewater treatment system in the soft drinks industry. *Water and Environment Journal* **1994,** *8* (3), 239-245.

2. Mata-Alvarez, J.; Mace, S.; Llabres, P., Anaerobic digestion of organic solid wastes. An overview of research achievements and perspectives. *Bioresource technology* **2000,** *74* (1), 3-16.

3. Blanc, F.; O'Shaughnessy, J.; Corr, S.; Smith, P. In *Treatment of soft drink bottling wastewater from bench-scale treatability to full-scale operation*, Proceedings of the... Industrial Waste Conference, Purdue University (USA), 1984.

4. Ng, K.-K.; Lin, C.-F.; Panchangam, S. C.; Hong, P.-K. A.; Yang, P.-Y., Reduced membrane fouling in a novel bio-entrapped membrane reactor for treatment of food and beverage processing wastewater. *Water research* **2011,** *45* (14), 4269-4278.

5. Liao, B.-Q.; Kraemer, J. T.; Bagley, D. M., Anaerobic Membrane Bioreactors: Applications and Research Directions. *Critical Reviews in Environmental Science and Technology* **2006,** *36* (6), 489-530.

6. Meng, F.; Chae, S.-R.; Drews, A.; Kraume, M.; Shin, H.-S.; Yang, F., Recent advances in membrane bioreactors (MBRs): Membrane fouling and membrane material. *Water Research* **2009,** *43* (6), 1489-1512.

7. Stuckey, D. C., *Anaerobic membrane reactors. In Fang, H., Environmental anaerobic technology: applications and new developments*. Imperial College Press: London, 2011.

8. Hu, A.; Stuckey, D., Treatment of Dilute Wastewaters Using a Novel Submerged Anaerobic Membrane Bioreactor. *Journal of Environmental Engineering* **2006,** *132* (2), 190-198.

9. Kalyuzhnyi, S. V.; Saucedo, J. V.; Martinez, J. R., The anaerobic treatment of soft drink wastewater in UASB and hybrid reactors. *Applied Biochemistry and Biotechnology* **1997,** *66* (3), 291-301.

10. Omil, F.; Garrido, J. M.; Arrojo, B.; Méndez, R., Anaerobic filter reactor performance for the treatment of complex dairy wastewater at industrial scale. *Water Research* **2003,** *37* (17), 4099-4108.

11. Tang, Y.-Q.; Fujimura, Y.; Shigematsu, T.; Morimura, S.; Kida, K., Anaerobic treatment performance and microbial population of thermophilic upflow anaerobic filter reactor treating awamori distillery wastewater. *Journal of Bioscience and Bioengineering* **2007,** *104* (4), 281-287.

12. Borja, R.; Banks, C., Kinetics of anaerobic digestion of soft drink wastewater in immobilized cell bioreactors. *Journal of chemical technology and biotechnology* **1994,** *60* (3), 327-334.

13. Peixoto, G.; Saavedra, N. K.; Varesche, M. B. A.; Zaiat, M., Hydrogen production from soft-drink wastewater in an upflow anaerobic packed-bed reactor. *international journal of hydrogen energy* **2011,** *36* (15), 8953-8966.

14. Kalyuzhnyi, S.; Saucedo, J. V.; Martinez, J. R., The anaerobic treatment of soft drink wastewater in UASB and hybrid reactors. *Applied biochemistry and biotechnology* **1997,** *66* (3), 291-301.

15. Lettinga, G.; Van Velsen, A.; Hobma, S. W.; De Zeeuw, W.; Klapwijk, A., Use of the upflow sludge blanket (USB) reactor concept for biological wastewater treatment, especially for anaerobic treatment. *Biotechnology and bioengineering* **1980,** *22* (4), 699-734.

16. Seghezzo, L.; Zeeman, G.; van Lier, J. B.; Hamelers, H.; Lettinga, G., A review: the anaerobic treatment of sewage in UASB and EGSB reactors. *Bioresource Technology* **1998,** *65* (3), 175-190.

17. Dupla, M.; Conte, T.; Bouvier, J.; Bernet, N.; Steyer, J.-P., Dynamic evaluation of a fixed bed anaerobic digestion process in response to organic overloads and toxicant shock loads. *Water Science and Technology* **2004,** *49* (1), 61-68.

18. Calli, B.; Mertoglu, B.; Roest, K.; Inanc, B., Comparison of long-term performances and final microbial compositions of anaerobic reactors treating landfill leachate. *Bioresource technology* **2006,** *97* (4), 641-647.

19. Louwe Kooijmans, J.; van Velsen, E. In *Application of the UASB process for the treatment of domestic sewage under sub-tropical conditions, the Cali case*, Anaerobic Treatment. A grown-up technology. Conference papers (Aquatech 1986), 1986; pp p423-436.

20. Draaijer, H.; Maas, J.; Schaapman, J.; Khan, A., Performance of the 5 MLD UASB reactor for sewage treatment at Kanpur, India. *Water Science and Technology* **1992,** *25* (7), 123-133.

21. Schellinkhout, A.; Collazos, C., Full-scale application of the UASB technology for sewage treatment. *Water Science and Technology* **1992,** *25* (7), 159-166.

22. Tandukar, M.; Machdar, I.; Uemura, S.; Ohashi, A.; Harada, H., Potential of a combination of UASB and DHS reactor as a novel sewage treatment system for developing countries: Long-term evaluation. *Journal of Environmental Engineering-Asce* **2006,** *132* (2), 166-172.

23. Tandukar, M.; Machdar , I.; Uemura, S.; Ohashi, A.; Harada, H., Potential of a Combination of UASB and DHS Reactor as a Novel Sewage Treatment System for Developing Countries: Long-Term Evaluation. *Journal of Environmental Engineering* **2006,** *132* (2), 166-172.

24. Tandukar, M.; Uemura, S.; Machdar, I.; Ohashi, A.; Harada, H., A low-cost municipal sewage treatment system with a combination of UASB and the" fourth-generation" downflow hanging sponge reactors. *Anaerobic Digestion X* **2005,** *52* (1), 323-329.

25. El-Tabl, A. S.; , R. W.; Younes, a. S. M.; , Downflow Hanging Sponge (DHS) Reactor as a Novel Post Treatment System for Municipal Wastewater. *Life Science Journal* **2013,** *10* (3), 409-414.

26. Tanaka, H.; Takahashi, M.; Yoneyama, Y.; Syutsubo, K.; Kato, K.; Nagano, A.; Yamaguchi, T.; Harada, H., Energy saving system with high effluent quality for municipal sewage treatment by UASB-DHS. *Water Science and Technology* **2012,** *66* (6), 1186-1194.

27. Uemura, S.; Suzuki, S.; Abe, K.; Kubota, K.; Yamaguchi, T.; Ohashi, A.; Takemura, Y.; Harada, H., Removal of organic substances and oxidation of ammonium nitrogen by a down-flow hanging sponge (DHS) reactor under high salinity conditions. *Bioresource Technology* **2010,** *101* (14), 5180-5185.

28. He, Y.; Xu, P.; Li, C.; Zhang, B., High-concentration food wastewater treatment by an anaerobic membrane bioreactor. *Water Research* **2005,** *39* (17), 4110-4118.

29. Skouteris, G.; Hermosilla, D.; López, P.; Negro, C.; Blanco, Á., Anaerobic membrane bioreactors for wastewater treatment: A review. *Chemical Engineering Journal* **2012,** *198–199* (0), 138-148.

30. Choo, K.-H.; Lee, C.-H., Membrane fouling mechanisms in the membrane-coupled anaerobic bioreactor. *Water Research* **1996,** *30* (8), 1771-1780.

31. Kang, I.-J.; Yoon, S.-H.; Lee, C.-H., Comparison of the filtration characteristics of organic and inorganic membranes in a membrane-coupled anaerobic bioreactor. *Water Research* **2002,** *36* (7), 1803-1813.

32. Teychene, B.; Guigui, C.; Cabassud, C.; Amy, G., Toward a better identification of foulant species in MBR processes. *Desalination* **2008,** *231* (1–3), 27-34.

33. Barker, D. J.; Mannucchi, G. A.; Salvi, S. M. L.; Stuckey, D. C., Characterisation of soluble residual chemical oxygen demand (COD) in anaerobic wastewater treatment effluents. *Water Research* **1999,** *33* (11), 2499-2510.

34. Jarusutthirak, C.; Amy, G., Role of soluble microbial products (SMP) in membrane fouling and flux decline. *Environmental Science & Technology* **2006,** *40* (3), 969-974.

35. Wei, Y. Y.; Liu, Y.; Zhang, Y.; Dai, R. H.; Liu, X. A.; Wu, J. J.; Zhang, Q. A., Influence of soluble microbial products (SMP) on wastewater disinfection byproducts: trihalomethanes and haloacetic acid species from the chlorination of SMP. *Environmental Science and Pollution Research* **2011,** *18* (1), 46-50.

36. Tang, H. L.; Chen, Y. C.; Regan, J. M.; Xie, Y. F. F., Disinfection by-product formation potentials in wastewater effluents and their reductions in a wastewater treatment plant. *Journal of Environmental Monitoring* **2012,** *14* (6), 1515-1522.

37. Reid, E.; Liu, X.; Judd, S. J., Effect of high salinity on activated sludge characteristics and membrane permeability in an immersed membrane bioreactor. *Journal of Membrane Science* **2006,** *283* (1–2), 164-171.

38. Chudoba, J., Inhibitory effect of refractory organic compounds produced by activated sludge micro-organisms on microbial activity and flocculation. *Water Research* **1985,** *19* (2), 197-200.

39. Ichihashi, O.; Satoh, H.; Mino, T., Effect of soluble microbial products on microbial metabolisms related to nutrient removal. *Water Research* **2006,** *40* (8), 1627-1633.

40. Azami, H.; Sarrafzadeh, M. H.; Mehrnia, M. R., Soluble microbial products (SMPs) release in activated sludge systems: a review. *Iranian Journal of Environmental Health Science & Engineering* **2012,** *9* (1), 30.

41. Laspidou, C. S.; Rittmann, B. E., A unified theory for extracellular polymeric substances, soluble microbial products, and active and inert biomass. *Water Research* **2002,** *36* (11), 2711-2720.

42. Liao, B. Q.; Bagley, D. M.; Kraemer, H. E.; Leppard, G. G.; Liss, S. N., A review of biofouling and its control in membrane separation bioreactors. *Water Environment Research* **2004,** *76* (5), 425-436.

43. Ni, B. J.; Zeng, R. J.; Fang, F.; Xie, W. M.; Sheng, G. P.; Yu, H. Q., Fractionating soluble microbial products in the activated sludge process. *Water Research* **2010,** *44* (7), 2292-2302.

44. Laabs, C. N.; Amy, G. L.; Jekel, M., Understanding the Size and Character of Fouling-Causing Substances from Effluent Organic Matter (EfOM) in Low-Pressure Membrane Filtration. *Environmental Science & Technology* **2006,** *40* (14), 4495-4499.

45. Ni, B. J.; Rittmann, B. E.; Fang, F.; Xu, J. A.; Yu, H. Q., Long-term formation of microbial products in a sequencing batch reactor. *Water Research* **2010,** *44* (13), 3787-3796.

46. Aquino, S. F.; Stuckey, D. C., Soluble microbial products formation in anaerobic chemostats in the presense of toxic compounds. *Water Research* **2004,** *38* (2), 255-266.

47. Aquino, S. F.; Stuckey, D. C., The effect of organic and hydraulic shock loads on the production of soluble microbial products in anaerobic digesters. *Water Environment Research* **2004,** *76* (7), 2628-2636.

48. Vyrides, I.; Stuckey, D. C., Effect of fluctuations in salinity on anaerobic biomass and production of soluble microbial products (SMPs). *Biodegradation* **2009,** *20* (2), 165-175.

49. Barker, D. J.; Salvi, S. M. L.; Langenhoff, A. A. M.; Stuckey, D. C., Soluble microbial products in ABR treating low-strength wastewater. *Journal of Environmental Engineering-Asce* **2000,** *126* (3), 239-249.

50. Ahn, Y. T.; Choi, Y. K.; Jeong, H. S.; Chae, S. R.; Shin, H. S., Modeling of extracellular polymeric substances and soluble microbial products production in a submerged membrane bioreactor at various SRTs. *Water Science and Technology* **2006,** *53* (7), 209-216.

51. Huang, Z.; Ong, S. L.; Ng, H. Y., Submerged anaerobic membrane bioreactor for low-strength wastewater treatment: Effect of HRT and SRT on treatment performance and membrane fouling. *Water Research* **2011,** *45* (2), 705-713.

52. Huang, G. T.; Jin, G.; Wu, J. H.; Liu, Y. D., Effects of glucose and phenol on soluble microbial products (SMP) in sequencing batch reactor systems. *International Biodeterioration & Biodegradation* **2008,** *62* (2), 104-108.

53. Jarusutthirak, C.; Amy, G., Understanding soluble microbial products (SMP) as a component of effluent organic matter (EfOM). *Water Research* **2007,** *41* (12), 2787-2793.

54. Jiang, T.; Myngheer, S.; De Pauw, D. J. W.; Spanjers, H.; Nopens, I.; Kennedy, M. D.; Amy, G.; Vanrolleghem, P. A., Modelling the production and degradation of soluble microbial products (SMP) in membrane bioreactors (MBR). *Water Research* **2008,** *42* (20), 4955-4964.

55. Dong, B.; Jiang, S. Y., Characteristics and behaviors of soluble microbial products in sequencing batch membrane bioreactors at various sludge retention times. *Desalination* **2009,** *243* (1-3), 240-250.

56. Costa, A. R.; de Pinho, M. N.; Elimelech, M., Mechanisms of colloidal natural organic matter fouling in ultrafiltration. *Journal of Membrane Science* **2006,** *281* (1-2), 716-725.

57. Okamura, D.; Mori, Y.; Hashimoto, T.; Hori, K., Identification of biofoulant of membrane bioreactors in soluble microbial products. *Water Research* **2009,** *43* (17), 4356-4362.

58. Lee, S.; Elimelech, M., Relating Organic Fouling of Reverse Osmosis Membranes to Intermolecular Adhesion Forces. *Environmental Science & Technology* **2006,** *40* (3), 980-987.

59. Metzger, U.; Le-Clech, P.; Stuetz, R. M.; Frimmel, F. H.; Chen, V., Characterisation of polymeric fouling in membrane bioreactors and the effect of different filtration modes. *Journal of Membrane Science* **2007,** *301* (1–2), 180-189.

60. Lee, S.; Elimelech, M., Salt cleaning of organic-fouled reverse osmosis membranes. *Water Research* **2007,** *41* (5), 1134-1142.

61. Park, H.; Choo, K. H.; Lee, C. H., Flux enhancement with powdered activated carbon addition in the membrane anaerobic bioreactor. *Separation Science and Technology* **1999,** *34* (14), 2781-2792.

62. Hu, A. Y.; Stuckey, D. C., Activated carbon addition to a submerged anaerobic membrane bioreactor: Effect on performance, transmembrane pressure, and flux. *Journal of Environmental Engineering-Asce* **2007,** *133* (1), 73-80.

63. Akram, A.; Stuckey, D. C., Flux and performance improvement in a submerged anaerobic membrane bioreactor (SAMBR) using powdered activated carbon (PAC). *Process Biochemistry* **2008,** *43* (1), 93-102.

64. Koseoglu, H.; Yigit, N. O.; Iversen, V.; Drews, A.; Kitis, M.; Lesjean, B.; Kraume, M., Effects of several different flux enhancing chemicals on filterability and fouling reduction of membrane bioreactor (MBR) mixed liquors. *Journal of Membrane Science* **2008,** *320* (1-2), 57-64.

65. Wu, B.; An, Y.; Li, Y.; Wong, F. S., Effect of adsorption/coagulation on membrane fouling in microfiltration process post-treating anaerobic digestion effluent. *Desalination* **2009,** *242* (1-3), 183-192.

66. Rittmann, B. E. a. M., P.L. , *Environmental Biotechnology: Principles and Applications*. McGraw-Hill: 2001.

67. Noguera, D. Soluble Microbial Products (SMP) Modeling in Biological Processes. MS thesis, Univ. of Illinois at Urbana-Champaign, Urbana, 1991.

68. Noguera, D.; Araki, N.; Rittmann, B., Soluble microbial products (SMP) in anaerobic chemostats. *Biotechnol Bioeng* **1994,** *44*, 1040 - 1047.

69. Okabe, S.; Kindaichi, T.; Ito, T., Fate of C-14-labeled microbial products derived from nitrifying bacteria in autotrophic nitrifying biofilms. *Appl. Environ. Microbiol.* **2005,** *71* (7), 3987-3994.

70. Kindaichi, T.; Ito, T.; Okabe, S., Ecophysiological interaction between nitrifying bacteria and heterotrophic bacteria in autotrophic nitrifying biofilms as determined by microautoradiography-fluorescence in situ hybridization. *Appl. Environ. Microbiol.* **2004,** *70* (3), 1641-1650.

71. Matsumoto, S.; Katoku, M.; Saeki, G.; Terada, A.; Aoi, Y.; Tsuneda, S.; Picioreanu, C.; Van Loosdrecht, M., Microbial community structure in autotrophic nitrifying granules characterized by experimental and simulation analyses. *Environmental microbiology* **2010,** *12* (1), 192-206.

72. Okamura, D.; Mori, Y.; Hashimoto, T.; Hori, K., Effects of Microbial Degradation of Biofoulants on Microfiltration Membrane Performance in a Membrane Bioreactor. *Environmental Science & Technology* **2010,** *44* (22), 8644-8648.

73. Cai, L.; Yu, K.; Yang, Y.; Chen, B.-w.; Li, X.-d.; Zhang, T., Metagenomic exploration reveals high levels of microbial arsenic metabolism genes in activated sludge and coastal sediments. *Applied Microbiology and Biotechnology* **2013,** *97* (21), 9579-9588.

74. Miura, Y.; Okabe, S., Quantification of cell specific uptake activity of microbial products by uncultured Chloroflexi by microautoradiography combined with fluorescence in situ hybridization. *Environmental Science & Technology* **2008,** *42* (19), 7380-7386.

75. Albertsen, M.; Hansen, L. B. S.; Saunders, A. M.; Nielsen, P. H.; Nielsen, K. L., A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. *Isme Journal* **2012,** *6* (6), 1094-1106.

76. Mason, O. U.; Hazen, T. C.; Borglin, S.; Chain, P. S. G.; Dubinsky, E. A.; Fortney, J. L.; Han, J.; Holman, H. Y. N.; Hultman, J.; Lamendella, R.; Mackelprang, R.; Malfatti, S.; Tom, L. M.; Tringe, S. G.; Woyke, T.; Zhou, J. H.; Rubin, E. M.; Jansson, J. K., Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *Isme Journal* **2012,** *6* (9), 1715-1727.

77. Ye, L.; Zhang, T.; Wang, T. T.; Fang, Z. W., Microbial Structures, Functions, and Metabolic Pathways in Wastewater Treatment Bioreactors Revealed Using High-Throughput Sequencing. *Environmental Science & Technology* **2012,** *46* (24), 13244-13252.

78. Yu, K.; Zhang, T., Metagenomic and Metatranscriptomic Analysis of Microbial Community Structure and Gene Expression of Activated Sludge. *Plos One* **2012,** *7* (5).

79. Daniel, R., The metagenomics of soil. *Nat Rev Microbiol* **2005,** *3* (6), 470-478.

80. Handelsman, J.; Rondon, M. R.; Brady, S. F.; Clardy, J.; Goodman, R. M., Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & biology* **1998,** *5* (10), R245-R249.

81. Ye, L.; Zhang, T., Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing. *Applied microbiology and biotechnology* **2013,** *97* (6), 2681-2690.

82. Lee, S.-H.; Kang, H.-J.; Lee, Y. H.; Lee, T. J.; Han, K.; Choi, Y.; Park, H.-D., Monitoring bacterial community structure and variability in time scale in full-scale anaerobic digesters. *Journal of Environmental Monitoring* **2012,** *14* (7), 1893-1905.

83. Wemheuer, B.; Wemheuer, F.; Daniel, R., RNA-based assessment of diversity and composition of active archaeal communities in the German Bight. *Archaea* **2012,** *2012*.

84. Baldrian, P.; Kolařík, M.; Štursová, M.; Kopecký, J.; Valášková, V.; Větrovský, T.; Žifčáková, L.; Šnajdr, J.; Rídl, J.; Vlček, Č., Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *The ISME journal* **2012,** *6* (2), 248-258.

85. Xia, Y.; Wang, Y.; Fang, H. H.; Jin, T.; Zhong, H.; Zhang, T., Thermophilic microbial cellulose decomposition and methanogenesis pathways recharacterized by metatranscriptomic and metagenomic analysis. *Sci Rep* **2014,** *4*, 6708.

86. Hultman, J.; Waldrop, M. P.; Mackelprang, R.; David, M. M.; McFarland, J.; Blazewicz, S. J.; Harden, J.; Turetsky, M. R.; McGuire, A. D.; Shah, M. B., Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* **2015**.

87. Stewart, F. J.; Ottesen, E. A.; DeLong, E. F., Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *The ISME journal* **2010,** *4* (7), 896-907.

88. Hu, Y.; Fu, C.; Huang, Y.; Yin, Y.; Cheng, G.; Lei, F.; Lu, N.; Li, J.; Ashforth, E. J.; Zhang, L., Novel lipolytic genes from the microbial metagenomic library of the South China Sea marine sediment. *FEMS microbiology ecology* **2010,** *72* (2), 228-237.

89. Shi, Y.; Tyson, G. W.; Eppley, J. M.; DeLong, E. F., Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *The ISME journal* **2011,** *5* (6), 999-1013.

90. Mason, O. U.; Hazen, T. C.; Borglin, S.; Chain, P. S.; Dubinsky, E. A.; Fortney, J. L.; Han, J.; Holman, H.-Y. N.; Hultman, J.; Lamendella, R., Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *The ISME journal* **2012,** *6* (9), 1715-1727.

91. Coolen, M. J.; Orsi, W. D., The transcriptional response of microbial communities in thawing Alaskan permafrost soils. *Frontiers in microbiology* **2015,** *6*.

92. Mackelprang, R.; Waldrop, M. P.; DeAngelis, K. M.; David, M. M.; Chavarria, K. L.; Blazewicz, S. J.; Rubin, E. M.; Jansson, J. K., Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* **2011,** *480* (7377), 368-371.

93. Cardenas, E.; Kranabetter, J. M.; Hope, G.; Maas, K. R.; Hallam, S.; Mohn, W. W., Forest harvesting reduces the soil metagenomic potential for biomass decomposition. *ISME J* **2015,** *9* (11), 2465-2476.

94. D'Argenio, V.; Casaburi, G.; Precone, V.; Salvatore, F., Comparative Metagenomic Analysis of Human Gut Microbiome Composition Using Two Different Bioinformatic Pipelines. *Biomed Research International* **2014**.

95. Brown, C. T.; Sharon, I.; Thomas, B. C.; Castelle, C. J.; Morowitz, M. J.; Banfield, J. F., Genome resolved analysis of a premature infant gut microbial community reveals a Varibaculum cambriense genome and a shift towards fermentation-based metabolism during the third week of life. *Microbiome* **2013,** *1* (1), 30-30.

96. Zhang, Y.; Zhao, F.; Deng, Y.; Zhao, Y.; Ren, H.-q., Metagenomic and metabolomic analysis of the toxic effects of trichloroacetamide-induced gut microbiome and urine metabolome perturbations in mice. *Journal of Proteome Research* **2015**.

97. He, S. M.; Ivanova, N.; Kirton, E.; Allgaier, M.; Bergin, C.; Scheffrahn, R. H.; Kyrpides, N. C.; Warnecke, F.; Tringe, S. G.; Hugenholtz, P., Comparative Metagenomic and

Metatranscriptomic Analysis of Hindgut Paunch Microbiota in Wood- and Dung-Feeding Higher Termites. *Plos One* **2013,** *8* (4).

98. Rodríguez, E.; García-Encina, P. A.; Stams, A. J.; Maphosa, F.; Sousa, D. Z., Meta-omics approaches to understand and improve wastewater treatment systems. *Reviews in Environmental Science and Bio/Technology* **2015,** *14* (3), 385-406.

99. Sales, C. M.; Lee, P. K., Resource recovery from wastewater: application of meta-omics to phosphorus and carbon management. *Current opinion in biotechnology* **2015,** *33*, 260-267.

# CHAPTER 2: MICROBIAL COMMUNITY ANALYSIS OF ANAEROBIC REACTORS TREATING SOFT DRINK WASTEWATER

## 2.1 Abstract

The AP and HP reactors containing methanogenic microbial consortia were applied to treat synthetic soft drink wastewater, which contains polyethylene glycol (PEG) and fructose as the primary constituents. The AP and HP reactors achieved high COD removal efficiency (>95%) after 80 and 33 days of the operation, respectively, and operated stably over 2 years. 16S rRNA gene pyrotag analyses on a total of 25 biofilm samples generated 98,057 reads, which were clustered into 2,882 operational taxonomic units (OTUs). Both AP and HP communities were predominated by *Bacteroidetes*, *Chloroflexi*, *Firmicutes*, and candidate phylum KSB3 that may degrade organic compound in wastewater treatment processes. Other OTUs related to uncharacterized *Geobacter* and *Spirochaetes* clades and candidate phylum GN04 were also detected at high abundance; however, their relationship to wastewater treatment has remained unclear. In particular, KSB3, GN04, *Bacteroidetes*, and *Chloroflexi* are consistently associated with the OLR increase to 1.5 g COD/L-d. Interestingly, KSB3 and GN04 dramatically decrease in both reactors after further OLR increase to 2.0 g COD/L-d. These results indicate that OLR strongly influences microbial community composition. This suggests that specific uncultivated taxa may take central roles in COD removal from soft drink wastewater depending on OLR.

## 2.2 Introduction

As the global consumption of soft drinks continues to grow, 687 billion liters in 2013, the global value reach 830 billion USD.[1] However, this incurs copious production (up to 2.0 trillion liters per year) and discharge of wastewater[2] containing high concentrations of sugar[3-5] and polyethylene glycol (PEG; $HO[CH_2\ CH_2\ O]_n\ H$), a detergent for bottle washing and equipment rinsing.[6] As such, the wastewater stream is characterized by high organic content with the COD ranging from 1.2 to 8.0 g/L and $BOD_5$ from 0.6 to 4.5 g/L,[3] and required to be treated to reduce COD to prevent the occurrence of contamination in the natural environment. Previous studies report physicochemical treatment, including reverse osmosis,[2] filtration,[2, 7] ion-exchange,[2, 7] and ozonation;[8] however, such approaches are relatively ineffective for removing soluble compounds (e.g., PEG and fructose) compared with biological methods.[5, 9-10] While aerobic biological

treatment systems have also been applied,[11-12] long HRT, high aeration requirement, extensive land requirement, high sludge production, and poor biomass settling are significant drawbacks.[13] Anaerobic biological treatment is a promising alternative due to its high capacity to degrade concentrated and recalcitrant substrates.[13-14] Several studies have successfully applied anaerobic bioprocesses to treat soft drink wastewater, including immobilized cell bioreactors,[15-16] UASB reactors,[13, 17] anaerobic filters,[18] and up-flow anaerobic pack-bed reactors.[19] Although these reactors achieved satisfactory COD removal, none of these studies report the microorganisms that facilitate degradation of the wastewater organic compounds. Without understanding of the microbial community structure and ecology, development of strategies to maintain and improve treatment efficiency and stability can be difficult. In the present study, we developed anaerobic bioreactors treating synthetic soft-drink-production wastewater and investigated the temporal change in microbial community structure during the operation through 16S rRNA gene pyrosequencing. Specifically, we identify organisms potentially related to reactor operational conditions.

## 2.3 Material and methods

### 2.3.1 Reactor operation

Two anaerobic up-flow bioreactors (7.6 L working volume) were operated separately at 35°C (Figure 2.1). The anaerobic packed-bed reactor (AP) and hybrid packed-bed reactor (HP) were filled with the Siporax ceramic media (LxDxH; 15x15x15mm) (Aquatic Eco Systems, Apopka, FL, USA) to fill 10.2% and 5.0% of their working volume, respectively. Seed sludge sample was taken from anaerobic digester at Urbana, IL, USA.

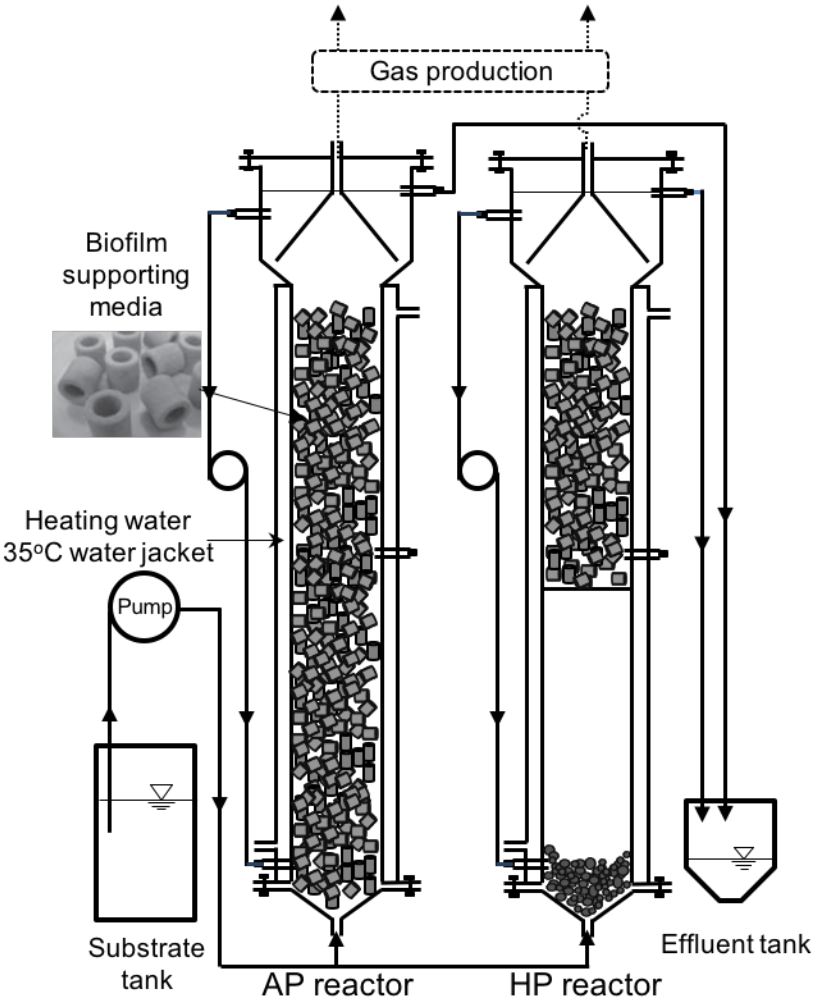



**Figure 2.1** Cross-section illustration of the AP and HP reactors. The reactors were equipped with water jacket and heated by water heater to kept at 35°C. The numbers in italics indicate size (mm).

The reactors were fed with 3,000 mg COD/L synthetic wastewater that mimicked the composition of wastewater discharged from soft drink-processing factory:[3, 5, 7, 15, 17, 20] 1,100 mg/L of polyethylene glycol 200 (PEG200); 1,500 mg/L of Corn Sweet High Fructose 55 (ADM, IL, USA); 30 mg/L of acetone; 30 mg/L of ethanol; 10 mg/L of silicone grease; 16 mg/L of $K_2HPO_4$; 19 mg/L of $FeSO_4.7H_2O$; 366 mg/L of $NaHCO_3$; 2 mg/L of NaF; 2.5 mg/L of NaOCl; and 28 mg/L of $NH_4HCO_3$. These components were dissolved in tap water, and pH was adjusted to 9.5–10.0 with 5M KOH to maintain the pH at 7.3–7.8 in the AP and HP. The internal circulation rates were 300 mL/min for both reactors. The reactors were operated under different HRT and organic loading rates (OLR) ranging from 1.5 to 6 days and from 0.5 to 2.0 g

SCOD/L/day, respectively (Figure 2.2). To avoid overloading of the organic compounds on initial microbial consortia, two reactors were operated for 11 days with constant recirculation of synthetic wastewater and no fresh influent. After day 11, both AP and HP reactors were fed with influent at a HRT of 6 days and an OLR of 0.5 g SCOD /L/day. For AP reactor, the HRT was decreased to 5, 4, and 3 days and the OLR gradually increased to 0.6, 0.75, and 1.0 g SCOD /L/day at 77, 91, and 115 days of the operation, respectively. For HP reactor, the HRT was decreased to 3 days and the OLR increased to 1.0 g SCOD /L/day after 31 days. After day 655, the HRT was decreased to 2 days and the OLR increased to 1.5 g SCOD /L/day for both reactors. Furthermore, the HRT of both reactors was decreased to 1.5 days and the OLR increased to 2.0 g SCOD /L/day after 744 days of the operation.

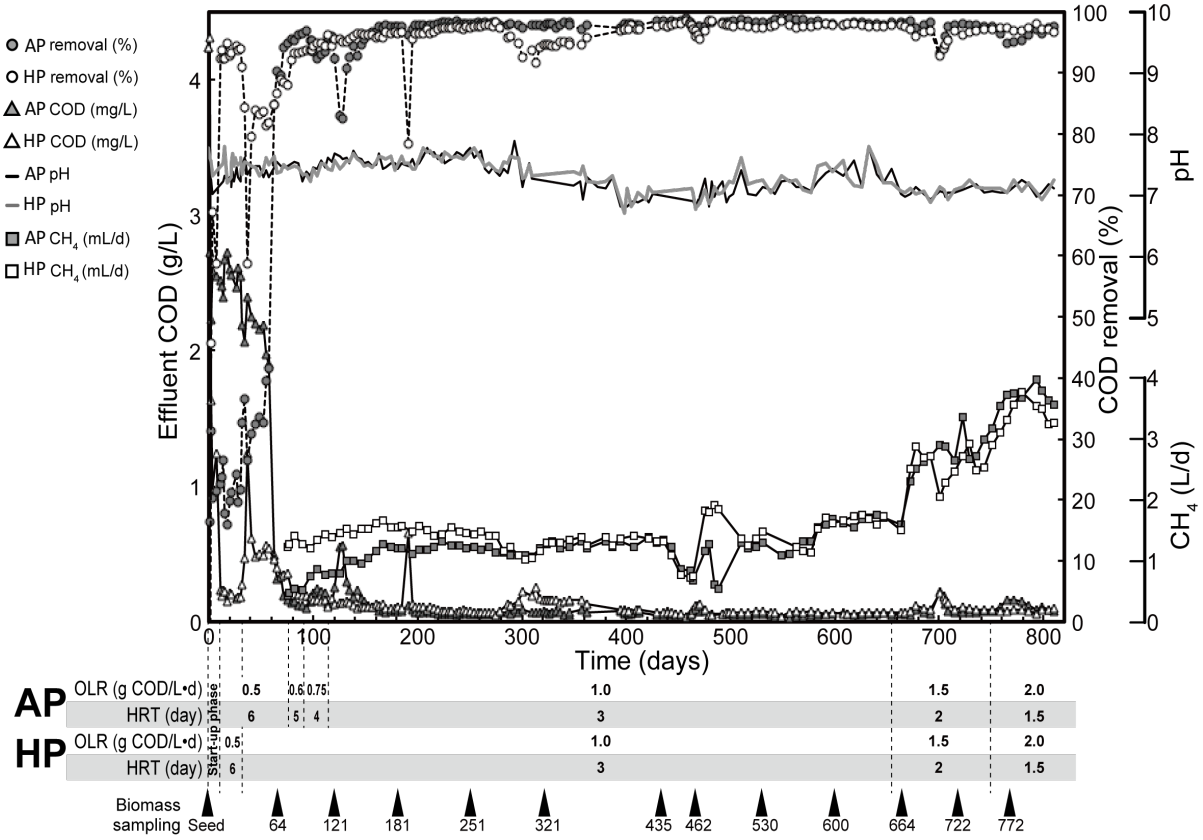



**Figure 2.2** Closed circle, COD removal (%) in AP; open circle, COD removal (%) in HP; closed triangle, effluent COD concentration (g/L) in AP; open triangle, effluent COD concentration (g/L) in HP; black line, pH in AP; gray line, pH in HP; closed square, methane gas production (L/day) in AP; open square, methane gas production (L/day) in HP. The reactors were operated at different OLR ranging from 0.5 to 2.0 g SCOD/L/day. The COD concentration of influent synthetic wastewater was decreased due to absence of polyethylene glycol 200 (1,100 mg/L) during days 398–411. The triangles in the bottom indicate the periods for biomass sampling from the reactors.

### 2.3.2 COD and methane gas measurements

The soluble COD was measured with COD digestion kit (HACH, Loveland, CO, USA) and DR/4000 U Spectrophotometer (HACH) according to the Standard Method 5220D.[21] Methane gas produced from the reactors was collected in gas sampling bag (Standard Tedlar PVF Bags, DuPont, DE, USA) and measured using a GC-2014 Gas Chromatograph (Shimadzu Scientific Instruments, Kyoto, Japan) equipped with a thermal conductivity detector (Shimadzu Scientific Instruments) and a Molecular Sieve 13X packed column (2,000xg 2 mm) (Restek, PA, USA).

### 2.3.3 Biomass sampling

Biomass samples for microbial community analysis were collected from AP and HP at 64, 121, 181, 251, 321, 435, 462, 530, 600, 664, 722, and 772 days of the operation (Figure 2.2). The ceramic media (ca. 5 pieces) were collected from 16 cm depth from effluent outlet with autoclaved forceps and put into 50-mL tube. After 10 mL of 1Å~ PBS was added, the media was vortexed rigorously to remove the biofilm. After centrifugation (8,500 xg, 3 min), the biomass samples were collected and stored in − 80°C freezer until DNA extraction.

### 2.3.4 DNA extraction, PCR, and pyrosequencing

DNA extraction, PCR, and pyrosequencing were performed as previously described.[22] Briefly, DNA was extracted using the FastDNA SPIN Kit for Soil (MP Biomedicals, Carlsbad, CA, USA). The 16S rRNA gene was amplified with the U515F forward primer and U909R reverse primer.[23] Pyrosequencing was performed using the GS-FLX Titanium platform (Roche/454 Life Sciences, Branford, CT, USA) at the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign (IL, USA).

### 2.3.5 Pyrosequencing data analysis

Raw 16S rRNA gene sequences were screened and trimmed with QIIME 1.8.0[24] using a sequence length (≥150 nt) and quality score (≥25) cut-off. The trimmed sequence data was clustered with the UCLUST algorithm using ≥97% sequence identity cut-off[25]. Representative sequences of each OTU were aligned using PyNAST[26] and chimeric sequences were removed using ChimeraSlayer.[27] The phylogenetic assignment of each OTU was carried out with a dataset

obtained from Greengenes_web_site (gg_13_5_otus; http://greengenes.secondgenome.com/ ).[28] The Chao1 index and rarefaction curve were calculated by EstimateS (version 9.1.0).[29] The coverage values were calculated using equation [1 – (n/N )], where n is the number of OTUs in a single read (singleton) and N is the total number of reads analyzed.[30] The weighted UniFrac distances were used for principal coordinate analysis (PCoA).[31] Phylogenetic trees for 16S rRNA gene pytotags and previously reported sequences were constructed with the ARB program based on the neighbor-joining algorithm.[32] Insertion of pyrotag sequences (ca. 370 bp) was performed with the parsimony insertion tool of the ARB program. The topology of the trees was estimated by 1,000 bootstrap replicates.[33]

### 2.3.6 Statistical analysis

In order to correlate microbial community profiles with reactor operational conditions (ORL, HRT and reactor type), statistical analysis including redundancy analysis (RDA) and correspondence analysis (CA) were performed using CANOCO software version 4.5 (Microcomputer Power, Ithaca, NY, USA).[34] According to the instruction of CANOCO, when the longest length lies between 3 and 4, it is reasonable to apply either linear method (RDA) or unimodal method (CA). All OTUs were used for calculation and major groups were picked out manually and plotted with operation conditions.

### 2.3.7 Nucleotide sequences accession number

The pyrosequencing data obtained in this study have been deposited under DDBJ/EMBL/GenBank accession no. DRA002423.

## 2.4 Results and discussion

### 2.4.1 Reactor operation

The operational performance of anaerobic packed-bed (AP) and hybrid packed-bed (HP) reactors treating synthetic soft drink wastewater is shown in Figure 2.2 and Table 2.1. AP and HP were continuously operated for more than 800 days. The removal efficiency of COD consistently maintained at 93–97% with an effluent COD mostly below 100 mg/L after 77 and 12 days of operation of AP and HP, respectively. After the days of operation, no apparent differences in performance were observed between AP and HP. The total volume of methane

increased gradually with an increase in OLR (Figure 2.2). The average values of pH were stable during the operation, implying no obvious acid accumulation in the reactors. These results indicated that enriched microbial consortia in AP and HP retain the stability against the feeding of synthetic soft drink wastewater at 2.0 g SCOD/L/day. Dark gray-black-colored biofilm was formed on the surface of ceramic media in both reactors. The biofilm samples were retrieved and used for microbial community analysis (Figure 2.2).

**Table 2.1** Operational parameter of AP and HP reactors.

| Parameters | AP | | | | | | | HP | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Day | 0-11 | 12-76 | 77-90 | 91-114 | 115-654 | 655-743 | 744-810 | 0-11 | 12-30 | 31-654 | 655-743 | 744-810 |
| HRT (day) | Batch | 6 | 5 | 4 | 3 | 2 | 1.5 | Batch | 6 | 3 | 2 | 1.5 |
| OLR (g SCOD/L/d) | Batch | 0.5 | 0.6 | 0.75 | 1.0 | 1.5 | 2.0 | Batch | 0.5 | 1.0 | 1.5 | 2.0 |
| COD removal (%) | 22.4±5.5 | 41.3± 27.5 | 95.9± 0.6 | 94.5± 1.6 | 97.4± 2.3 | 97.2± 1.6 | 96.4± 1.0 | 66.0± 19.7 | 93.5± 0.9 | 95.1± 5.0 | 96.2± 1.4 | 97.2± 0.3 |
| Methane (L/day) | N.D. | N.D. | 514.4± 32.5 | 793.6± 53.9 | 1216.1± 242.8 | 2657.1± 424.6 | 3642.2± 218.4 | N.D. | N.D. | 1402.5± 237.3 | 2478.1± 391.7 | 3352.2 ± 260.2 |
| pH | 7.3±0.3 | 7.4±0.1 | 7.4±0.2 | 7.5±0.1 | 7.3±0.3 | 7.0±0.1 | 7.1±0.1 | 7.6±0.2 | 7.5±0.2 | 7.4±0.3 | 7.1±0.1 | 7.1±0.1 |

AP, anaerobic packed-bed reactor; HP, hybrid packed-bed reactor; COD: chemical oxygen demand; HRT, hydraulic retention time; OLR, organic loading rate; N.D. not determine

### 2.4.2 Overview of 16S rRNA gene pyrosequencing

16S rRNA gene pyrotag libraries were constructed for twelve AP and HP biofilm samples each and their seed sludge. A total of 98,057 16S rRNA gene pyrotag reads were retrieved and further classified into 2,882 OTUs using a 97% sequence identity cut-off (Table A.1). Although the rarefaction curves of most samples were insufficient to achieve the plateau (Figure A.1), the high Good's coverage values (>93%) suggested that obtained OTUs adequately estimated the microbial diversity of the reactors. According to the Chao1 indexes, the biofilm may contain approximately 1.53– 2.23-fold more OTUs than detected. Comparing microbial community composition between samples, unweighted UniFrac-based PCoA clearly showed that the community composition varied with time (Figure 2.3). Specifically, the microbial constituents continuously change over 321 days and reached stable structure only after 462 days, based on Jackknife clustering analysis, weighted UniFrac-based PCoA and correspondence analysis (CA) (Figure A.2, A.3, and A.4). Despite the dynamic community structure, the steady COD removal indicates that the enriched microbial consortia at all stages were suitable for soft drink wastewater treatment at the respective operation conditions (Figure 2.2). Using OTU-level phylogenetic analyses, we identify dominant organisms (Figure 2.4) and discuss their potential ecological roles below.

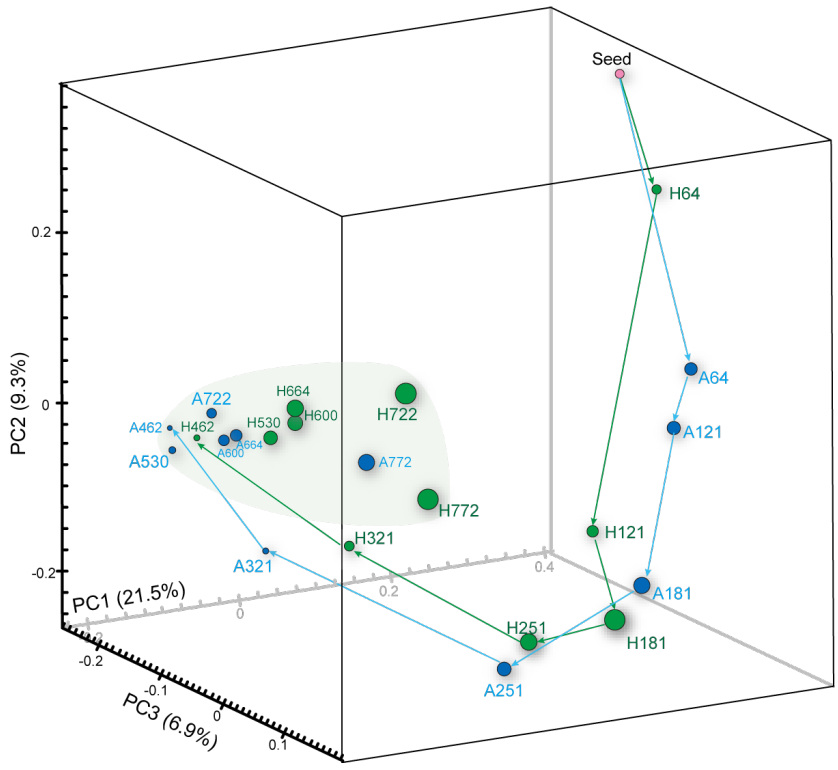



**Figure 2.3** PCoA based on the abundances of 16S rRNA gene OTUs (unweighted UniFrac). For this analysis, observed 16S rRNA gene OTUs were normalized to 1,400 reads per sample. A and H indicate the samples taken from the AP and HP reactors. The numbers following A and H indicate days of the operation for biomass sampling.

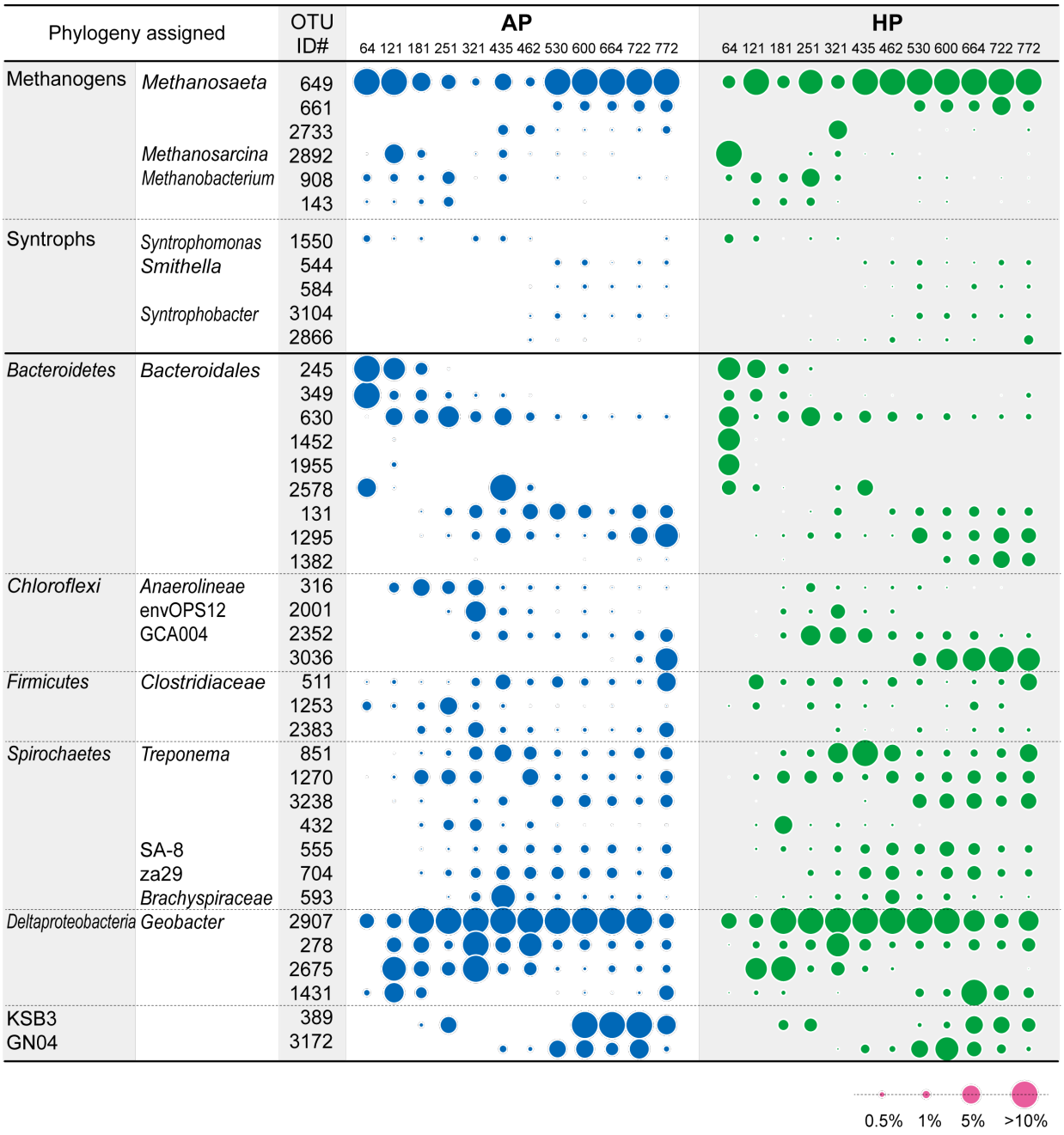



**Figure 2.4** The numbers below AP and HP on top row indicate days of the operation for biomass sampling.

### 2.4.3 Bacteroidetes, Chloroflexi, Firmicutes, and Spirochaetes

Phyla thought to take part in the anaerobic digestion nexus,[35-39] *Bacteroidetes*, *Chloroflexi* , *Firmicutes*, and *Spirochaetes* , were detected in all samples (Table A.2 and Figure 2.4).

Firmicutes family *Clostridiaceae* (OTU1253, 2383, and 2853) were found in seed sludge and consistently observed throughout operation. On the other hand, although only two abundant OTUs (349 and 2758) were found in seed sludge within the phylum *Bacteroidetes*, other seven major OTUs emerged during the operation and their abundances behaved differently over time: OTUs 245, 349, 630, 1452, and 1955 predominated before day 435 and decreased in the later stages while OTUs 131, 1295, and 1382 increased after 321 days. Despite no dominant *Chloroflexi*-related OTUs in seed sludge, the abundances of three *Chloroflexi*-type OTUs (OTU316, 2001, and 2352) were frequently detected at day 121-435 and decreased after day 530, while OTU3036 predominated in later stage. Based on redundancy analysis (RDA) to correlate the abundance of major OTUs with operational conditions (Figure A.5), HRT, OLR, and reactor type were the major explanatory variables; further, this RDA plot supported the fluctuation of the discussed *Bacteroidetes*, *Chloroflexi*, and *Firmicutes* OTUs (Figure A.5). The members of the phyla may be responsible for fermentative degradation of protein and, more importantly, sugar to VFAs, based on previous reports.[40-42] In addition, Bacteroidetes found in the reactor may perform PEG degradation as a Bacteroidetes member, Bacteroides sp. PG1, has been observed to degrade PEG1000 axenically or in co-culture with *Methanobacterium* sp. DG1.[43] While *Spirochaetes* is neither known to degrade sugars nor PEG, related OTUs (555, 704, 851, 1270, and 3238) were consistently observed after 121 days (Figure 2.4 and Figure A.5), indicating that relatively high OLR condition (> 1.0 g SCOD L/day) facilitated their proliferation in the reactors. Although studies have reported *Spirochaetes* populations performing syntrophic acetate oxidation[44] and acetogenesis[45] in methanogenic environments, their ecological function still remains unclear.

### 2.4.4 Candidate phyla KSB3 and GN04

Besides such phyla widely associated with anaerobic digestion, we also observed populations of candidate phyla KSB3 and GN04 during later stages of operation (Figure 2.4). After 600 days, KSB3 (OTU389) predominated up to 38.3% and 4.8% in AP and HP respectively. This KSB3 closely relates to a clone (99.2% similarity to clone SwB25fl, accession no. AB266941) associated with a mesophilic UASB reactors treating sugar-containing wastewater (Figure 2.5).[35] Further, KSB3 was also previously observed to degrade carbohydrates (i.e., glucose and maltose), especially in association with increase in influent sugar

concentration.[46-47] Thus, KSB3 likely participates in fructose degradation in both AP and HP reactors. The GN04-related OTU3172 was detected in the AP (2.6-5.6%) and HP (1.5-8.1%) reactors after 530 days operation (Figure 2.4). Like KSB3, this GN04 OTU is related to a lineage (specifically MSB-5A5) associated with mesophilic UASB reactors treating sugar-containing wastewater (e .g ., 99.5% identity with clone N2B95fl; accession no. AB266976) (Figure 2.6).[35] However, in both cases, their physiology and in situ functions remains largely unknown. The RDA plot indicated that GN04 and KSB3 populations are positioned close to the origin of the axes, indicating that their appearance could not be explained by the environmental factors tested. Further study on metagenomic and single-cell genomic analyses would provide more useful information to elucidate the ecophysiological traits of these functionally unknown microbes.

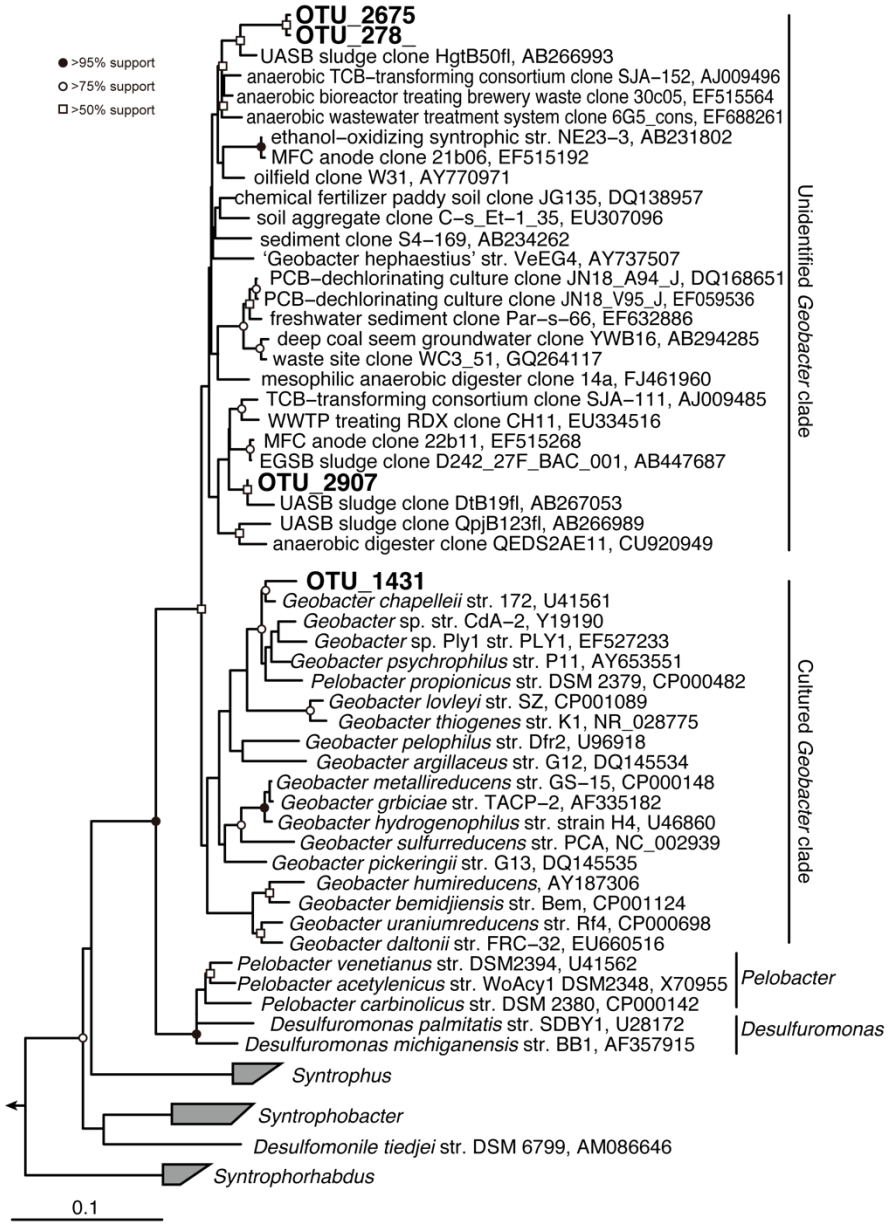



**Figure 2.5** Distance matrix tree of 16S rRNA gene sequences assigned to the candidate phyla GN04 and KSB3 retrieved from anaerobic reactors based on the neighbor-joining method. Boldface indicates the sequences obtained in this study. The 16S rRNA gene sequences of *Methanosaeta harundinacea* 8Ac (AY817738), *Methanosaeta pelagica* 03d30q (AB679167), *Methanosaeta concilii opfikon* (X51423) were used as outgroup. The bar indicates 10% base substitution. Branching points supported probabilities >95%, >75%, and >50% by bootstrap analyses (based on 1,000 replicates) are indicated by solid circle, open circles, and open square, respectively.

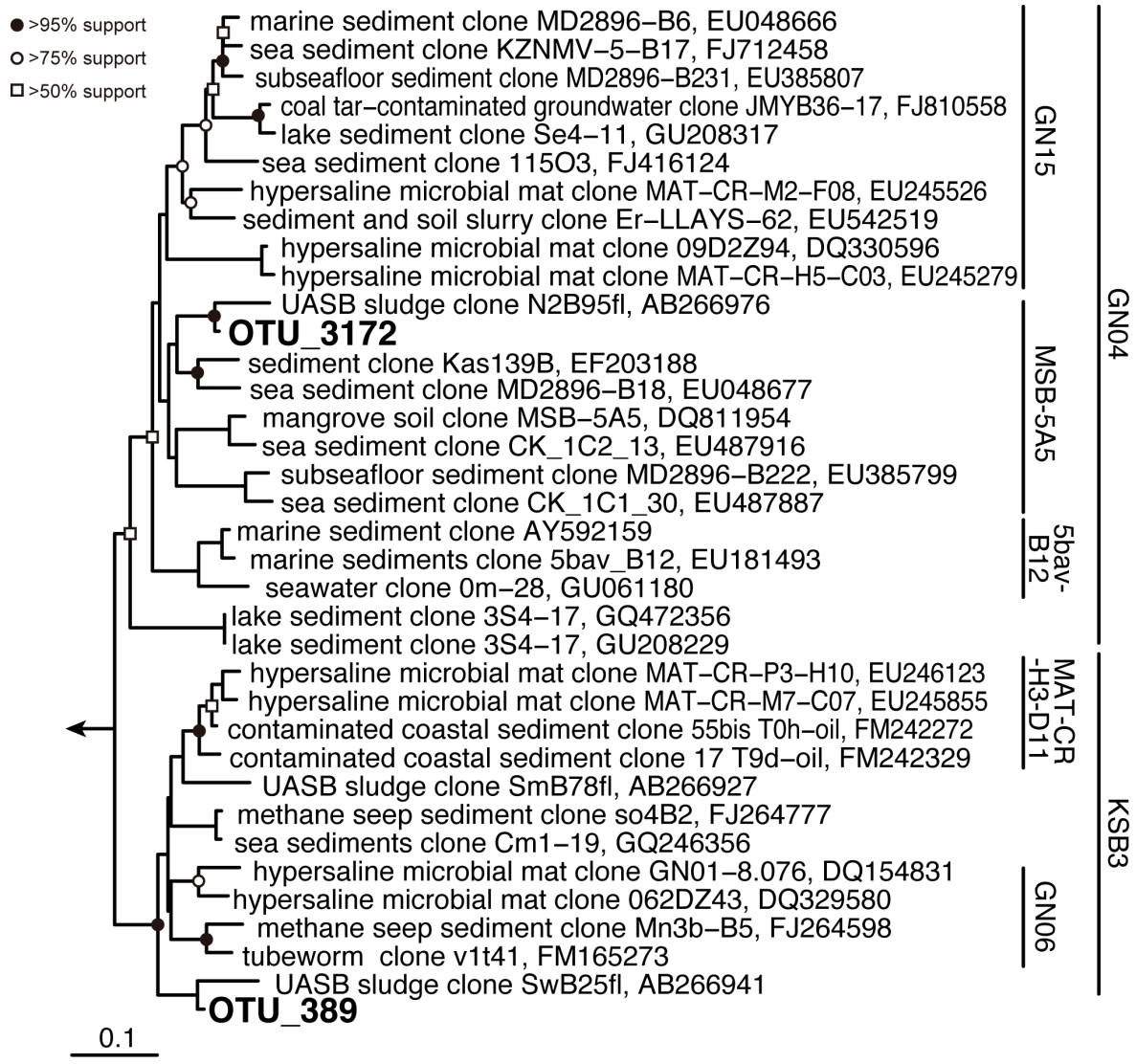



**Figure 2.6** Distance matrix tree of 16S rRNA gene sequences assigned to the *Geobacter* retrieved from anaerobic reactors based on the neighbor- joining method. Boldface indicates the sequences obtained in this study. The 16S rRNA gene sequences of *Thermodesulfobacterium commune* DSM 2178 (AF418169), *Thermodesulfobacterium hveragerdense* DSM 12571 (NR_029311), and *Thermodesulfobacterium hydrogeniphilum* DSM 14290 (NR_025146) were used as outgroup. The bar indicates 10% base substitution. Branching points supported probabilities >95%, >75%, and >50% by bootstrap analyses (based on 1,000 replicates) are indicated by solid circle, open circles, and open square, respectively.

### 2.4.5 Methanogens and syntrophs

In order to accomplish complete conversion of sugar to $CH_4$ and $CO_2$, it is necessary to further degrade $H_2$, acetate, and other volatile fatty acids (VFAs; e.g., propionate and butyrate) likely generated from sugar fermentation by the aforementioned organisms. Specific methanogen

clades are known to individually degrade H2 and acetate to CH4 and CO2. On the other hand, degradation of VFAs is thermodynamically limited in methanogenic environments,[48-50] and syntrophs and methanogens are known to form obligate mutualistic metabolic interactions to accomplish such degradation. As expected, OTUs associated with known methanogens and syntrophs were consistently observed in AP and HP during operation (Figure 2.4). For methanogens, *Methanobacterium* (OTU143 and 908) was the dominant H2 -oxidizing methanogen throughout reactor operation. Similarly, aceticlastic *Methanosaeta*-related OTU649 was found not only in all sludge samples (1.0– 27.1% of the total population) but also in seed sludge (3.1%), likely degrading acetate derived from fructose and/or PEG.[51-52] An OTU (2892) related to *Methanosarcina*, capable of both acetate- and H2 -oxidation, was detected at relatively higher abundances at day 121 in AP (5.5%) and day 64 in HP (12.4%). RDA revealed that *Methanosarcina*- and *Methanobacterium*-related OTUs (OTU143, 908, and 2892) were represented by relatively short arrows in the direction of HRT, indicating their proliferation at higher HRT conditions. For *Methanosaeta* populations, OTU649 had no significant correlation with HRT and OLR. In contrast, the OTU661 was strongly correlated with OLR. For *Methanosaeta* populations, OTU649 had no significant correlation with HRT and OLR. It has been reported that the affinity for acetate could be relevant to the growth of aceticlastic methanogens, and under high acetate concentrations, *Methanosarcina* spp. often outcompete *Methanosaeta* spp.[53-54] While the acetate concentration was not measured in the reactor, it was likely very low due to the dilution of substrate concentration from internal circulation and reactor volume right after entering the reactor. Even in such low acetate concentration, *Methanosaeta* related OTU661 might be affected by different OLR conditions. As for degradation of VFAs, in both reactors, we found known syntrophic populations, including *Syntrophomonas* (OTU1550), *Syntrophobacter* (OTU2866 and 3104), and *Smithella* (OTU544 and 584) (Figure 2.4). Among them, *Syntrophomonas*-related OTU1550 was found in seed sludge as a major syntrophic population (0,44%). Based on characteristics of these genera,[48, 55] they are most likely involved in the degradation of butyrate (*Syntrophomonas*) and propionate (*Syntrophobacter* and *Smithella*) through with syntrophic partnership with methanogens (e.g., *Methanobacterium*). Such VFAs may be produced by butyrate- or propionate producing fermentative bacteria, such as the members of the phyla Firmicutes and Bacteroidetes.[56-60] The relatively low abundances of syntrophic bacteria (< 1.6% of the total populations) are in good accordance with the results of

quantitative analyses of anaerobic bioreactors with membrane hybridization[61] and sequence-specific 16S rRNA cleavage method.[62] These results suggest that hydrogenotrophic methanogens and syntrophs observed here might play a supporting role in the VFA removal to maintain process stability. RDA plot of known syntrophs showed that the OTUs associated with propionate-oxidizing syntrophs (OTU544 and 584, 2866, and 3104) shared similar trend going along with OLR axis (Figure A.5A). Given that these microbes utilize propionate as major substrate for syntrophic metabolism,[48] it is reasonable to conclude that propionate fermentation might be the dominant sugar degradation pathway as OLR increased. *Syntrophomonas*-related OTU1550 that primarily utilizes butyrate, showed opposite trending with propionate oxidizers, implying a major role of butyrate fermentation in lower OLR condition.

### 2.4.6 Geobacter

Unlike most other methanogenic environments, *Geobacter*-related organisms were frequently observed in the AP and HP reactor pyrotag libraries, although they were minor populations in seed sludge (< 0.31%) (Figure 2.4). OTU1431 closely related to *G. chapelleii* strain 172 (99.5% sequence identity; accession no. U41561), a non-fermentative, iron-reducing bacterium capable of oxidizing acetate, formate, ethanol, and lactate (Figure 2.5).[63] RDA indicated that OLR correlated with the abundance of the OTU1431 (Figure A.5), suggesting that *G. chapelleii*-related organism might contribute to oxidizing acid (i .e ., formate, acetate, and lactate) or alcohol (i .e ., ethanol) possibly produced by fermentative degradation of sugar and PEG. Three other OTUs (278, 2675, and 2907) were distantly related to known *Geobacter* isolates (i.e., OTU278 has 98.0% identity with *G. argillaceus* strain G12; accession no. NR_043575, and OTU2675 and 2907 have 99.0% identity with *G. daltonii* strain FRC-32; accession no. NR_074916), and clustered with environmental clones that retrieved from mesophilic UASB reactors treating wastewater discharged from sugar- and amino acid-processing factories (Figure 2.6).[35] These observations suggested the importance of these *Geobacter*-related organisms in anaerobic processes treating food-processing wastewater. Within this poorly characterized Geobacter clade, 16S rRNA gene sequence of a syntrophic ethanol-oxidizing bacterium NE23-3 (accession no. AB231802) was deposited. Albeit no report on its physiology has yet been published, such unidentified Geobacter may oxidize ethanol (and possibly other syntrophic substrates) in association with hydrogenotrophic methanogens. RDA

plot showed that these OTUs had no correlations with OLR/HRT. It is puzzling that *Geobacter* predominated the reactor community despite no substantial addition of oxidized metals (e .g ., $Fe^{3+}$ and $Mn^{4+}$ ). However, recent studies suggest that *Geobacter* may thrive under methanogenic conditions through interspecies electron transfer with methanogens.[64-65] In short, while we suspect they ought play an important role in the treatment of soft drink wastewater based on their consistent presence, more studies are necessary to investigate their ecological contribution.

## 2.5 Conclusions

We successfully operated AP and HP reactors to treat synthetic soft drink wastewater. Based on the 16S rRNA gene pyrotag analyses, we identified core microbial constituents and assigned their possible function based on previously known physiological characteristics: *Methanosaeta*, *Methanosarcia*, and *Methanobacterium* as major methanogenic archaea; *Bacteroidetes*, *Chloroflexi*, *Firmicutes*, and KSB3 as fermentative bacteria; *Bacteroidetes* as PEG degrader. *Syntrophs*, *Syntrophomonas*, *Syntrophobacter*, and *Smithella* may support degradation of VFAs derived from sugar and PEG degradation by the fermenters. While we also identify *Geobacter*, *Spirochaetes*, and GN04 members prevalent in the reactor, their ecological role in soft drink wastewater treatment remains unclear. Interestingly, many of these organisms, especially KSB3 and GN04, appear to be strongly influenced by operational conditions, indicating that specific organisms may be adapted to and responsible for sugar/PEG degradation under specific conditions.

## 2.6 Acknowledgements

We are grateful to Huijie Lu, Lin Ye, and Priscilia Tunas for technical assistance. We also thank Chris L. Wright at the Roy J. Carver Biotechnology Center, the University of Illinois at Urbana- Champaign for pyrosequencing analysis.

## 2.7 References

1. Euromonitor International. Passport: Global Market Information Database. .

2. Haroon, H.; Waseem, A.; Mahmood, Q., Treatment and reuse of wastewater from beverage industry. *J Chem Soc Pak* **2013,** *35*, 5-10.

3. Chen JP, S. S., Hung YT. Soft Drink Waste Treatment. In: Wang LK, Hung Y-T, Lo HH, Yapijakis C, editors, *Waste Treatment in the Food Processing Industry*. CRC Press: Florida, 2006.

4. Hanover, L. M.; White, J. S., Manufacturing, composition, and applications of fructose. *The American journal of clinical nutrition* **1993,** *58* (5), 724S-732S.

5. Hsine, E.; Benhammou, A.; Pons, M., Water resources management in soft drink industry-water use and wastewater generation. *Environmental technology* **2005,** *26* (12), 1309-1316.

6. Hemm, D.; Hellmann, G.; Wilbert, K., Compact cleaning agent for industrial dish washing machines. Google Patents: 2001.

7. Camperos, E. R.; Nacheva, P. M.; Díaz Tapia, E., Treatment techniques for the recycling of bottle washing water in the soft drinks industry. *Water Science & Technology* **2004,** *50* (2), 107-112.

8. Garcia-Morales, M.; Roa-Morales, G.; Barrera-Diaz, C.; Balderas-Hernandez, P., Treatment of soft drink process wastewater by ozonation, ozonation-H2O2 and ozonation-coagulation processes. *Journal of Environmental Science and Health, Part A* **2012,** *47* (1), 22-30.

9. Hsine, E. A.; Benhammou, A.; Marie-NoeUe, P., Design of a beverage Industry wastewater treatment facility using process simulation. *Computer Applications in Biotechnology 2004* **2005**, 299–302.

10. Tebai, L.; Hadjivassilis, I., Soft drinks industry wastewater treatment. *Water Science and Technology* **1992,** *25* (1), 45-51.

11. Blanc, F.; O'Shaughnessy, J.; Corr, S.; Smith, P. In *Treatment of soft drink bottling wastewater from bench-scale treatability to full-scale operation*, Proceedings of the... Industrial Waste Conference, Purdue University (USA), 1984.

12. Ng, K.-K.; Lin, C.-F.; Panchangam, S. C.; Hong, P.-K. A.; Yang, P.-Y., Reduced membrane fouling in a novel bio-entrapped membrane reactor for treatment of food and beverage processing wastewater. *Water research* **2011,** *45* (14), 4269-4278.

13. Housley, J.; Zoutberg, G., Application of the ‘biothane’wastewater treatment system in the soft drinks industry. *Water and Environment Journal* **1994,** *8* (3), 239-245.

14. Mata-Alvarez, J.; Mace, S.; Llabres, P., Anaerobic digestion of organic solid wastes. An overview of research achievements and perspectives. *Bioresource technology* **2000,** *74* (1), 3-16.

15. Borja, R.; Banks, C., Kinetics of anaerobic digestion of soft drink wastewater in immobilized cell bioreactors. *Journal of chemical technology and biotechnology* **1994,** *60* (3), 327-334.

16. Borja, R.; Banks, C., Semicontinuous anaerobic digestion of soft drink wastewater in immobilised cell bioreactors. *Biotechnology letters* **1993,** *15* (7), 767-772.

17. Kalyuzhnyi, S.; Saucedo, J. V.; Martinez, J. R., The anaerobic treatment of soft drink wastewater in UASB and hybrid reactors. *Applied biochemistry and biotechnology* **1997,** *66* (3), 291-301.

18. Harper, S. R.; Pohland, F. G., Microbial consortia selection in anaerobic filters operated in different reactor configurations. *Water Science and Technology* **1997,** *36* (6), 33-39.

19. Peixoto, G.; Saavedra, N. K.; Varesche, M. B. A.; Zaiat, M., Hydrogen production from soft-drink wastewater in an upflow anaerobic packed-bed reactor. *international journal of hydrogen energy* **2011,** *36* (15), 8953-8966.

20. Miyaki, H.; Adachi, S.; Suda, K.; Kojima, Y., Water recycling by floating media filtration and nanofiltration at a soft drink factory. *Desalination* **2000,** *131* (1), 47-53.

21. *Chemical Oxygen Demand (COD)*; US Environmental Protection Agency: 1997.

22. Tamaki, H.; Wright, C. L.; Li, X.; Lin, Q.; Hwang, C.; Wang, S.; Thimmapuram, J.; Kamagata, Y.; Liu, W.-T., Analysis of 16S rRNA amplicon sequencing options on the Roche/454 next-generation titanium sequencing platform. *PLoS One* **2011,** *6* (9), e25263.

23. Wang, Y.; Qian, P.-Y., Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PloS one* **2009,** *4* (10), e7401.

24. Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Pena, A. G.; Goodrich, J. K.; Gordon, J. I., QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **2010,** *7* (5), 335-336.

25. Edgar, R. C., Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010,** *26* (19), 2460-2461.

26. Caporaso, J. G.; Bittinger, K.; Bushman, F. D.; DeSantis, T. Z.; Andersen, G. L.; Knight, R., PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **2010,** *26* (2), 266-267.

27. Haas, B. J.; Gevers, D.; Earl, A. M.; Feldgarden, M.; Ward, D. V.; Giannoukos, G.; Ciulla, D.; Tabbaa, D.; Highlander, S. K.; Sodergren, E., Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome research* **2011,** *21* (3), 494-504.

28. McDonald, D.; Price, M. N.; Goodrich, J.; Nawrocki, E. P.; DeSantis, T. Z.; Probst, A.; Andersen, G. L.; Knight, R.; Hugenholtz, P., An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal* **2012,** *6* (3), 610-618.

29. EstimateS, C. R. EstimateS: Statistical estimation of species richness and shared species from samples. Statistical estimation of species richness and shared species from samples. Version 9 and earlier. User's Guide and application. [WWW document]. http://purl.oclc.org/estimates.

30. Good, I. J., The population frequencies of species and the estimation of population parameters. *Biometrika* **1953,** *40* (3-4), 237-264.

31. Hamady, M.; Lozupone, C.; Knight, R., Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *The ISME journal* **2010,** *4* (1), 17-27.

32. Saitou, N.; Nei, M., The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **1987,** *4* (4), 406-425.

33. Felsenstein, J., Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **1985**, 783-791.

34. Lepš, J.; Šmilauer, P., *Multivariate analysis of ecological data using CANOCO*. Cambridge university press: 2003.

35. Narihiro, T.; Terada, T.; Kikuchi, K.; Iguchi, A.; Ikeda, M.; Yamauchi, T.; Shiraishi, K.; Kamagata, Y.; Nakamura, K.; Sekiguchi, Y., Comparative analysis of bacterial and archaeal communities in methanogenic sludge granules from upflow anaerobic sludge blanket reactors treating various food-processing, high-strength organic wastewaters. *Microbes and Environments* **2009,** *24* (2), 88-96.

36. Sekiguchi, Y.; Kamagata, Y., Microbial community structure and functions in methane fermentation technology for wastewater treatment. *Strict and Facultative Anaerobes: Medical and Environmental Aspects* **2004**, 361-384.

37. Riviere, D.; Desvignes, V.; Pelletier, E.; Chaussonnerie, S.; Guermazi, S.; Weissenbach, J.; Li, T.; Camacho, P.; Sghir, A., Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. *The ISME journal* **2009,** *3* (6), 700-714.

38. Nelson, M. C.; Morrison, M.; Yu, Z., A meta-analysis of the microbial diversity observed in anaerobic digesters. *Bioresource technology* **2011,** *102* (4), 3730-3739.

39. Narihiro, T.; Nobu, M. K.; Kim, N. K.; Kamagata, Y.; Liu, W. T., The nexus of syntrophy-associated microbiota in anaerobic digestion revealed by long-term enrichment and community survey. *Environmental microbiology* **2015,** *17* (5), 1707-1720.

40. Kampmann, K.; Ratering, S.; Kramer, I.; Schmidt, M.; Zerr, W.; Schnell, S., Unexpected stability of Bacteroidetes and Firmicutes communities in laboratory biogas reactors fed with different defined substrates. *Applied and environmental microbiology* **2012,** *78* (7), 2106-2119.

41. Ito, T.; Yoshiguchi, K.; Ariesyady, H. D.; Okabe, S., Identification and quantification of key microbial trophic groups of methanogenic glucose degradation in an anaerobic digester sludge. *Bioresource technology* **2012,** *123*, 599-607.

42. Ariesyady, H. D.; Ito, T.; Okabe, S., Functional bacterial and archaeal community structures of major trophic groups in a full-scale anaerobic sludge digester. *Water research* **2007,** *41* (7), 1554-1568.

43. Dwyer, D. F.; Tiedje, J. M., Metabolism of polyethylene glycol by two anaerobic bacteria, Desulfovibrio desulfuricans and a Bacteroides sp. *Applied and environmental microbiology* **1986,** *52* (4), 852-856.

44. Lee, S.-H.; Park, J.-H.; Kang, H.-J.; Lee, Y. H.; Lee, T. J.; Park, H.-D., Distribution and abundance of Spirochaetes in full-scale anaerobic digesters. *Bioresource technology* **2013,** *145*, 25-32.

45. Graber, J. R.; Breznak, J. A., Physiology and nutrition of Treponema primitia, an H2/CO2-acetogenic spirochete from termite hindguts. *Applied and environmental microbiology* **2004,** *70* (3), 1307-1314.

46. Yamada, T.; Yamauchi, T.; Shiraishi, K.; Hugenholtz, P.; Ohashi, A.; Harada, H.; Kamagata, Y.; Nakamura, K.; Sekiguchi, Y., Characterization of filamentous bacteria, belonging to candidate phylum KSB3, that are associated with bulking in methanogenic granular sludges. *The ISME journal* **2007,** *1* (3), 246-255.

47. Yamada, T.; Kikuchi, K.; Yamauchi, T.; Shiraishi, K.; Ito, T.; Okabe, S.; Hiraishi, A.; Ohashi, A.; Harada, H.; Kamagata, Y., Ecophysiology of uncultured filamentous anaerobes belonging to the phylum KSB3 that cause bulking in methanogenic granular sludge. *Applied and environmental microbiology* **2011,** *77* (6), 2081-2087.

48. Schink, B.; Stams, A. J., *Syntrophism among prokaryotes*. Springer: 2013.

49. Kato, S.; Watanabe, K., Ecological and evolutionary interactions in syntrophic methanogenic consortia. *Microbes and Environments* **2010,** *25* (3), 145-151.

50. Hattori, S., Syntrophic acetate-oxidizing microbes in methanogenic environments. *Microbes and Environments* **2008,** *23* (2), 118-127.

51. Schink, B.; Stieb, M., Fermentative degradation of polyethylene glycol by a strictly anaerobic, gram-negative, nonsporeforming bacterium, Pelobacter venetianus sp. nov. *Applied and Environmental Microbiology* **1983,** *45* (6), 1905-1913.

52. Dwyer, D. F.; Tiedje, J. M., Degradation of ethylene glycol and polyethylene glycols by methanogenic consortia. *Applied and environmental microbiology* **1983,** *46* (1), 185-190.

53. Min, H.; Zinder, S. H., Kinetics of acetate utilization by two thermophilic acetotrophic methanogens: Methanosarcina sp. strain CALS-1 and Methanothrix sp. strain CALS-1. *Applied and Environmental Microbiology* **1989,** *55* (2), 488-491.

54. Grady Jr, C. L.; Daigger, G. T.; Love, N. G.; Filipe, C. D., *Biological wastewater treatment*. CRC press: 2011.

55. Sieber, J. R.; McInerney, M. J.; Gunsalus, R. P., Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. *Annual review of microbiology* **2012,** *66*, 429-452.

56. Eeckhaut, V.; Van Immerseel, F.; Pasmans, F.; De Brandt, E.; Haesebrouck, F.; Ducatelle, R.; Vandamme, P., Anaerostipes butyraticus sp. nov., an anaerobic, butyrate-producing bacterium from Clostridium cluster XIVa isolated from broiler chicken caecal content, and emended description of the genus Anaerostipes. *International journal of systematic and evolutionary microbiology* **2010,** *60* (5), 1108-1112.

57. De Vuyst, L.; Leroy, F., Cross-feeding between bifidobacteria and butyrate-producing colon bacteria explains bifdobacterial competitiveness, butyrate production, and gas production. *International journal of food microbiology* **2011,** *149* (1), 73-80.

58. Qiu, Y.-L.; Kuang, X.-Z.; Shi, X.-S.; Yuan, X.-Z.; Guo, R.-B., Paludibacter jiangxiensis sp. nov., a strictly anaerobic, propionate-producing bacterium isolated from rice paddy field. *Archives of microbiology* **2014,** *196* (3), 149-155.

59. Ueki, A.; Akasaka, H.; Suzuki, D.; Ueki, K., Paludibacter propionicigenes gen. nov., sp. nov., a novel strictly anaerobic, Gram-negative, propionate-producing bacterium isolated from plant residue in irrigated rice-field soil in Japan. *International journal of systematic and evolutionary microbiology* **2006,** *56* (1), 39-44.

60. Chen, S.; Dong, X., Proteiniphilum acetatigenes gen. nov., sp. nov., from a UASB reactor treating brewery wastewater. *International journal of systematic and evolutionary microbiology* **2005,** *55* (6), 2257-2261.

61. Zheng, D.; Angenent, L.; Raskin, L., Monitoring granule formation in anaerobic upflow bioreactors using oligonucleotide hybridization probes. *Biotechnology and bioengineering* **2006,** *94* (3), 458-472.

62. Narihiro, T.; Terada, T.; Ohashi, A.; Kamagata, Y.; Nakamura, K.; Sekiguchi, Y., Quantitative detection of previously characterized syntrophic bacteria in anaerobic wastewater treatment systems by sequence-specific rRNA cleavage method. *Water research* **2012,** *46* (7), 2167-2175.

63. Coates, J. D.; Bhupathiraju, V. K.; Achenbach, L. A.; Mclnerney, M.; Lovley, D. R., Geobacter hydrogenophilus, Geobacter chapellei and Geobacter grbiciae, three new, strictly anaerobic, dissimilatory Fe (III)-reducers. *International Journal of Systematic and Evolutionary Microbiology* **2001,** *51* (2), 581-588.

64. Morita, M.; Malvankar, N. S.; Franks, A. E.; Summers, Z. M.; Giloteaux, L.; Rotaru, A. E.; Rotaru, C.; Lovley, D. R., Potential for direct interspecies electron transfer in methanogenic wastewater digester aggregates. *MBio* **2011,** *2* (4), e00159-11.

65. Kato, S.; Hashimoto, K.; Watanabe, K., Methanogenesis facilitated by electric syntrophy via (semi) conductive iron-oxide minerals. *Environmental microbiology* **2012,** *14* (7), 1646-1654.

# CHAPTER 3: ENRICHMENT AND CHARACTERIZATION OF MICROBIAL CONSORTIA DEGRADING SOLUBLE MICROBIAL PRODUCTS DISCHARGED FROM ANAEROBIC METHANOGENIC BIOREACTORS

## 3.1 Abstract

SMP produced in bioprocesses have been known as a main cause to decrease treatment efficiency, lower effluent quality and promote membrane fouling in water reclamation plants. In this study, biological degradation of SMP using selectively enriched microbial consortia in a DHS reactor was introduced to remove SMP discharged from anaerobic methanogenic reactors. On average, 68.9 to 87.5% SMP removal was achieved by the enriched microbial consortia in the DHS reactor for >800 days. The influent SMP fed to the DHS reactor exhibited a bimodal MW distribution with 14-20 kDa and <4 kDa. Between these two types of SMP, the small MW SMP were biodegraded in the upper part of the reactor, together with most of the large MW SMP. Using 16S rRNA gene pyrosequencing technology, the microbial community composition and structure were characterized and correlated with operational factors, such as hydraulic retention time, organic loading rate, and removal of soluble COD at different depths of the reactor by performing network and redundancy analyses. The results revealed that *Saprospiraceae* was strongly correlated to the increasing SMP loading condition, indicating positive co-occurrences with neighboring bacterial populations. Different microbial diversity along with the depth of the reactor implies that stratified microbial communities could participate in the process of SMP degradation. Taken together, these observations indicate that the spatial and temporal variability of the enriched microbial community in the DHS reactor could effectively treat SMP with respect to changes in the operational factors.

## 3.2 Introduction

Biological treatment processes have been extensively used to treat wastewater containing dissolved organic materials. In these treatment processes, microbial cells are enriched to high concentrations (>1-2 g/L) to effectively degrade and mineralize organic matters to carbon dioxide. Concurrently, energy is derived for the growth of microbial cells, and soluble microbial products (SMP) are secreted into the bulk solution. SMP generally

contain a wide range of soluble, complex, and heterogeneous compounds with a MW ranging from 0.5-1000 kDa.[1-3] They are present in the effluent discharged from the treatment processes and are primary substances contributing to the increase in effluent COD.[2, 4] They can be a cause of increasing toxicity of the effluent by themselves[5] and an environmental hazard by acting as a precursor of disinfection by-products.[5-7] Accumulation of SMP in AS processes not only decreases respiration rates but also reduces efficiencies in flocculation, settling ability, and dewaterbility of AS by affecting physical properties in the processes, such as sludge structure, turbidity, and viscosity[8-9]. Increase of SMP in tertiary treatments can also have inhibitory effects on nitrification.[10]

The presence of SMP in discharged water can potentially have negative impacts to water reclamation processes. In membrane bioreactors and in membrane separation processes for water purification, SMP are reported to be responsible for membrane fouling by accumulating on the surface of membranes, blocking pores, and subsequently reducing the water flux through the membranes.[11-13] To remove foulants deposited on the membrane surface and restore the water flux passing through the membrane, membrane backwashing or chemical cleaning is often used. In extreme cases, these foulants can no longer be removed from the membrane surface. As a result, replacement of new membrane modules is required, which can increase operation costs.

It is important to develop strategies to effectively control and remove SMP in membrane-based water treatment and reclamation processes.[14-15] In these processes, adsorption and coagulation as pretreatments are often used to reduce SMP, and this can prevent or minimize the extent of fouling taking place on the membrane surface.[16-20] The most commonly used adsorbent is activated carbon in a form of granules or powders,[17-18, 20-21] and its use prior to microfiltration and ultrafiltration is reported as the most effective pretreatment to control SMP in secondary effluent.[22-23] However, the long-term application of activated carbon can be limited by its adsorption capacity.[18, 20-21, 23]

Alternatively, biologically degrading SMP has been suggested to control the amount of SMP in water treatment systems.[24] Biodegradation of SMP is feasible but at a slow rate due to the large MW and complex chemical structures.[25-26] However, when appropriate conditions are provided, effective degradation of SMP discharged from an anaerobic reactor can be achieved with efficiency up to 96% on high MW SMP (> 100 kDa); the degradation

efficiency of SMP is observed to be more effective under aerobic conditions than anaerobic conditions.[27] Microbial community compositions related to anaerobic SMP degradation have not been characterized, but a few previous studies are limited to identifying phylogenetic groups of heterotrophic bacteria utilizing SMP produced by nitrifying bacteria at the phylum and class levels.[28-30] The decrease of SMP in the system was just speculated to have a correlation with the abundance of *Klebsiella* in a biological activated carbon reactor[31] and *Chloroflexi* in a membrane bioreactor (MBR).[32]

Several reports have also described the use of a down-flow hanging sponge (DHS) reactor as a post-treatment to treat the effluent discharged from UASB processes treating domestic wastewater.[33-37] Additional 70-80 % reduction in COD by the DHS reactor was reported. Although no measurement was performed to confirm the molecular size of the COD present in the UASB effluent, it is possible that the majority of the COD was primarily made of SMP, and a large fraction of the SMP was biodegraded by the microbial populations selectively enriched in the DHS reactors.

In this study, to understand biological degradation of SMP in a DHS reactor as a post-treatment process to the effluent of anaerobic methanogenic reactors, the spatial and temporal variability of the community composition and structure of the enriched microbial consortia was characterized using 16S rRNA-based pyrosequencing. In addition, the key microbial populations involved in SMP degradation and their relationships with the operational factors were identified and evaluated by applying network and redundancy analyses.

## 3.3 Material and methods

### 3.3.1 Experimental set up

Figure 1A illustrates the use of a DHS reactor to enrich microbial consortia that could degrade SMP present in the effluent of anaerobic bioreactors. To produce the required SMP-containing effluent, two anaerobic reactors, named an anaerobic packed-bed reactor (AP) and a hybrid packed-bed reactor (HP), were operated to treat synthetic wastewater that mimicked the wastewater composition discharged from soft drink production plants (Table B.1). The detailed information of the system performance of the AP and the HP reactors is described elsewhere.[38] Briefly, the OLR increased from 0.5 g SCOD/L/day to 2.0 g

SCOD/L/day, and the removal efficiency of SCOD was 93-97% with consequent average effluent SCOD 121.9 ± 106.5 mg/L and 123.4 ± 74.9 mg/L in the AP and the HP reactors, respectively. The DHS reactor was fed with combined effluent discharged from the anaerobic reactors as the sole substrate. With a working volume of 10 L, the DHS reactor was filled with polyurethane sponge media (porosity 0.985 vol./vol.) covered by net cylinder-shape polyethylene cases (L34xD34xH34, unit mm). AS from the Urbana-Champaign sanitary district at Urbana, Illinois was inoculated into fifteen pieces of sponge and randomly placed at the top, middle, and bottom parts of the DHS reactor. The HRT was decreased from 1.82 to 0.52 days stepwise with a consistent internal-circulation rate of 50 ml/min (Table 3.1). In total, five phases based on HRT and OLR were defined for the operation conditions. The reactor was maintained at room temperature without additional aeration and pH adjustment.

**Table 3.1** Operational conditions and performance of the DHS reactors in the five phases.

| Phase | | Day | Average influent SCOD (mg/L) | HRT (Day) | Influent SCOD (mg/L) | OLR (mg SCOD/L/day) | SCOD removal (%) | SCOD removal (mg SCOD /L/day) |
|---|---|---|---|---|---|---|---|---|
| I | i | 0-135 | 106.5 ± 9.6 | 1.82 | 106.5 ± 9.6 | 58.5 ± 5.3 | 65.8 ± 22.5 | 38.5 ± 5.1 |
| | ii | 136-230 | 66.9 ± 11.3 | 1.82 | 66.9 ± 11.3 | 36.8 ± 6.2 | 66.4 ± 5.3 | 24.4 ± 3.0 |
| | iii | 231-335 | 105.2 ± 13.8 | 1.82 | 105.2 ± 13.8 | 57.8 ± 7.6 | 76.8 ± 5.2 | 44.4 ± 3.7 |
| | iv | 336-461 | 55.9 ± 15.5 | 1.82 | 55.9 ± 15.5 | 30.7 ± 8.5 | 65.8 ± 7.0 | 20.2 ± 4.2 |
| II | | 462-603 | 48.7 ± 8.3 | 1.21 | 48.7 ± 8.3 | 40.2 ± 6.8 | 72.9 ± 5.5 | 29.3 ± 2.3 |
| III | | 604-691 | 112.8 ± 60.8 | 0.91 | 112.8 ± 60.8 | 124.0 ± 66.8 | 72.9 ± 6.3 | 90.4 ± 25.0 |
| IV | | 692-753 | 92.9 ± 23.8 | 0.67 | 92.9 ± 23.8 | 138.7 ± 35.5 | 73.9 ± 5.6 | 102.5 ± 32.7 |
| V | | 754-824 | 225.3 ± 65.3 | 0.52 | 225.3 ± 65.3 | 433.3 ± 125.6 | 87.5 ± 6.5 | 379.1 ± 142.4 |

### 3.3.2 Analytical procedures

Soluble COD (SCOD) in the samples was characterized using COD digestion kits (HACH, 2125815) with a UV/VIS spectrophotometer (DR/4000 U Spectrophotometer_115 Vac, HACH Company, USA) after filtration with 0.22 μm filters (Millex-GP, Millipore, MA, USA). Dissolved organic carbon in the filtered sample was measured using an automated total organic carbon analyzer (TOC_Vcph, Shimadzu, Japan). The filtered

soluble samples were subjected to MW distribution of SMP, using a high performance liquid chromatography (HPLC) – size exclusion chromatography (SEC) (P680A LPG-2, Dionex, US) equipped with a Zorbax GF-250 column. SMP were detected with the UV detector at a wavelength of 254 nm. 0.01M phosphate solution filtered through 0.22 μm filters was used as the mobile phase at a flow rate of 1.0 mL/min. A standard curve was generated using a protein-based molecular weight marker kit (MWGF-200, MW ranges 12-200 kDa, Sigma) and Q-Dextran (MW 4 kDa, Sigma). The peak areas were calculated based on the peak intensity and the peak retention time in the chromatograms.

### 3.3.3 Biomass sampling, DNA extraction, amplification of 16S rRNA genes, and pyrosequencing analysis

To sample the biomass from the DHS reactor, a piece of sponge media was collected bimonthly from the upper (depth, 0.16 m) and lower (depth, 0.94 m) parts of the reactor. The biomass in the sponge was suspended in 25 ml of 1x phosphate buffered saline (PBS) solution by vortexing, pelleted by centrifugation (10000 rpm, 3 min), and stored in -80$^{o}$C prior to DNA extraction. Biomass was taken at days 82 (only from the upper part), 136, 196, 258, 373, and 454 in Phase I, days 528 and 602 in Phase II, day 648 in Phase III, day 723 in Phase IV, and day 798 in Phase V.

DNA was extracted using a FastDNA spin kit for soil (MP Biomedicals, Carlsbad, CA, USA) and purified with the Promega Wizard DNA clean up system. A LIB-L kit (454, Roche, Basel) with a primer set targeting the 515F-909R region of 16S rRNA gene sequences[39] was used for PCR amplification. PCR products were separated by 1.5% low melting gel electrophoresis and extracted with a Wizard SV Gel and PCR clean-up system (Promega, USA). A Qubit fluorometer (Invitrogen, USA) was used to quantify the PCR products. Equal amounts of the PCR products were combined and analyzed by a 454 Genome Sequencer FLX Titanium platform (Roche/454 Life Sciences, Branford, CT, USA) at the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign (IL, USA) for pyrosequencing. The pyrosequencing data have been deposited in NCBI-Sequence Read Archive (accession no. SRP056366). The Quantitative Insights Into Microbial Ecology (QIIME) pipeline was used to process the pyrosequencing data.[40]

Pyrosequencing results were processed using the Quantitative Insights Into Microbial Ecology (QIIME) pipeline[40] with the default settings as follows: after the sequences were denoised, operational taxonomic units (OTUs) were assigned by the UCLUST algorithm (≥ 97% pairwise identity). Representative sets of sequences from each OTU were then formed with identical sequences collapsing. Then, PyNAST[41] was used to align the representative sequences of each OTU to the Greengenes imputed core reference sequences. Chimeric sequences were removed by Chimera Slayer.[42] The PyNAST alignment was filtered to remove gaps and a phylogenetic tree was built using FastTree.[43] The taxonomy of each OTU was assigned using the Ribosomal Database Project (RDP) classifier with a Greengenes-based training dataset at a confidence threshold of 0.8. Chao1 richness estimator,[44] Good's coverage,[45] Equitability, phylogenetic diversity (PD), and Shannon diversity indices were calculated in QIIME.

### 3.3.4 Statistical methods

A weighted UniFrac beta-diversity distance matrix was constructed from the phylogenetic tree and subjected to the non-metric multidimensional scaling (NMDS) analysis. Redundancy analysis (RDA) was used to evaluate the correlation of the operational factors, including HRT, OLR, SCOD removal and sampling locations with the temporal and spatial variability of the microbial community at the genus (Table 3.1). RDA was conducted using CANOCO v.4.5 (Microcomputer Power, Ithaca, NY). To determine statistically significant variables ($P<0.05$), the forward selection method was conducted using the Monte Carlo test (499 permutations).[46]

### 3.3.5 Association network analysis

Network analysis of operational taxonomic units (OTUs) with the operational factors was performed using CoNet.[47] OTUs that had a relative abundance of at least 4% of the total in any community were considered (Table B.2). The association between i) the abundances of any two OTUs and ii) the abundance of an OTU and the value of an operational factor was calculated based on the Pearson correlation. The association network analysis was conducted with both operational factors, HRT and OLR, respectively and the respective two networks were subsequently merged.

### 3.3.6 Phylogenetic analysis

In order to assess the phylogenetic affiliation of the OTUs that indicated a significant correlation with HRT and OLR in the network analysis, the aligned sequences of the OTUs were imported into ARB and then added to the Greengenes ARB database (Greengenes_16S_2011_1.arb) using ARB parsimony method. A phylogenetic tree for the aligned sequences with their neighboring sequences was built using the neighbor-joining algorithm with Jukes-Cantor correction. Bootstrap values were calculated based on 1000 replications.

### 3.3.7 Principal component analysis of 16S rRNA gene sequences data sets from DHS and other ecosystems

The 16S rRNA gene sequences of AS, sewage, a pilot-scale DHS process, and soil were collected from the National Center for Biotechnology Information (NCBI) and GenBank nucleotide databases (Table B.3). QIIME pipeline was used for phylogenetic analysis of the 35 microbial communities with the default settings described previously. The principal component analysis (PCA) was carried out based on the relative abundance of phylogenetic groups at the family level in each sample using the Multibase program (Numerical Dynamics; www.numericaldynamics.com).

## 3.4 Results and discussion

### 3.4.1 Operational conditions and SCOD removal

The DHS operation during the 824 days was divided into five phases according to the decrease in HRTs from 1.82 days to 0.52 days (Figure 3.1 and Table 3.1). Under an HRT of 1.82 days, Phase I was further divided into four different sub-phases based on OLRs. Despite fluctuation in the OLR, the effluent SCOD concentration was stabilized at 23.1 ± 5.8 mg/L, and the average SCOD removal efficiency was 68.9 ± 10%. As the HRT was decreased, the OLR was increased almost 10 times from Phase I (46.3 ± 14.2 mg SCOD/L/day) to Phase V (433.3 ± 125.6 mg SCOD/L/day). Still, the effluent SCOD concentration remained around 23.9 ± 9.9 mg/L, and the SCOD removal efficiency increased from 68.9 ± 10.0% in Phase I to 87.5 ± 6.5% in Phase V. During the overall

operation, the performance of the DHS reactor in terms of SCOD removal was not seriously affected by the fluctuation or a sudden increase in influent organic loading, suggesting that DHS reactors could effectively polish the effluent quality and produce stable and low SCOD effluent under various organic loading conditions. During the 824 days of operation, either sloughing or detachment of biomass was not observed.

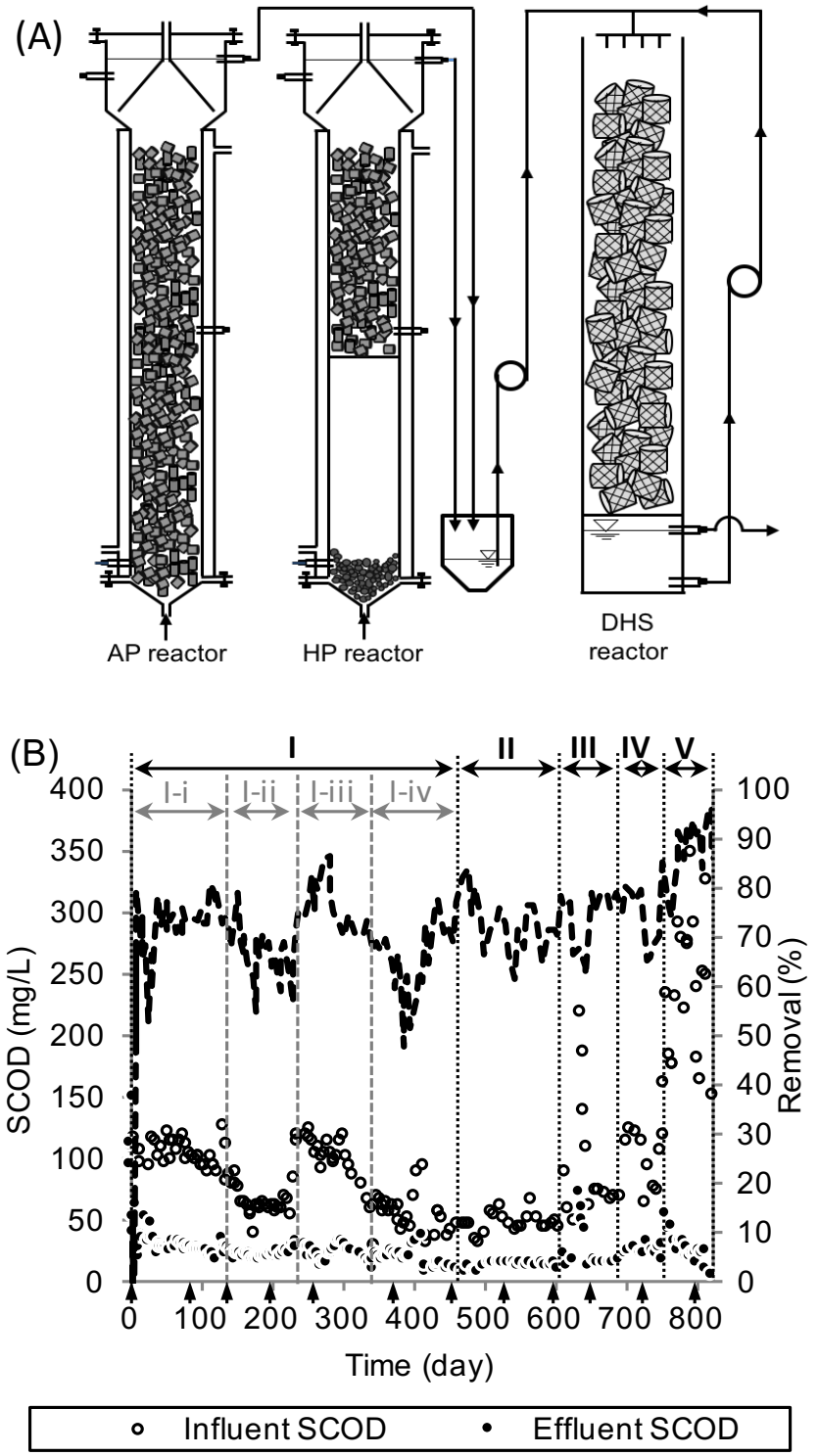



**Figure 3.1** (A) Schematic diagram of the anaerobic packed-bed reactors the DHS reactor. (B) SCOD removal of the DHS reactor in five different phases. Phase I was divided into four different subphases based on OLRs. The arrows along with the time axis indicate the operational days when biomass in the supporting sponge media was collected for microbial community analysis.

### 3.4.2 SMP removal

SCOD reduction in the DHS reactor suggested that SMP released from the anaerobic reactors could be effectively degraded. To analyze and compare changes in SMP profiles before and after the DHS treatment, samples from the effluent of the AP and HP reactors and the DHS reactors were collected in each phase. The MW distribution of the SMP showed a bimodal distribution with the major MW between 14-20 kDa, and the minor MW less than 4 kDa (Figure 3.2). Considering that the major constituents in the synthetic wastewater (Table B.1) were glucose, fructose, and polyethylene glycol (MW, 200 Da), of which the MWs were less than 200 Da, the observed SMP with a MW of 14-20 kDa in the anaerobic effluent were likely to be SMP-derived from metabolism of the anaerobic biomass.[2, 48-49] Figure 3.2 further indicated the relative SMP removal of the large and small MW fractions in the DHS reactor, respectively. Although quantitative comparison of SMP reduction in the DHS reactor could not be made without standard compounds for quantification at each MW range, it was observed that the SMP associated with the peak areas (14-20 kDa) in the chromatogram in the AP and HP effluent were greatly reduced in the DHS effluent. Most of the SMP with an MW less than 4 kDa were degraded through the DHS in most of the phases. The pertinent chromatograms are shown in Figure B.1.

The SMP degradation profile, after combined with the recirculation, along the reactor depth was further investigated (Figure 3.2). For the SMP with an MW of 14-20 kDa (n=4), 80.4% reductions in the chromatogram area was observed in the upper part of the reactor with a depth between 0.0 m and 0.32 m, and 8.6% reductions occurred in the lower part (0.65-1.14 m). For the small MW SMP (< 4 kDa), reduction was mainly observed in the upper part of the reactor. A SCOD profile for the degradation of the total organic compounds also indicated that 64.5% and 15.3% of influent SCOD were removed in the upper and lower part of the reactor, respectively (Figure B.2).

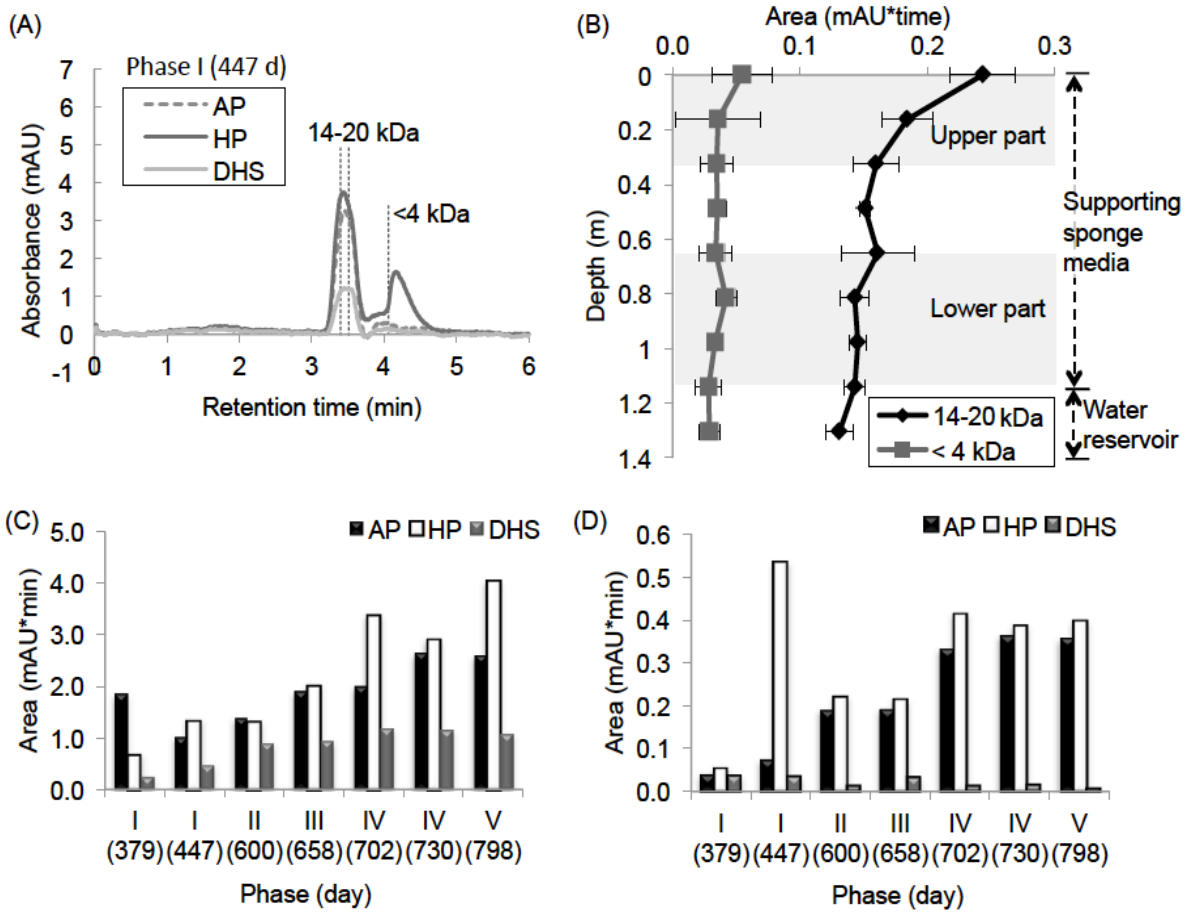



**Figure 3.2** HPLC-SEC analyses of the effluent SMP from the AP and HP reactors, and the effluent from the DHS reactor in the five phases: (A) a chromatogram in Phase I at day 447, (B) degradation profiles of SMP sub-fractions along with the DHS reactor depth (n=4), (C) chromatogram peak areas corresponding to MW range of 14-20 kDa, and (D) chromatogram peak areas corresponding to MW range of <4 kDa. The peak areas were calculated based on the peak intensity and the peak retention time in the chromatograms. The number in parentheses indicates the days when the samples were collected.

### 3.4.3 Microbial community dynamics

A NMDS plot based on weighted UniFrac distances (stress value = 0.046) indicated that samples collected from the DHS reactor tended to cluster together and were separated from the inoculated AS and the influent biomass (Figure 3.3). A significant shift in microbial composition for the samples collected from the upper part of the reactor was observed over time. Four of the six samples taken in Phase I were closely clustered, and this cluster could be differentiated from those taken in Phases II-III and Phases IV-V, which also formed individual distinct clusters. The samples collected from the lower part of the reactor clustered separately from the samples taken from the upper part, likely due to differences in substrate concentration at the upper and lower parts of the reactor. Among these samples, there is no clear shift in community structure along with the different phases (I-V).

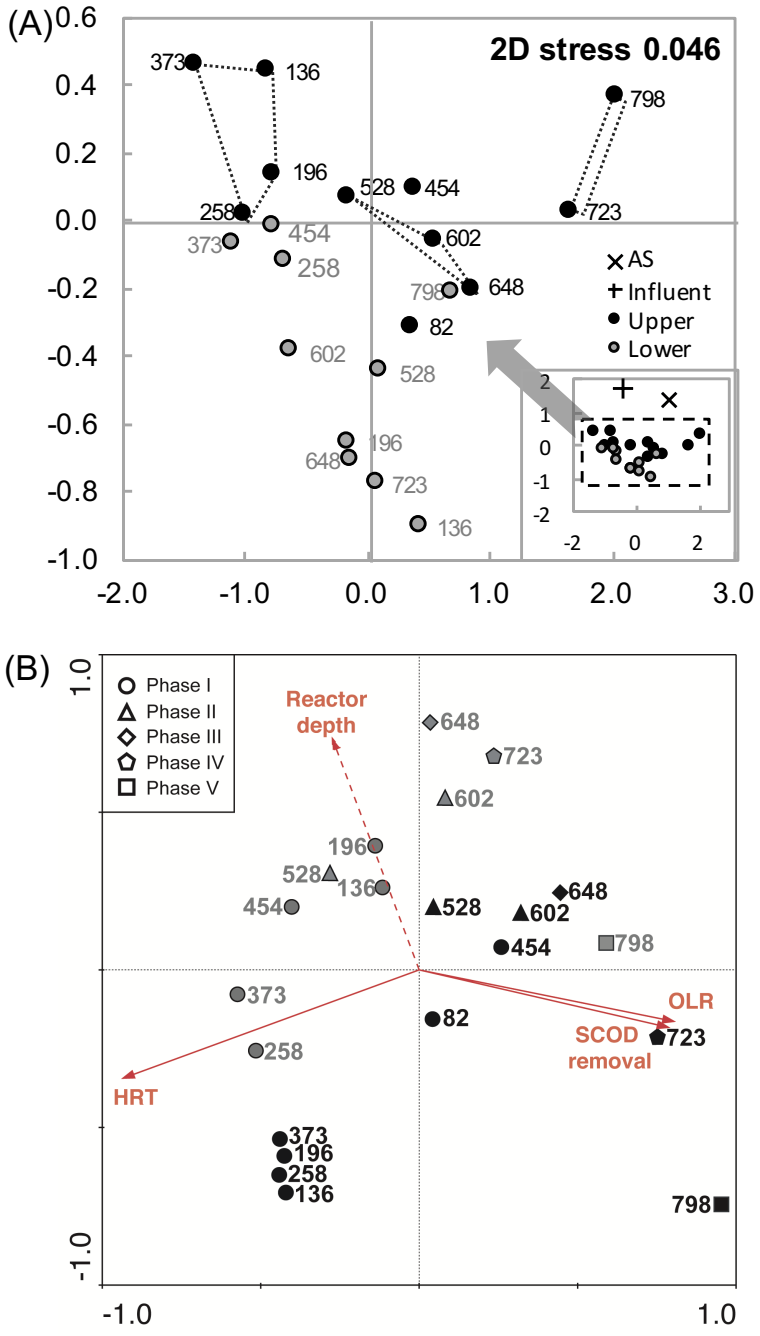



**Figure 3.3** Ordination diagrams of (A) non-metric multidimensional scaling and (B) redundancy analysis for the samples from the upper and lower parts of the DHS reactor. The upper part samples are indicated in black circles with the sampling days in black letters, and the lower part samples are indicated in gray circles with the sampling days in gray letters. 'AS' stands for the inoculated activated sludge. HRT, OLR, and SCOD removal are indicated by red arrows due to their statistical significance ($P < 0.05$), and reactor depth is indicated by a red dotted arrow ($P > 0.05$).

To determine the factors affecting the spatial and temporal differences in the microbial communities observed above, RDA was applied with four explanatory variables (HRT, OLR, SCOD removal, and reactor depth). Results indicated that HRT, OLR, and SCOD removal were the statistically significant variables ($P < 0.05$) correlating with the microbial community compositions classified by genus (Figure 3.3). Due to the stable reactor performance in terms of SCOD removal (Table 3.1), OLR and SCOD removal indicated a high positive correlation with each other. The four samples (136, 196, 258, and 373) collected in Phase I from the upper part of the reactor exhibited the strongest correlation with HRT but exhibited negative correlations with the rest of the variables, whereas the upper part samples in the later phases came to have positive correlations with OLR and SCOD removal but rather a negative correlation with HRT. Comparing the upper and lower part samples of the same sampling days, as expected the samples from the lower part of the reactor had stronger correlations with the reactor depth than the upper ones. In this comparison, the lower samples, projecting away from the increasing direction of OLR and SCOD removal in the ordination, were not immediately affected by changes in the organic loading. This is possibly because the microorganisms at the lower part were exposed to the partially degraded organics derived during the SMP degradation from the upper part of the reactor. However, in Phase V when the organic loading significantly increased, the lower part sample was also affected by OLR and SCOD removal the most, rather than the other variables.

### 3.4.4 Microbial community composition

The microbial community profile indicates clear shifts of dominant microbial populations in the individual samples at the genus level (relative abundance >3% of the total sequencing reads in any sample) with respect to the different phases and the sampling depth (Figure B.3). *Gordonia* (14.5-34.3%) and *Ectothiorhodospiraceae* (9.9-25.4%) were most abundant in both the upper and lower parts of the reactor in Phase I, and rapidly decreased to < 1% in Phases II-V. The genus *Cytophaga* (12.1-22.8%) also dominated in Phase I and was more abundant at the lower part than at the upper part. In Phases II-V, the abundances of *Flavobacteria* and *Saprospiraceae* in samples taken from the upper part increased from 4.4% to 25.9% and from 12.1% to 30.1%, respectively. In addition, *Bacteroidales*-related

genus, *Dechloromonas*, and *Geobacter*, which were hardly detected in Phases I-III, increased up to 25.9%, 5.1%, and 8.7% in Phase IV-V, respectively. Unclassified *Sphingobacteriales* genus (2.7-12.5%) was found evenly in all samples collected at the different depths and the phases.

We further compared the microbial composition in the DHS with that present in the anaerobic reactor effluent. Out of 4613 OTUs detected in the DHS community, 4328 were unique to the DHS (Figure B.5). Among the shared OTUs, three most-abundant OTUs detected in the anaerobic effluent were selected and their abundances plotted along with the DHS operation (Figure B.5). The abundance profiles of these three OTUs in the DHS did not exhibit the same trends as observed in the anaerobic reactor effluent. Similar observations could be made with all dominant OTUs found in DHS and the anaerobic effluent during the DHS operation (Table B.7). Thus, it could be concluded that microbial populations enriched in the reactor primarily represented the microbial community in the DHS reactor, and the microbial populations carried over from the anaerobic reactor effluent had insignificant impact on the microbial composition in the DHS.

To investigate the correlation of the microbial population shifts with HRT and OLR at the different depths of the reactor, microbial association networks with respect to the operational factors were constructed based on relatively abundant OTUs (Table B.2). Twelve OTUs were determined to have direct significant correlations with either HRT or OLR (Figure 3.4). The detailed phylogenetic examination of the OTUs was performed by constructing a neighbor-joining tree with previously reported sequences (Figure B.4). OLR was positively associated with proliferation of *Saprospiraceae*-related OTUs (OTUs 1229, 2194, and 588), *Flavobacteriales*-related OTU2195, *Geobacter*-related OTU6110, and *Azobacter*-related OTU3311. Of those, OTUs 1229, 2194, 2195, and 588 were negatively associated with HRT together with *Phycisphaerae*-related OTU2325. OTUs showing strong positive correlations with HRT were *Ectothiorhodospiraceae*-related OTU1933, *Gordonia*-related OTU5208, *Cytophaga*-related OTUs 1661 and 3624, and *Nitrospira*-related OTU225. When the relative abundance of these OTUs between the upper and lower reactor was compared, OTUs 1229, 2194, 2195, and 6110 adapted to the increasing OLR faster in the upper part of the reactor than the lower part, whereas OTUs 3311 and 588 became proliferative in the lower part. Among the OTUs that were strongly correlated to the HRT,

OTUs 1993 and 5208 were abundant in the upper part of the reactor while OTUs 1661, 225, and 3624 were abundant in the lower part.

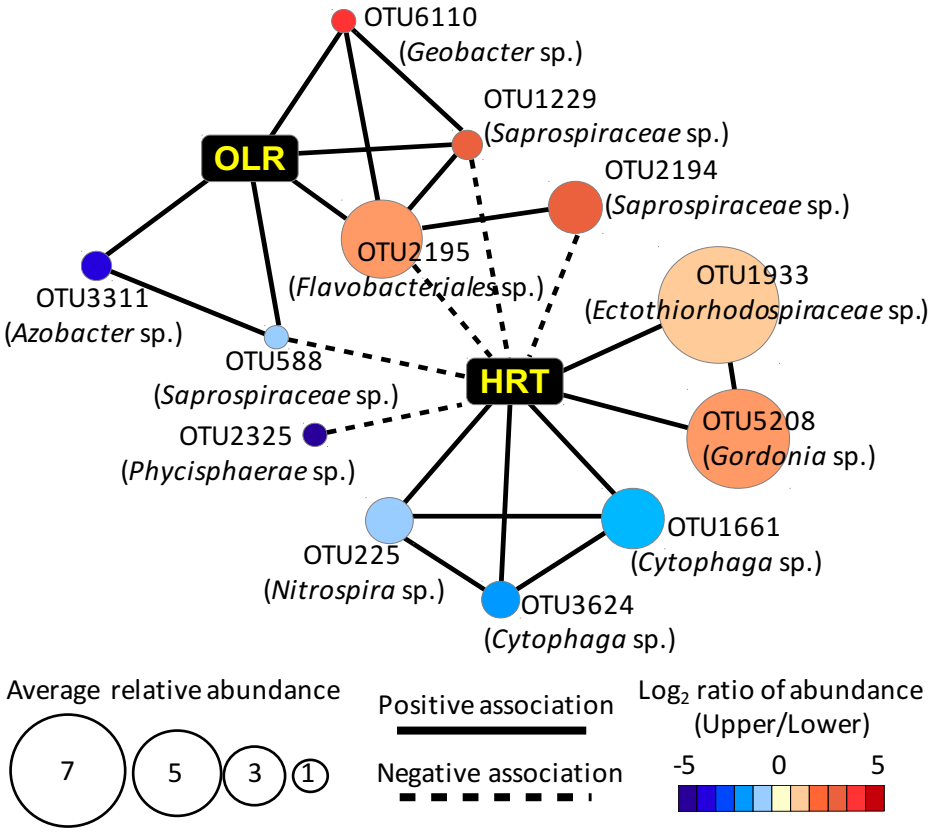



**Figure 3.4** Association network of OTUs with two operational factors (HRT and OLR). Each node represents an OTU (defined at 97% identity level). The size of the node represents the average relative abundance of the OTU across all communities. The node is color-coded (see key) by the mean fold change (logarithmic scale) of the average relative abundance at the upper parts compared to the lower parts. Each edge represents a positive or negative association with a correlation coefficient higher than 0.5 or less than -0.5 and P value of < 0.05. OTUs that were the first neighbors of either HRT or OLR are shown. The taxonomic classification of the OTU is provided in parentheses.

Similar results to the network analysis were observed in the RDA ordination with the relatively abundant genera (Figure B.5). *Dechloromonas*, *Geobacter*, and genera in *Bacteroidales*, *Flavobacteria*, and *Saprospiraceae* were strongly correlated to OLR and SCOD removal, whereas *Ectotiorhodospiraceae*-related genus, *Gordonia*, and *Cytophaga* were closely correlated to HRT. The genera showing the most positive correlation with reactor depth included *Nitrospira* and *Caldilineaceae*-related genus. *Sphingobacteriales*-

related genus (OTU1682) that was abundant throughout all phases did not show a strong correlation with any of the variables.

### 3.4.5 Characteristics of SMP

Carbohydrate and protein complexes are often identified as the main components of SMP.[3, 50] These SMP were identified to contain long-chain alkenes, alkanes, aromatic compounds, esters, humic acids, uronic acids, and nucleic acids using advanced analytical methods such as gas chromatography-mass spectrometry (GC-MS) and matrix-assisted laser desorption/ionization (MALDI)-time of flight (TOF)/mass spectrometry (MS).[2, 51-54] To describe how SMP originate, they are classified into utilization-associated products (UAP), which are produced directly from substrate utilization, and biomass-associated products (BAP), which are formed from cell lysis and decay.[14] Furthermore, SMP can be characterized based on the MW distribution. A bimodal MW distribution was verified as a generic phenomenon in various studies, although the compositions and ranges of the MW greatly varied depending on feed water compositions, sludge mixtures, and operational and environmental conditions that a system was exposed to.[1, 3, 25, 55-57] Several studies among them reported that most SMP found in a range of <1 and >10 kDa as SMP < 4 kDa and between 14-20 kDa was observed in our study.[3, 55-57] The skewed bimodal distribution to the large MW SMP can be observed in a system operated at a long solid retention time (SRT), as were the anaerobic reactors in this study.[58] To correlate the MW of SMP and the origin, it was suggested that UAP were made of low MWs, and BAP tended to contain high MWs and accumulated in the bioreactor.[58] Taking this into consideration, we surmise that SMP detected in the range of < 4 kDa MW could likely represent UAP, and those detected in the range of 14-20 kDa MW could likely represent BAP. The small MW SMP were almost completely removed, whereas the large MW ones were partially degraded. The difference in the degradability of the large and small MW fractions was likely due to the difference in the structure complexity of SMP; the biodegradability of the BAP fraction was relatively lower than that of the UAP fraction.[59] Despite the low biodegradability of large MW SMP, Barker et al. (2000) reported that relatively high removal (74%) of large MW SMP (> 10 kDa) produced from an anaerobic reactor could be achieved in a following aerobic treatment. This

observation was relevant to our results that more than half of the peak areas of the 14-20 kDa MW SMP in the chromatograms were reduced by the degradation in the DHS reactor.

### 3.4.6 Uniqueness of SMP-degrading microbial community in DHS

Based on the efficiency of SMP degradation obtained in this study, we conclude that SMP-degrading microbial community was successfully enriched in the DHS reactor. It is not clear how different the SMP-degrading microbial community is from those found in conventional wastewater treatment processes. To address this, PCA was used to compare the SMP-degrading microbial communities in the DHS reactor with those (Table B.3) observed in AS, integrated fixed-film AS (IFAS), a pilot-scale DHS reactor (PS-DHS) treating effluent discharged from a UASB reactor, raw sewage, and soil at the family level (Figure 3.5).[37, 60-63] The microbial communities in the DHS reactor were relatively close to the PS-DHS, sewage, and IFAS, and distinctly distant from the AS and soil. Especially, the microbial communities in the upper part of the PS-DHS were the closest related ones to the DHS. This result is reasonable as the microorganisms in the upper part of the PS-DHS could effectively degrade organic matter in the UASB effluent that were likely made of mainly SMP [37]. However, the detailed phylogenetic affiliation of the most abundant OTUs in the upper part of the PS-DHS showed a different observation from our study. In the PS-DHS, the majority of the sequences were related to *Dechloromonas* in the early phase and then to *Firmicutes* and *Xanthomonas* later, whereas these were generally minor in the DHS even though the abundance of *Dechloromonas* and *Fusibacter* in *Firmicutes* slightly increased in Phase V. Also, the methane-oxidizing family significantly found in the PS-DHS was nearly undetectable in the DHS. Microbial communities from the middle and lower parts of the PS-DHS, where nitrifiers became abundant for ammonia and nitrite oxidation, formed a more distant cluster from the DHS than those in the upper part. The overall results suggest that the microbial consortia in the DHS were uniquely enriched for SMP degradation, which were apparently different from the other wastewater treatment processes.

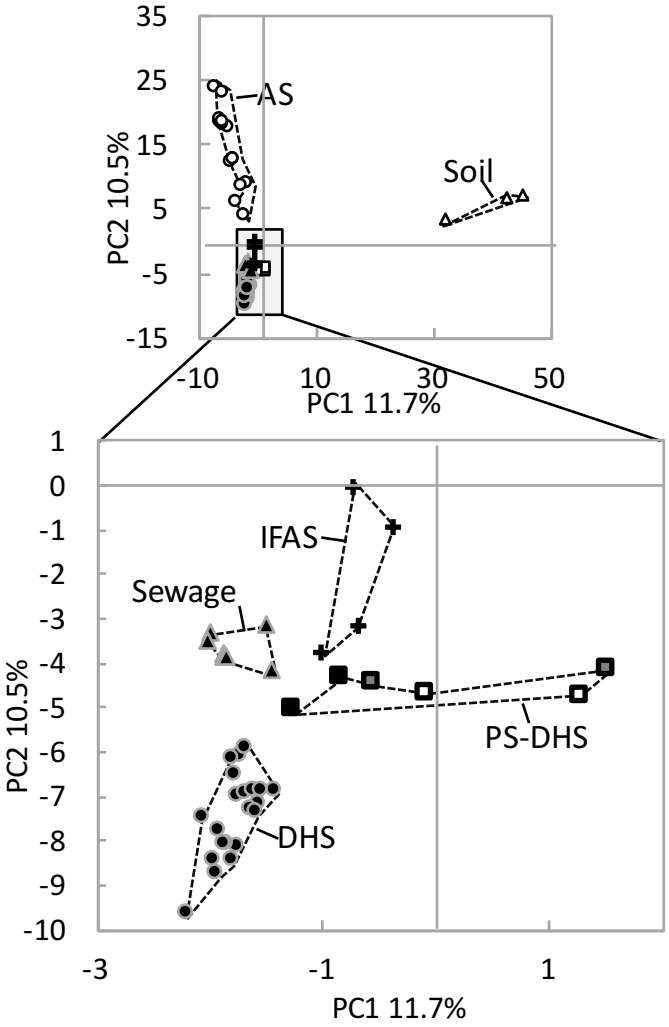



**Figure 3.5** Principal component analysis of the six different environmental samples using the relative abundance of bacterial 16S rRNA gene sequences at the family level. DHS stands for the samples collected in this study. AS, IFAS, and PS-DHS stand for activated sludge, integrated fixed-film activated sludge, and pilot-scale DHS, respectively. Among the samples from the PS-DHS, the samples collected from the upper part of the reactor are indicated by black squares, the samples collected from the middle part by gray squares, and the samples collected from the lower part by white squares.

### 3.4.7 Microbial populations involved in SMP degradation

The unique microbial community detected in the DHS reactor strongly suggested certain microbial populations were capable of directly or indirectly utilizing SMP. So far, there is very little information describing the composition and diversity of microbial populations degrading SMP prior to this study. Using microautoradiography combined with fluorescence in situ hybridization (MAR-FISH) analysis, it was revealed that several

bacterial groups, including *Chloroflexi*, *Cytophaga-Flavobacterium* cluster, *α-Proteobacteria*, and *γ-Proteobacteira*, were able to take up $^{14}$C-labeled SMP released from nitrifiers in nitrification reactors.[28] Specifically, it was suggested that *Chloroflexi*-related populations took up BAP, and the *Cytophaga-Flavobacterium*-related populations gradually ingested both UAP and BAP from the nitrifying bacteria. Miura et al.[32, 64] further reported that the abundance of *Chloroflexi* was correlated to the decrease of microbial products accumulated in a MBR, suggesting that *Chloroflexi* could alleviate membrane fouling by utilizing SMP from other heterotrophs. Unlike the high abundance of *Chloroflexi* observed in SMP-degrading MBR reactors (14-26%),[64] our study detected low abundance of *Chloroflexi* (0.02-9.2%) in the DHS reactor. This difference clearly suggests that microbial populations other than *Chloroflexi* could play an important role in the degradation of SMP in the DHS reactor.

In this study, a rapid increase in the abundance of *Flavobacteriales* and *Saprospiracea* was observed to correlate with an increase in OLR (Figure 3.4). These populations were key active members in SMP degradation in the DHS reactor, in particular at the upper part. The family *Saprospiraceae* was reported as active protein hydrolyzers, showing epiphytic growth attached to filamentous bacteria, like the phylum *Chloroflexi*.[65] The most abundant *Saprospiraceae*-related OTU2194 (average relative abundance 2.5%) showed a 99% similarity to *Candidatus Epiflobacter* sp. (EF523446), constituting a deep branch with the sequences in this genus that specifically utilizes amino acids as energy and carbon sources rather than other types of macromolecules (Figure B.6).[66] Co-occurrence of commensal *Flavobacteria* and *Sphingobacteria* with *Saprospiraceae* was observed in AS treating municipal wastewater, which leads to a speculation that the commensal bacteria rely on amino acids hydrolyzed from proteins by *Saprospiraceae*.[67] Considering that members of *Saprospiracea* could behave like a micropredator to obtain nutrients under conditions where organic substrates are limited,[68-69] the proliferation of *Saprospiracea* in the DHS reactor was likely due to the degradation of protein-like SMP released from biomass lysis, so that the neighboring *Flavobacteriales*, *Geobacter*, and *Azobacter*-related OTUs in the network were to be cross-fed on intermediates in the protein degradation. Enhancing abundance and activity of these key microbial populations, specifically responding to increase in the SMP

loading, in enriched consortia would be helpful to improve the removal efficiency of SMP in practical bioprocesses.

Although, in previous studies, *Cytophaga* and *Flavobacteria* in *Bacteroidetes* as a cluster have been often targeted together for utilization of dissolved organic materials and microbial products,[28, 70] *Cytophaga*-related OTUs (OTU1661 (average relative abundance 3.2%) and OTU3624 (average relative abundance 1.3%)) in our study, unlike the *Flavobacteriales*, were observed to be most abundant at the lower part of the reactor in the early phases when 3 to 10 times less SMP were fed compared with Phase IV and Phase V. The OTUs were also positively associated with HRT that made a negative correlation with OTU2195 in the network (Figure 3.4). This suggests that *Cytophaga* might have higher substrate affinity than *Flavobacteria* when low SMP were available.

The other OTUs strongly correlated with HRT were OTU5208 and OTU1933. OTU5208 is closely related to the *Gordonia* species (i.e., 99% similarity to *G. hydrophobica* (X87340)) that are capable of oxidizing various types of aliphatic and aromatic hydrocarbons including recalcitrant natural compounds (Figure B.4).[71] Considering that various refractory alkenes, alkanes, and aromatic compounds were produced from anaerobic reactors fed with simply biodegradable substrates,[51] it is speculated that the *Gordonia* sp. might contribute to the degradation of carbohydrate- and aromatic-like fractions of the SMP that were not preferably consumed by other bacteria under the long HRT condition. OTU1933 is distantly related to *Ectothiorhodospiraceae*-related genera, such as *Thioalkalivibrio* which is known as autotrophic halo-alkaliphilic sulfur-oxidizing bacteria[72] (Figure B.4). Since the abundance of OTU1933 was concordant with OTU5208 over all phases and both had strong association with each other in the network (Figure 3.4), *Ectothiorhodospiraceae*-related OTU5208 might be commensal to the *Gordonia* sp., relying on intermediates released from sulfur-containing aromatic compounds.

### 3.4.8 Microbial diversity affected by the operational factors

This study observed that the microbial community in the upper part of the reactor became less diverse than that in the lower part over phases. The phenomenon was pronounced in Phase IV-V, when the OLR substantially increased compared with the previous phases, as the microbial communities from the upper part had fewer observed and

estimated OTUs and lower evenness at similar coverage than those from the lower part (Table B.4). This difference in the microbial diversity was likely due to the stratification of SMP degradation developed along with the reactor depth. Small MW SMP tended to be readily degraded by the microbial community at the upper part of the reactor, whereas degradation of large MW SMP, which likely represents BAP, could still occur at the lower part and might require a complex microbial community (Figure 3.2). This is supported by the strong correlation of *Caldilineaceae* in the phylum *Chloroflexi* with the reactor depth in the RDA (Figure B.5), considering that *Chloroflexi* was known for preferential utilization of BAP.[28] In addition, *Nitrospira* was more abundant in the lower part of the reactor than the upper part, which had a considerable correlation with reactor depth in the RDA analysis (Figure B.5). It indicates that nitrification occurred in this part of the reactor despite the lack of ammonia in the substrate and ammonia-oxidizing bacteria present in the microbial community. In consideration of the high transcription activity of ammonia-oxidizing bacteria at low abundance,[73] we infer that the presence of more *Nitrospira* at the lower DHS was due to the ammonia released from protein-like SMP degradation. A similar pattern of microbial diversities at different stratified layers was reported in the pilot-scale DHS reactor for oxidation of the organic matters and nitrification.[37] Moreover, the difference in the microbial diversity along the depth could be resulted from the differences in oxygen, nutrients, and SMPs availability in individual sponges along the reactor. Since the dissolved oxygen of the trickling flow in the DHS reactor was observed to vary from zero to 6-8 mg/L, likely due to differences in biomass concentration in each sponge and air diffusion into it,[74] microaerophilic and anaerobic condition could take place in the upper part of the reactor toward to the later phases (IV and V) of the operation. This could contribute to the increase in the abundances of facultative and obligate anaerobes, like *Dechloromonas* and *Geobacter*, in the upper part.

### 3.5 Conclusion

The microbial community selected in the DHS reactor effectively utilized SMP produced from anaerobic methanogenic reactors. Microbial community analysis further indicated that unique microbial communities different from other wastewater treatment processes were developed. Microbial groups that were not previously reported for taking

SMP, such as *Flavobacteriales*-, *Saprospiraceae*-, and *Gordonia*-related microorganisms, could play important roles in SMP degradation together with the microorganisms known for SMP utilization, like *Chloroflexi* and *Cytophaga*. The abundance of these microbial groups was significantly affected by HRT, OLR, and SCOD removal, and the microbial diversity was influenced by reactor depth. Although the microbial community effectively degrading SMP was revealed in this study, it is still unclear how the individual microbial populations participated in the degradation of SMP. Future studies should focus on the mechanisms of SMP degradation and the role of individual microbial populations by applying function-driven genomic approaches (e.g., metagenomics and metatranscriptomics). Meanwhile, these findings suggest that SMP reduction by enriched microbial consortia in a DHS reactor will be a useful post-treatment of anaerobic processes, in the aspect of improving effluent quality as well as treatment performance and efficiency in bioprocesses and preventing fouling from a following membrane process for water reclamation.

### 3.6 Acknowledgments

We thank Dr. Hideki Harada at Tohoku University for kindly providing the supporting sponge media of the DHS reactor. This research was supported by a research grant from PepsiCo Inc.

### 3.7 References

1. Barker, D. J.; Stuckey, D. C., A review of soluble microbial products (SMP) in wastewater treatment systems. *Water Research* **1999,** *33* (14), 3063-3082.

2. Jarusutthirak, C.; Amy, G., Role of soluble microbial products (SMP) in membrane fouling and flux decline. *Environmental Science & Technology* **2006,** *40* (3), 969-974.

3. Ni, B. J.; Zeng, R. J.; Fang, F.; Xie, W. M.; Sheng, G. P.; Yu, H. Q., Fractionating soluble microbial products in the activated sludge process. *Water Research* **2010,** *44* (7), 2292-2302.

4. Barker, D. J.; Mannucchi, G. A.; Salvi, S. M. L.; Stuckey, D. C., Characterisation of soluble residual chemical oxygen demand (COD) in anaerobic wastewater treatment effluents. *Water Research* **1999,** *33* (11), 2499-2510.

5. Magbanua, B. S.; Bowers, A. R., Characterization of soluble microbial products (SMP) derived from glucose and phenol in dual substrate activated sludge bioreactors. *Biotechnology and Bioengineering* **2006,** *93* (5), 862-870.

6. Wei, Y. Y.; Liu, Y.; Zhang, Y.; Dai, R. H.; Liu, X. A.; Wu, J. J.; Zhang, Q. A., Influence of soluble microbial products (SMP) on wastewater disinfection byproducts: trihalomethanes and haloacetic acid species from the chlorination of SMP. *Environmental Science and Pollution Research* **2011,** *18* (1), 46-50.

7. Tang, H. L.; Chen, Y. C.; Regan, J. M.; Xie, Y. F. F., Disinfection by-product formation potentials in wastewater effluents and their reductions in a wastewater treatment plant. *Journal of Environmental Monitoring* **2012,** *14* (6), 1515-1522.

8. Chudoba, J., Inhibitory effect of refractory organic compounds produced by activated sludge micro-organisms on microbial activity and flocculation. *Water Research* **1985,** *19* (2), 197-200.

9. Reid, E.; Liu, X.; Judd, S. J., Effect of high salinity on activated sludge characteristics and membrane permeability in an immersed membrane bioreactor. *Journal of Membrane Science* **2006,** *283* (1–2), 164-171.

10. Ichihashi, O.; Satoh, H.; Mino, T., Effect of soluble microbial products on microbial metabolisms related to nutrient removal. *Water Research* **2006,** *40* (8), 1627-1633.

11. Choo, K. H.; Lee, C. H., Membrane fouling mechanisms in the membrane-coupled anaerobic bioreactor. *Water Res.* **1996,** *30* (8), 1771-1780.

12. Kang, I. J.; Yoon, S. H.; Lee, C. H., Comparison of the filtration characteristics of organic and inorganic membranes in a membrane-coupled anaerobic bioreactor. *Water Res.* **2002,** *36* (7), 1803-1813.

13. Liao, B. Q.; Kraemer, J. T.; Bagley, D. M., Anaerobic membrane bioreactors: applications and research directions. *Critical Reviews in Environmental Science and Technology* **2006,** *36* (6), 489-530.

14. Laspidou, C. S.; Rittmann, B. E., A unified theory for extracellular polymeric substances, soluble microbial products, and active and inert biomass. *Water Research* **2002,** *36* (11), 2711-2720.

15. Liao, B. Q.; Bagley, D. M.; Kraemer, H. E.; Leppard, G. G.; Liss, S. N., A review of biofouling and its control in membrane separation bioreactors. *Water Environment Research* **2004,** *76* (5), 425-436.

16. Park, H.; Choo, K. H.; Lee, C. H., Flux enhancement with powdered activated carbon addition in the membrane anaerobic bioreactor. *Separation Science and Technology* **1999,** *34* (14), 2781-2792.

17. Hu, A. Y.; Stuckey, D. C., Activated carbon addition to a submerged anaerobic membrane bioreactor: Effect on performance, transmembrane pressure, and flux. *Journal of Environmental Engineering-Asce* **2007,** *133* (1), 73-80.

18. Akram, A.; Stuckey, D. C., Flux and performance improvement in a submerged anaerobic membrane bioreactor (SAMBR) using powdered activated carbon (PAC). *Process Biochemistry* **2008,** *43* (1), 93-102.

19. Koseoglu, H.; Yigit, N. O.; Iversen, V.; Drews, A.; Kitis, M.; Lesjean, B.; Kraume, M., Effects of several different flux enhancing chemicals on filterability and fouling reduction of membrane bioreactor (MBR) mixed liquors. *Journal of Membrane Science* **2008,** *320* (1-2), 57-64.

20. Wu, B.; An, Y.; Li, Y.; Wong, F. S., Effect of adsorption/coagulation on membrane fouling in microfiltration process post-treating anaerobic digestion effluent. *Desalination* **2009,** *242* (1-3), 183-192.

21. Trzcinski, A. P.; Ofoegbu, N.; Stuckey, D. C., Post-treatment of the permeate of a submerged anaerobic membrane bioreactor (SAMBR) treating landfill leachate. *Journal of Environmental Science and Health Part a-Toxic/Hazardous Substances & Environmental Engineering* **2011,** *46* (13), 1539-1548.

22. Kim, S. L.; Chen, J. P.; Ting, Y. P., Study on feed pretreatment for membrane filtration of secondary effluent. *Separation and Purification Technology* **2002,** *29* (2), 171-179.

23. Pandey, S. R.; Jegatheesan, V.; Baskaran, K.; Shu, L., Fouling in reverse osmosis (RO) membrane in water recovery from secondary effluent: a review. *Reviews in Environmental Science and Bio-Technology* **2012,** *11* (2), 125-145.

24. Cai, L.; Yu, K.; Yang, Y.; Chen, B.; Li, X.; Zhang, T., Metagenomic exploration reveals high levels of microbial arsenic metabolism genes in activated sludge and coastal sediments. *Applied Microbiology and Biotechnology* **2013,** *97* (21), 9579-9588.

25. Jiang, T.; Myngheer, S.; De Pauw, D. J. W.; Spanjers, H.; Nopens, I.; Kennedy, M. D.; Amy, G.; Vanrolleghem, P. A., Modelling the production and degradation of soluble microbial products (SMP) in membrane bioreactors (MBR). *Water Research* **2008,** *42* (20), 4955-4964.

26. Rittmann, B. E.; McCarty, P. L., *Environmental biotechnology: principles and applications*. Tata McGraw-Hill Education: 2012.

27. Barker, D. J.; Salvi, S. M. L.; Langenhoff, A. A. M.; Stuckey, D. C., Soluble microbial products in ABR treating low-strength wastewater. *Journal of Environmental Engineering-Asce* **2000,** *126* (3), 239-249.

28. Okabe, S.; Kindaichi, T.; Ito, T., Fate of C-14-labeled microbial products derived from nitrifying bacteria in autotrophic nitrifying biofilms. *Appl. Environ. Microbiol.* **2005,** *71* (7), 3987-3994.

29. Kindaichi, T.; Ito, T.; Okabe, S., Ecophysiological interaction between nitrifying bacteria and heterotrophic bacteria in autotrophic nitrifying biofilms as determined by

microautoradiography-fluorescence in situ hybridization. *Appl. Environ. Microbiol.* **2004,** *70* (3), 1641-1650.

30. Matsumoto, S.; Katoku, M.; Saeki, G.; Terada, A.; Aoi, Y.; Tsuneda, S.; Picioreanu, C.; Van Loosdrecht, M., Microbial community structure in autotrophic nitrifying granules characterized by experimental and simulation analyses. *Environmental microbiology* **2010,** *12* (1), 192-206.

31. Cai, L.; Yu, K.; Yang, Y.; Chen, B.-w.; Li, X.-d.; Zhang, T., Metagenomic exploration reveals high levels of microbial arsenic metabolism genes in activated sludge and coastal sediments. *Applied Microbiology and Biotechnology* **2013,** *97* (21), 9579-9588.

32. Miura, Y.; Okabe, S., Quantification of cell specific uptake activity of microbial products by uncultured Chloroflexi by microautoradiography combined with fluorescence in situ hybridization. *Environmental Science & Technology* **2008,** *42* (19), 7380-7386.

33. Tandukar, M.; Machdar, I.; Uemura, S.; Ohashi, A.; Harada, H., Potential of a combination of UASB and DHS reactor as a novel sewage treatment system for developing countries: Long-term evaluation. *Journal of Environmental Engineering-Asce* **2006,** *132* (2), 166-172.

34. Tanaka, H.; Takahashi, M.; Yoneyama, Y.; Syutsubo, K.; Kato, K.; Nagano, A.; Yamaguchi, T.; Harada, H., Energy saving system with high effluent quality for municipal sewage treatment by UASB-DHS. *Water Science and Technology* **2012,** *66* (6), 1186-1194.

35. Abdou Saad El-Tabl; Rifaat Abed Wahaab; Younes, a. S. M.; , Downflow Hanging Sponge (DHS) reactor as a novel post treatment system for municipal wastewater. *Life Science Journal* **2013,** *10* (3), 409-414.

36. Onodera, T.; Matsunaga, K.; Kubota, K.; Taniguchi, R.; Harada, H.; Syutsubo, K.; Okubo, T.; Uemura, S.; Araki, N.; Yamada, M.; Yamauchi, M.; Yamaguchi, T., Characterization of the retained sludge in a down-flow hanging sponge (DHS) reactor with emphasis on its low excess sludge production. *Bioresource Technology* **2013,** *136*, 169-175.

37. Kubota, K.; Hayashi, M.; Matsunaga, K.; Iguchi, A.; Ohashi, A.; Li, Y.-Y.; Yamaguchi, T.; Harada, H., Microbial community composition of a down-flow hanging sponge (DHS) reactor combined with an up-flow anaerobic sludge blanket (UASB) reactor for the treatment of municipal sewage. *Bioresource Technology* **2014,** *151* (0), 144-150.

38. Narihiro, T.; Kim, N. K.; Mei, R.; Nobu, M. K.; Liu, W. T., Microbial community analysis of anaerobic reactors treating soft drink wastewater. *PLoS ONE* **2015,** *10* (3), e0119131.

39. Tamaki, H.; Wright, C. L.; Li, X. Z.; Lin, Q. Y.; Hwang, C. C.; Wang, S. P.; Thimmapuram, J.; Kamagata, Y.; Liu, W. T., Analysis of 16S rRNA amplicon sequencing options on the Roche/454 next-generation titanium sequencing platform. *Plos One* **2011,** *6* (9), e25263.

40. Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Pena, A. G.; Goodrich, J. K.; Gordon, J. I.; Huttley, G. A.; Kelley, S. T.; Knights, D.; Koenig, J. E.; Ley, R. E.; Lozupone, C. A.; McDonald, D.; Muegge, B. D.; Pirrung, M.; Reeder, J.; Sevinsky, J. R.; Tumbaugh, P. J.; Walters, W. A.; Widmann, J.; Yatsunenko, T.; Zaneveld, J.; Knight, R., QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **2010,** *7* (5), 335-336.

41. Caporaso, J. G.; Bittinger, K.; Bushman, F. D.; DeSantis, T. Z.; Andersen, G. L.; Knight, R., PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **2010,** *26* (2), 266-267.

42. Haas, B. J.; Gevers, D.; Earl, A. M.; Feldgarden, M.; Ward, D. V.; Giannoukos, G.; Ciulla, D.; Tabbaa, D.; Highlander, S. K.; Sodergren, E., Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome research* **2011,** *21* (3), 494-504.

43. Price, M. N.; Dehal, P. S.; Arkin, A. P., FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one* **2010,** *5* (3), e9490.

44. Colwell, R. K.; Chao, A.; Gotelli, N. J.; Lin, S. Y.; Mao, C. X.; Chazdon, R. L.; Longino, J. T., Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *J Plant Ecol-Uk* **2012,** *5* (1), 3-21.

45. Good, I. J., The Population Frequencies of Species and the Estimation of Population Parameters. *Biometrika* **1953,** *40* (3-4), 237-264.

46. Braak, C. t.; Šmilauer, P., CANOCO reference manual and CanoDraw for Windows user's guide: software for canonical community ordination (version 4.5). *Section on Permutation Methods. Microcomputer Power, Ithaca, New York* **2002**.

47. Detti, A.; Blefari Melazzi, N.; Salsano, S.; Pomposini, M. In *CONET: a content centric inter-networking architecture*, Proceedings of the ACM SIGCOMM workshop on Information-centric networking, ACM: 2011; pp 50-55.

48. Park, N.; Kwon, B.; Kim, I. S.; Cho, J. W., Biofouling potential of various NF membranes with respect to bacteria and their soluble microbial products (SMP): Characterizations, flux decline, and transport parameters. *Journal of Membrane Science* **2005,** *258* (1-2), 43-54.

49. Ni, B. J.; Rittmann, B. E.; Yu, H. Q., Soluble microbial products and their implications in mixed culture biotechnology. *Trends in Biotechnology* **2011,** *29* (9), 454-463.

50. Liang, S.; Liu, C.; Song, L. F., Soluble microbial products in membrane bioreactor operation: Behaviors, characteristics, and fouling potential. *Water Research* **2007,** *41* (1), 95-101.

51. Aquino, S. F.; Stuckey, D. C., Characterization of soluble microbial products (SMP) in effluents from anaerobic reactors. *Water Science and Technology* **2002,** *45* (10), 127-132.

52. Zhou, W. L.; Wu, B. T.; She, Q. H.; Chi, L.; Zhang, Z. J., Investigation of soluble microbial products in a full-scale UASB reactor running at low organic loading rate. *Bioresource Technology* **2009,** *100* (14), 3471-3476.

53. Mesquita, P. L.; Aquino, S. F.; Xavier, A. L. P.; da Silva, J. C. C.; Afonso, R. C. F.; Silva, S. Q., Soluble microbial product (SMP characterization in bench-scale aerobic and anaerobic CSTRs under different operational conditions *Brazilian Journal of Chemical Engineering* **2010,** *27* (1), 101-111.

54. Wu, B. T.; Zhou, W. L., Investigation of soluble microbial products in anaerobic wastewater treatment effluents. *Journal of Chemical Technology and Biotechnology* **2010,** *85* (12), 1597-1603.

55. Aquino, S. F.; Stuckey, D. C., The effect of organic and hydraulic shock loads on the production of soluble microbial products in anaerobic digesters. *Water Environment Research* **2004,** *76* (7), 2628-2636.

56. Huang, G. T.; Jin, G.; Wu, J. H.; Liu, Y. D., Effects of glucose and phenol on soluble microbial products (SMP) in sequencing batch reactor systems. *International Biodeterioration & Biodegradation* **2008,** *62* (2), 104-108.

57. Kuo, W. C.; Parkin, G. F., Characterization of soluble microbial products from anaerobic treatment by molecular weight distribution and nickel-chelating properties. *Water Research* **1996,** *30* (4), 915-922.

58. Ni, B. J.; Rittmann, B. E.; Fang, F.; Xu, J. A.; Yu, H. Q., Long-term formation of microbial products in a sequencing batch reactor. *Water Research* **2010,** *44* (13), 3787-3796.

59. Jarusutthirak, C.; Amy, G., Understanding soluble microbial products (SMP) as a component of effluent organic matter (EfOM). *Water Research* **2007,** *41* (12), 2787-2793.

60. Roesch, L. F. W.; Fulthorpe, R. R.; Riva, A.; Casella, G.; Hadwin, A. K. M.; Kent, A. D.; Daroub, S. H.; Camargo, F. A. O.; Farmerie, W. G.; Triplett, E. W., Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME Journal* **2007,** *1* (4), 283-290.

61. Kwon, S.; Kim, T.-S.; Yu, G. H.; Jung, J.-H.; Park, H.-D., Bacterial community composition and diversity of a full-scale integrated fixed-film activated sludge system as investigated by pyrosequencing. *J Microbiol Biotechnol* **2010,** *20* (12), 1717-1723.

62. McLellan, S. L.; Huse, S. M.; Mueller-Spitz, S. R.; Andreishcheva, E. N.; Sogin, M. L., Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environmental Microbiology* **2010,** *12* (2), 378-392.

63. Zhang, T.; Shao, M. F.; Ye, L., 454 Pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. *Isme Journal* **2012,** *6* (6), 1137-1147.

64. Miura, Y.; Watanabe, Y.; Okabe, S., Significance of Chloroflexi in performance of submerged membrane Bioreactors (MBR) treating municipal wastewater. *Environmental Science & Technology* **2007,** *41* (22), 7787-7794.

65. Xia, Y.; Kong, Y.; Nielsen, P. H., In situ detection of protein-hydrolysing microorganisms in activated sludge. *FEMS Microbiology Ecology* **2007,** *60* (1), 156-165.

66. Xia, Y.; Kong, Y.; Thomsen, T. R.; Nielsen, P. H., Identification and ecophysiological characterization of epiphytic protein-hydrolyzing Saprospiraceae ("Candidatus Epiflobacter" spp.) in activated sludge. *Appl. Environ. Microbiol.* **2008,** *74* (7), 2229-2238.

67. Ju, F.; Zhang, T., Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. *The ISME journal* **2015,** *9* (3), 683-695.

68. Lewin, R. A., *Saprospira grandis*: a flexibacterium that can catch bacterial prey by "Ixotrophy". *Microb Ecol* **1997,** *34* (3), 232-236.

69. Saw, J. H. W.; Yuryev, A.; Kanbe, M.; Hou, S. B.; Young, A. G.; Aizawa, S. I.; Alam, M., Complete genome sequencing and analysis of *Saprospira grandis* str. *Lewin*, a predatory marine bacterium. *Standards in Genomic Sciences* **2012,** *6* (1), 84-93.

70. Cottrell, M. T.; Kirchman, D. L., Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacter cluster consuming low-and high-molecular-weight dissolved organic matter. *Appl. Environ. Microbiol.* **2000,** *66* (4), 1692-1697.

71. Arenskotter, M.; Broker, D.; Steinbuchel, A., Biology of the metabolically diverse genus *Gordonia*. *Appl. Environ. Microbiol.* **2004,** *70* (6), 3195-3204.

72. Imhoff, J. F., The family Ectothiorhodospiraceae. In *The Prokaryotes*, Springer: 2006; pp 874-886.

73. Yu, K.; Zhang, T., Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge. *PloS one* **2012,** *7* (5), e38183.

74. Machdar, I.; Sekiguchi, Y.; Sumino, H.; Ohashi, A.; Harada, H., Combination of a UASB reactor and a curtain type DHS (downflow hanging sponge) reactor as a cost-effective sewage treatment system for developing countries. *Water Science and Technology* **2000,** *42* (3-4), 83-88.

# CHAPTER 4: PYLOGENETIC AND FUNCTIONAL CHARACTERIZATION OF THE MICROBIAL COMMUNITY DEGRADING SOLUBLE MICROBIAL PRODUCTS IN A DHS REACTOR USING METAGENOMIC AND METATRANSCRIPTOMIC APPROACHES

## 4.1 Abstract

Soluble microbial products (SMP), ubiquitously found in bioprocesses, have been identified as a main cause for the decreasing efficiency of water and wastewater treatment systems. Despite recent attempts for the biological removal of SMP to control the negative impacts of their accumulation, the mechanisms of SMP degradation and the roles of the microbial community remain unsolved. To gain a better understanding of biological SMP degradation in a down-flow hanging sponge reactor that treats SMP, and to profile the active functions of the microbial populations therein, the metabolically active microbial communities were assessed by comparative metagenomic and metatranscriptomic analyses. Taxonomic classification identified that the dominant microbial populations shifted from *Sphingobacteriales*, *Flavobacteriaceae*, and *Cytophaga*-relatives to *Saprospiraceae*, *Dechloromonas*, and *Geobacter*-relatives with increasing SMP loading. In the lower part of the reactor, other populations, besides *Sphingobacteriales*, *Opitutus,* and *Nitrospira*, contributed to a significant distribution. Global functionality and gene expression annotation based on SEED subsystems exhibited that, despite these phylogenetically disparate microbial communities with SMP loading, a functional convergence was observed for the stratified SMP degradation, including amino acids and derivatives, carbohydrates, and protein metabolisms. Further, functional gene expression analysis, focusing on carbohydrate-active enzymes with respect to assembled genome bins, revealed that cell associated enzyme-related genes, specific to polysaccharide components of peptidoglycan, were significantly represented by the dominant assembled genome bins in *Bacteroidetes*. The observations reflect that the microbial communities, degrading SMP in the down-flow hanging sponge (DHS) reactor, were selectively enriched for the utilization of detrital cell structural components, such as peptidoglycan and lipopolysaccharides, that contributed to biomass associated products (BAP).

## 4.2 Introduction

Soluble microbial products (SMP) are ubiquitously present in water and wastewater treatment processes based on mixed culture biotechnology, contributing to predominant constituents of the organic fraction of discharged effluent.[1] Since SMP are soluble organic compounds produced from microbial metabolism and decay, the predominant presence of them is inevitable in the bioprocesses. SMP primarily consist of polysaccharides and proteins [2]. Depending on the mechanism of SMP formation, they are classified into two sub-groups, utilization-associated products (UAP) produced from substrate utilization during biomass growth and biomass-associated products (BAP) derived from cell lysis and decay.[1-2] SMP were reported to be a cause of negative impacts to the treatment processes since they contributed to effluent chemical oxygen demand (COD),[3-5] effluent toxicity,[6] and a potential precursor of disinfection by-products.[7-8] Their accumulation not only hinders efficient respiration, flocculation, and settling ability of activated sludge (AS) by deforming the physical properties of the AS, but it also inhibits nitrification efficiency. Last, but most importantly, SMP are a major obstacle impeding application of membrane bioreactors (MBRs) for water reclamation causing membrane fouling.[9-11]

Numbers of studies were conducted to investigate the characteristics of SMP, their adsorption and coagulation properties, and effects by operational conditions in a way of developing strategies to reduce the accumulation of SMP in the bioprocesses and enhancing the understanding of them.[12-17] Nevertheless, a strategy providing a long-term stable application for SMP removal has not been established likely because the composition and property of SMP vary depending on substrates, sludge mixtures, and operational conditions.[2, 12, 18-21]

Several studies have confirmed that various microorganisms were capable of utilization of SMP as their sole energy and carbon source, suggesting biological removal of SMP as alternatives. First, it was reported that microbial products derived from nitrifiers without organic substrates could be used as carbon sources for the growth of heterotrophs, mainly *Cytophaga-Flavobacterium-Bacteroides* (CFB) cluster.[22-25] Next, effective reduction of SMP released into the MBRs by inoculating either a specific microorganism or microbial consortium was reported.[26-27] Miura and Okabe (2008) also showed that the population dynamics of *Chloroflexi* was inversely related to the SMP concentration in the MBR.[28]

Based on these findings, in our previous study, a long-term and stable removal of SMP were demonstrated in a DHS reactor using selectively enriched microbial consortia for SMP degradation.[29] The temporal and spatial diversity and dynamics of the microbial community were characterized using 16S rRNA-based pyrosequencing, and revealed that highly specialized microbial community with predominant populations related to *Bacteroidetes* had been enriched. Despite these community-wise informative findings, the microbial functions essential for degrading SMP remain to be elucidated. Metagenomic and metatranscriptomic sequencing based on Next Generation Sequencing (NGS) enables to fill the gaps of the ecological roles of microorganisms detected in the DHS reactor by providing microbial structure, functional potential, and identified gene expression.[30-33] An in-depth resolution of the genetic information using coupled metagenomic and metatranscriptomic approaches would be not only helpful to unveil the unique characteristics of selectively enriched microbial communities in the DHS reactor, but beneficial to develop biological strategies to control the accumulation of SMP in the system. The aims of this study, therefore, were i) to determine the complementary phylogenetic characteristic of the microbial community structure to that of targeted 16S rRNA gene sequencing; ii) to explore microbial metabolic potential and expression; and iii) to disclose the active roles of various microbial populations involved in SMP degradation.

## 4.3 Material and methods

### 4.3.1 DHS microbial consortia

Biomass was collected from the DHS reactor treating SMP generated from the anaerobic packed-bed bioreactors at day 648 in Phase III and day 798 in Phase V (Figure 4.1). The detailed operational conditions and the system performance were described in elsewhere.[29] The samples collected from the upper (depth, 0.16 m) and lower (depth, 0.94 m) part of the DHS reactor at day 648 were named U648 and L648, respectively. The samples at day 798 were collected from the upper part of the reactor, named U798.

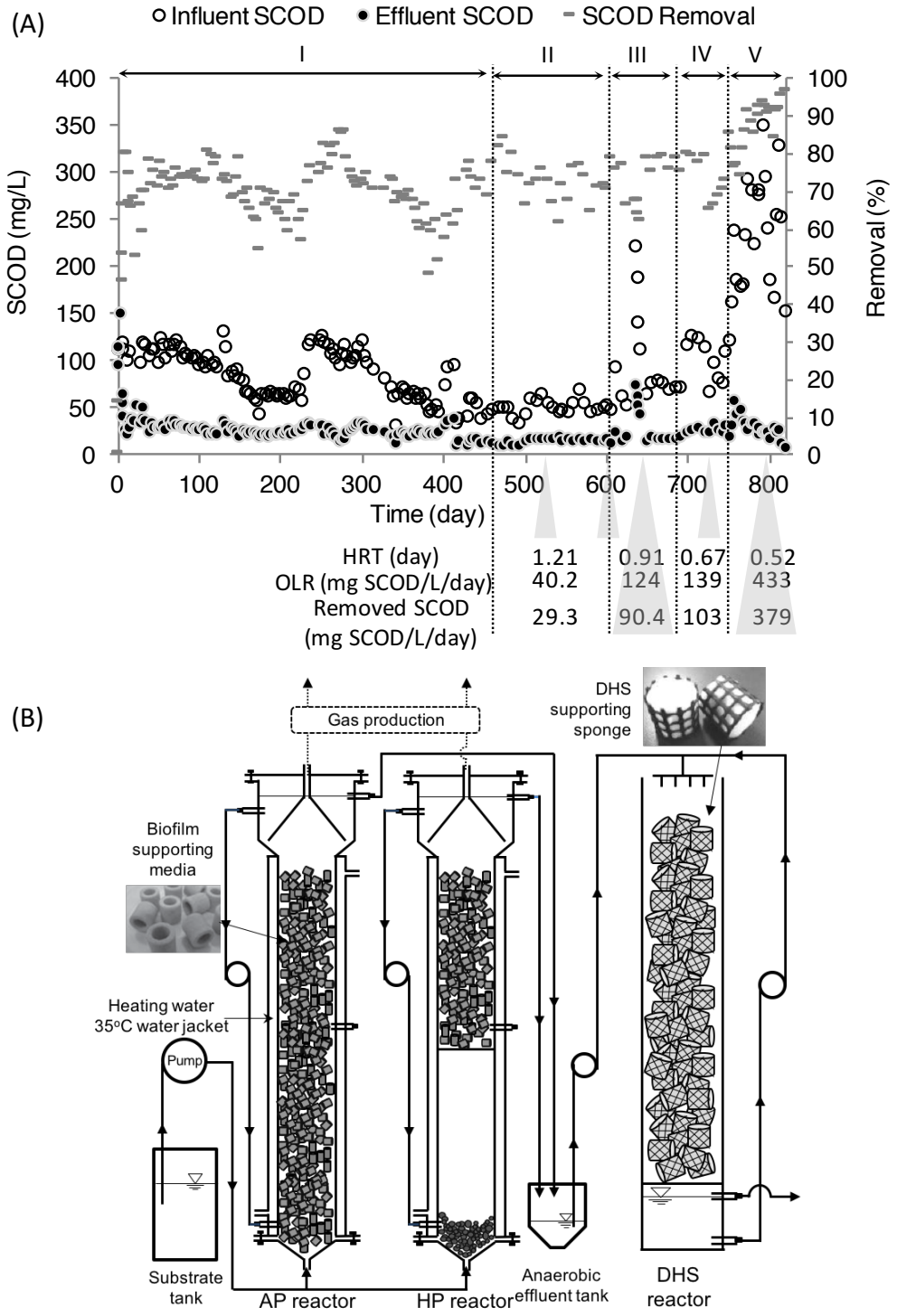



**Figure 4.1** Sampling for metagenomic and metatranscriptomic analyses. (A) The operational factors in Phase II, III, IV, and V were shown with the performance of the DHS reactor. (B) sampling locations of the DHS reactor was indicated in the schematic diagram of the system.

### 4.3.2 DNA and RNA extraction

The biomass in the sponge was suspended in 25 ml of 1x phosphate buffered saline (PBS) solution by vortexing, pelleted by centrifugation (10000 rpm, 3 min), and stored in -80$^{o}$C. Genomic DNA was extracted using the FastDNA SPIN kit for soil (MP biomedicals, USA) according to the manufacturer's instruction and resuspended in TE (10 mM Tris-HCl with 1 mM EDTA) buffer at the last step. The samples collected in Phase V was duplicated. DNA in one of them, U798_1, was extracted using the same kit as the previous samples were treated, and DNA in the other, U798_2, was extracted following the established protocol (Bacterial genomic DNA isolation using CTAB, http://my.jgi.doe.gov/general/). The concentration and the purity of DNA were assessed using a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

Three sponges were collected for biological triplicates of RNA extraction at each sampling time and location. The biomass was suspended in 25ml of soluble reactor effluent filtered with 0.22 μm filters (Millex-GP, Millipore, MA, USA) by vortexing at 4 $^{o}$C. Two volumes of RNAprotect bacterial reagent (Qiagen, CA) were added into 2ml of the suspended biomass, immediately mixed by vortexing, and incubated for 5 min at room temperature. The pellets were harvested by centrifugation (7000 rpm, 10 min) at 4 $^{o}$C. For enzymatic lysis of the biomass, the pellets were resuspended with 200 μL TE (30 mM Tris-Cl, 1 mM EDTA, pH 8.0) buffer containing lysozyme (15 mg/ml) and 20 μL Proteinase K (Qiagen, CA) by vortexing and incubated at room temperature for 15 min. RNA was extracted using the RNeasy Mini Kit (Qiagen, CA). Genomic DNA during the RNA preparation was excluded by using RNase-free DNase I (Qiagen, CA. The concentrations of RNA in the samples were measured, which were 103.7 ng/ml, 86.8 ng/ml, and 55.0 ng/ml for the triplicates of U648, 54.4 ng/ml, 58.5 ng/ml, and 49.8 ng/ml for L648, 49.8 ng/ml, 221.3 ng/ml, and 165.5 ng/ml for U798 by a Nanodrop 1000 spectrophotometer. The integrity of the extracted DNA and RNA was verified by running 100ng of each sample with a DNA molecular weight marker (1kb DNA ladder, Promega) on a 1% denaturing formaldehyde agarose gel for electrophoresis prior to sequencing (Figure C.1).

### 4.3.3 DNA and cDNA library construction and sequencing

The extracted DNA and RNA samples were submitted to the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign (IL, USA) for sequencing and DNA and cDNA library construction. The DNA libraries were constructed for each sample using the TruSeq DNA sample prep kit (Illumina Inc. San Diego, CA), and the pooled libraries were quantitated by qPCR and sequenced on one lane for 101 cycles from each end of the fragments on a HiSeq2000 sequencer (Illumina, San Diego, CA, USA) using a TruSeq SBS sequencing kit v3 (Illumina Inc. San Diego, CA). The genomic libraries were analyzed with Casava1.8.2. The triplicate RNA libraries, after removal of rRNA with the Ribo-Zero™ rRNA Removal Kit (Meta-Bacteria, Illumina, WI, USA), were prepared with the TruSeq Stranded RNA Sample Prep kit (Illumina Inc. San Diego, CA). The rest of the process was the same as described for the DNA library construction.

### 4.3.4 Metagenomic and metatranscriptomics sequence analysis

#### 4.3.4.1 Quality control, rRNA subtraction, and 16S rRNA gene reconstruction

The raw genomic and transcriptomic reads were trimmed using a Q 13 Phred quality score cutoff and screened with minimum length 50 bp cutoff using SolexaQA v3.1.2[34] for a quality control (QC) (Table C.1). The post-QC reads have been deposited in MG-RAST (http://metagenomics.anl.gov/?page=MetagenomeProject&project=9993, of which MG-RAST projetc ID (mgp9993) and MG-RAST library IDs were listed in Table C.1. Sequences encoding rRNA genes in the genomic and transcriptomic datasets were separated from the coding-DNA sequences and non-rRNA sequences in the post-QC datasets by SortMeRNA v.2.0[35] with default settings against SILVA SSU and LSU databases (release 119).[36] The post QC genomic datasets were used to reconstruct full length of 16S rRNA using EMIRGE[37] with 0.99 OTU identity and default settings for the rest of conditions to reveal the microbial community compositions. The reconstructed genes for the three datasets were combined and subjected to an operational taxonomic units (OTUs) assignment by the UCLUST algorithm (≥ 97% pairwise identity) in Quantitative Insights Into Microbial Ecology (QIIME).[38] The representative sets of the OTUs were formed and Chimeric sequences were removed by Chimera Slayer.[39] PyNAST[40] was used to align the representative sequences to the Greengenes imputed core dataset. The phylogenetic

affiliation of the representative sequences was classified in the Greengenes ARB database (Greengenes_16S_2011_1.arb) using ARB parsimony method. A phylogenetic tree for the aligned sequences with their neighboring sequences was built using the neighbor-joining algorithm with Jukes-Cantor correction. Bootstrap values were calculated based on 1000 replications. The relative abundance of the representative sequences in each genomic dataset was expressed in percentage of the raw sequencing reads mapped to the representative sequences using Blastn with a cutoff of 95% identity and the parameters of X = 150, q = −1 and F = F at default settings.

#### 4.3.4.2 Assembly of the metagenomic datasets

The Velvet[41], SOAPdenovo2[42], and IDBA-UD[43] assemblers were used to preassemble Illumia short reads of each dataset into contigs using different k-mer sizes (49-65 for Velvet; 49-83 for SOAPdenovo2; 45-95 for IDBA-UD). The k-value used for the preassemblies and their statistics were described in the Table C.2. The preassemblies for each dataset that provided longer maximum contig sizes and n50s were chosen and assembled in to one final assembly using Newbler v2.9.[44] In the process of getting the final assembly, the U798_1 and U798_2 datasets were merged since separated preliminary analysis showed that the two datasets were proved to be exact replicates.

#### 4.3.4.3 Protein encoding gene prediction

The assembled contigs longer than 300 bp were submitted to the MG-RAST pipeline[45] and subjected to protein encoding genes (PEG) prediction (MG-RAST ID, 4579439.3).[46] Taxonomic annotation was performed against the SEED database using a Best Hit Classification approach with a maximum e-value cutoff of 1E-5, a similarity cutoff of 60%, and a minimum alignment length of 15 measured in amino acids for protein and base pairs for RNA databases. Functional annotation was conducted by comparison to the subsystems using a hierarchical classification algorithm with a maximum e-value cutoff of 1E-5, a similarity cutoff of 60%, and a minimum alignment length of 15 amino acids. The PEGs longer than 300 bp were applied to the further expression analysis.

#### 4.3.4.4 Normalization for expression analysis

For gene expression analysis, the relative abundance of PEGs was estimated by following steps; the coding-DNA and non-rRNA sequences were mapped to the PEGs using Blastn with at least 95% identity and 50% query length coverage and the parameters of X = 150, q = −1 and F = F at default settings. The length of the coding DNA and non-RNA sequences mapped to each PEG were summed and divided by the length of the PEG to calculate the genomic and transcriptomic coverage of each PEG. The coverage per 1Mb of sequences was computed to provide the genomic relative abundance and the transcriptional activity of a gene independent of the size of the datasets. The ratio of the transcriptional relative abundance to the genomic relative abundance was calculated to estimate the absolute transcriptional activity regardless of the genomic relative abundance in the microbial community.

#### 4.3.4.5 Assembled genome bins from the metagenomic datasets

Assembled contigs greater than 1000 bases were subjected to cluster into genome bins based on metagenomic read coverage, tetranucleotide frequency, and occurrence of essential single copy genes, using MaxBin (v 2.0)[47] MetaBat,[48] and MetaWatt (v 3.5.2).[49] The overlapped contigs among the clustered contigs using each binning tool were manually extracted to group into draft bins. CheckM (v 1.0.5)[50] was used to estimate the completeness and contamination of the draft genomic bins based on number of single-copy marker genes identified in each bins. Bins with more than 10% contamination or less than 20% completeness were discarded from further analyses. The taxonomic affiliation of the assembled genome bins was carried out using AMPHORA2,[51] and the resulted marker lineage was reported when 75% of the classifications reached a consensus taxonomic level.[52] A genome-wide phylogenetic analysis of the assembled genome bins was conducted using PhyloPhlAn.[53] The predicted protein encoding genes for the assembled genome bins were identified and aligned on a subset of 400 conserved protein sequences. The assembled genome bins and reference genomes were integrated into the tree of life with 3,171 microbial genomes. The bins were submitted under the MG-RAST project (ID: mgp9993). The annotated information of each assembled genome bin was retrieved from RAST[54] and PATRIC.[55]

#### 4.3.4.6 Carbohydrate-active enzyme annotation

The clustered contigs for the major assembled genome bins were subjected to gene prediction using FragGeneScan v1.30.[56] A carbohydrate-active enzyme (CAZy)-family specific hidden Markov model (HMMs) were downloaded from the dbCAN database (http://csbl.bmb.uga.edu/dbCAN/)[57] and used in screening amino acid sequences of the predicted ORFs for similarity to 192 families (9 auxiliary activity (AA), 32 carbohydrate-binding module (CBM), 16 carbohydrate esterase (CE), 79 glycoside hydrolase (GH), 42 glycosyl transferase (GT), and 12 polysaccharide lyase (PL) families) in the CAZy database.[58] The protein sequences were compared to the profile HMMs by employing hmmscan of the HMMER 3.1.b2 software package (hmmer.org). As instructed in the dbCAN database, the overlapping hits and the hits with a higher e-value and a coverage of less than 30% of the respective HMM were removed. The remaining hits were processed with an e-value cutoff of $1e^{-5}$ for alignments longer than 80 amino acids and $1e^{-3}$ for alignments shorter than 80 amino acids. Duplicated hits found in the CAZy-families were manually removed. The major assembled genome bins based on the hits of the CAZy genes and families were clustered, and the values of the hits were plotted using the heatmap.2 function of the gplots package (v 3.0.1) in R. To compare the genomic encoding and transcriptional expression of the CAZymes in the assembled genome bins, the relative abundance of the CAZymes was normalized to 1Mb of sequences as described in the section, 4.2.4.4.

## 4.4 Results and discussion

### 4.4.1 Microbial phylogenetic community structure in the DHS reactor

Out of 820 days of the DHS reactor operation for SMP degradation, three biomass samples U648 and L648 in Phase III and U798 in Phase V were collected from the supporting media for metagenomic and metatranscriptomics analyses. Because of the decreasing HRT in the system, the organic loading rate (OLR) to the DHS reactor in Phase V, 433 mg SCOD/L/day, was almost four times higher than that in Phase III, 124 mg SCOD/L/day. Along with the OLR increase, the SCOD removal in the DHS reactor also increased about four times (Figure 4.1). The detailed operational conditions and the system

performance were described in elsewhere.[29] The analytical workflow of metagenomic and metatranscriptomic datasets was illustrated in Figure C.2. The Illumina sequencing of SMP degrading microbial community provided paired-end metagenomic reads (100 bp; a range of fragment size, 380bp to 640bp; 0.9 x $10^8$ reads for U648, 1.1 x $10^8$ reads for L 648, and 2.0 x $10^8$ reads for U798) and single-end metatranscriptomic reads (100 bp; a range of fragment size, 130bp to 480bp; average $1.5x10^7$, $1.5x10^7$, and $1.7x10^7$ reads for U648, L648, and U798, respectively) (Table C.1). Approximately 0.2% of post QC genomic reads were sorted out as rRNA sequences in the genomic dataset (Table C.1)[35] and blasted to the EMIRGE-based reconstructed 16S rRNA gene sequences[37] and aligned to the reference 16S sequences in the SILVA database (release 119).[59] A strong consistency between the community structures based on EMIRGE and SILVA was observed in all three datasets (Pearson correlation coefficient $r > 0.9$). The most abundant bacteria found at the phylum level were *Bacteroidetes* and *Proteobacteria* in all datasets (Figure C.3). At the class level, *Sphingobacteria* (EMIRGE, 25.0% and raw 16S rRNA reads, 19.3%) in U648, *Alphaproteobacteria* (EMIRGE, 19.9% and raw 16S rRNA reads, 19.4%) in L648, and *Betaproteobacteria* (EMIRGE, 31.8% and raw 16S rRNA reads, 26.5%) in U798 were the most dominant populations.

To see a consensus of the microbial community structures between 16S rRNA gene-based PCR-dependent and independent assays, a phylogenetic tree was constructed with dominant EMIRGE-constructed sequences and representative operational taxonomic units (OTUs) of the amplified 16S rRNA genes by pyrosequencing in the previous study[29] (relative abundance >1% of the total number of bacterial 16s rRNA gene sequences in any sample) (Figure C.4). Most of the EMIRGE-constructed sequences constituted a deep branch together with paired OTUs. Similar distribution in the relative abundance of the paired EMIRGE-constructed sequences and OTUs was observed (correlation coefficient $r=0.88$ for U648, $r=0.71$ for L648, and $r=0.72$ for U798), suggesting that the microbial community structures derived from the metagenomes concurred with the results of pyrosequencing despite an inherent bias of 16S rRNA gene amplification. In U648, *Sphingobacteriales*-related members (DHS_OTU 1435, 7.5% and DHS_Emg 87, 8.2%) were most abundant followed by *Flavobacteriaceae*-related members (DHS_OTU 246, 4.1%, DHS_Emg 78, 5.8%, DHS_OTU 567, 3.6%, and DHS_Emg 19, 4.3%) and

*Cytophaga*-related members (DHS_OTU 897, 5.6% and DHS_Emg 72, 3.6%). The abundance of these dominant members in *Bacteroidetes* decreased to be minor except for the *Cytophaga* relatives in U798. Shifts in the abundant members to *Saprospiraceae*-related members (DHS_OTU 660, 9.6% and DHS_Emg 21, 3.6%) and another *Sphingobacteriales* relatives (DHS_OTU 150, 9.4% and DHS_Emg 121, 4.3%) were observed. Furthermore, *Dechloromonas*-related members (DHS_OTU 1496, 4.6% and DHS_Emg 99, 5.8%) in *Rhodocyclaceae* and *Geobacter*-related members (DHS_OTU 331, 4.7% and DHS_Emg 12, 3.2%) increased in U798. Unlike the microbial community in the upper part of the reactor, L648 indicated a more diverse community with relatively even distribution such as high abundances in *Opitutus* (DHS_OTU 666, 3.7% and DHS_Emg 117, 7.1%) and *Nitrospira* (DHS_OTU 248, 6.6% and DHS_Emg 51, 4.5%) besides *Sphingobacteriales* relatives.

Microbial community composition from metagenomes is also assessed based on taxonomic homology of PEGs.[60-61] In comparison with the community compositions derived from the 16S rRNA sequence based analyses, a consistent taxonomic composition was observed in U648 and U798 metagenomic datasets (Figure C.5). *Spingobacteriales* (11.9%), *Flavobacteriales* (7.1%), and *Cytophagles* (8.6%) were the most abundant order groups in U648, and the dominant orders shifted from these Bacteroidetes to *Desulforomonadales* (7.4%) and *Rhodocyclales* (14.1%), predominantly *Geobacter* and *Dechloromonas* in U798, respectively. In contrast, in L648 *Burkholeriales* (8.4%) and *Rhizobiales* (11.2%) were the most abundant members rather than *Nitrospira* and *Opitutus*. Meanwhile, we observed the taxonomic origins of the non-rRNA transcript sequences to display the transcriptional activity in the microbial community determined based on the metagenomes. In the two datasets form the upper part of the reactor, the taxonomic composition from the metatranscriptomes showed a similar distribution to those from the metagenomes. However, in L648 the highest transcriptional activity in *Solibactrales*, predominantly *Candidatus Solibacter*, was detected.

### 4.4.2 Global functionality and expressions of the DHS microbial communities

58.4%, 51.5% and 58.5% of the 67, 77, and 143 million reads of total metagenomic sequences in the three datasets (in order of U648, L648, and U798) were included in the de novo assembly integrated by MetaVelvet, SoupDenovo2, IDBA-UD, and Newbler (Table

C.2). The assembly generated a metagenome of 45,392 contigs with total sequences of 440.6 Mb, N50 of 25,560 bp, and N90 of 3,186 bp. 45,037 assembled contigs longer than 300 bp were subjected to MG-RAST functional annotation, predicting 272,083 ORFs. Among these ORFs, 200,515 were annotated with putative protein functions and 166,555 were assigned to a functional classification. of them 57.0% of features were classified as SEED Subsystems-based PEGs (Table C.4). Among the PEGs by Subsystems, 19.6-49.6% showed transcriptional activity with at least one aligned non-rRNA sequence (Table C.6).

Highly encoded and expressed metabolic pathways of the DHS microbial communities at the SEED Subsystems level 1 were exhibited (Figure 4.2). Genomically, the two systems, Cluster-based Subsystems (11.8-12.2%) and Carbohydrates (9.3-10.5%) were most abundantly encoded in all three metagenomic datasets, followed by Amino acids and derivatives (8.0-8.3%) and Protein metabolism (7.6-7.9%). On the other hand, the transcriptional pattern indicated that the systems, Protein metabolism (17.1-22.4%), Clustering-based subsystems (9.5-10.5%), and Carbohydrates (7.4-8.2%) were highly expressed in the two dataset from the upper part of the reactor. In the metatranscriptomic dataset from the lower part of the reactor, the systems, Amino acids and derivatives (27.1%) and Protein metabolism (12.5%) were the most abundantly expressed systems. An interesting clustering observed in the pattern of SEED Subsystems level1 was that, among the metagenomes, the two libraries collected in Phase III were closely clustered than the other library in Phase V whereas, in the metatranscriptional analysis, the two triplicate datasets from the upper part of the reactor in Phase III and V were more closely clustered than the dataset collected from the lower part of the reactor at the same operational phase. This indicated that phylogenetically and genomically disparate microbial communities, U648 and U798, resulted in similar transcriptional consequences. The same clustering was observed in the detailed systemic categories of the metabolic pathway, SEED Subsystems at level 3 that were significantly abundant at the 98% confidence level (Figure C.6).

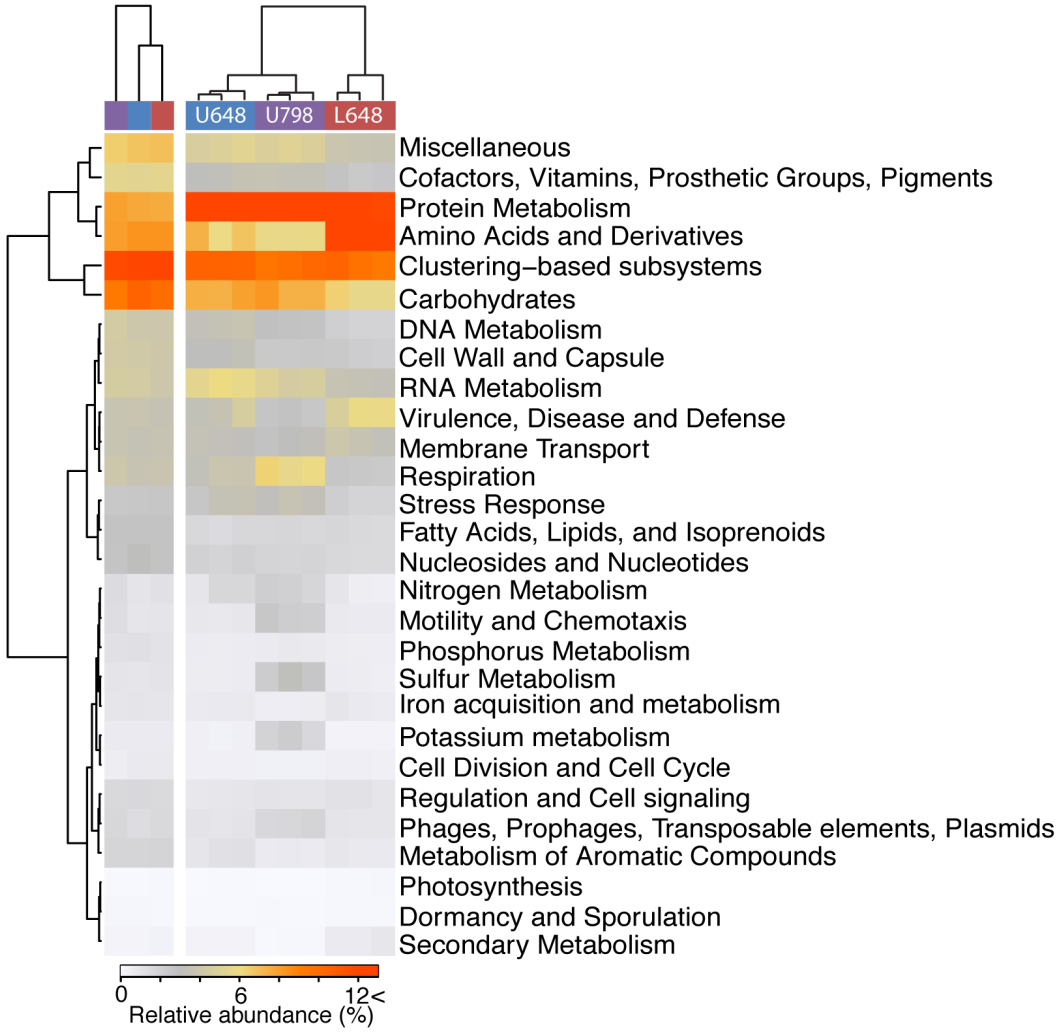



**Figure 4.2** Global analysis of metabolic potential and functional activities in the DHS communities. Clustering of the three metagenomic and triplicated metatranscriptomic datasets based on normalized relative abundance of SEED Subsystems level 1. Hierarchical clustering of the metagenomic and the metatranscriptomic datasets were separately conducted with Euclidean distance using R package (Stats v3.2.0).

Among the most enriched systems in the metagenomic and metatranscriptomic libraries, the transcriptional activity profiles of the three systems that are related to SMP catabolism, Amino acid and derivatives, Carbohydrates, and Protein metabolism, were further presented at the Subsystems level 3 for each datasets and the ten most dominant orders (Figure C.7 and Table C.7). In the sub-level of the amino acids derivatives system, alanine biosynthesis, predominantly branched-chain amino acid aminotransferase (EC 2.6.1.42), was the most actively expressed in L648 and more specifically by *Solibacterales*. Another function showing a substantial transcription was glutamate dehydrogenases (EC 1.4.1.2) by *Sphingobacteriales* especially in U648. Ornithine degradation was activated by Unclassified *Opitutae* in U648 and U798 and by *Bacteroidales* and *Clostridiales* in L648. Lysine utilization was found to be highly expressed by *Desulfuromonadales* and *Cytophagales* in U798. Threonine anaerobic utilization were activated by *Desulfuromonadales* mostly in L648.

In the subcategories of the Carbohydrates system, oligosaccharides and polysaccharides utilization transported from the outer membrane was overrepresented in U648; trehalose uptake and utilization that was mostly genes involved in glucose-specific phosphotransferase system (EC 2.7.1.69) showed high transcriptional activity together with fructose-bisphosphate aldolase (EC 4.1.2.13). An ABC transporter gene among N-acetylglucosamine catabolic operon was also overrepresented. The activity of alanine dehydrogenase (EC 1.4.1.1) to convert alanine to pyruvate in the system, pyruvate alanine serine interconversions, was induced by various orders such as *Flavobacteriales*, *Solibacterales*, *Desulfuromonadales*, *Cytophagales*, and *Bacteroidales*. Cellulosome, more specifically SusC like outer membrane binding protein for extracellular polysaccharides, was highly expressed in all three datasets especially by *Flavobacteriales*, *Rhizobiales, Cytophagales,* and *Bacteroidales*. In U798, catabolism of these extracellular carbohydrates by fermentation was highly activated by *Desulfuromonadales*. Since the profile was for transcriptional activity of the most abundant orders, the biosynthesis and processing-related subgroups in Protein biosynthesis were activated by all of the top ten orders. In the protein degradation-related subgroups, bacteria-driven proteolysis and proteasome were highly expressed in all datasets. Dipeptidases and aminopeptidases were significantly expressed by *Bacteroidales* in U648.

### 4.4.3 Potential encoding and expression of CAZy families

To further investigate the key metabolic features and physiology of the dominant microbial populations in the DHS reactor, a metagenomic binning was conducted to reconstruct assembled genome bins. Sixty nine high quality bins with less than 10% contamination and more than 20% completeness were observed.[52] Of them, 28 bins, which contributed top 50% of relative abundance of PEG in the metagenomic and metatranscriptomic datasets, were listed in Figure 4.4. AMPHORA2[51] software with 31 conserved bacterial phylogenetic protein marker genes were employed to identify a taxonomic affiliation of the major assembled genome bins; a total of 850 marker genes were assigned. 12 bins were affiliated with *Proteobacteria*, and the second most abundant taxonomic bin classification was *Bacteroidetes* with 7 bins. The rest of bins were classified into *Verrucomicrobia* (3), *Planctomycetes* (2), *Acidobacteria* (1), *Firmicutes* (1), *Gemmatinomonadetes* (1), and *Nitrospira* (1) (Figure 4.3 and Table C.8). The relative abundance of gene encoding and expression by the assembled genome bins basically reflected the population structure as observed in the phylogenetic analysis of the DHS microbial communities. The genome-wide phylogenetic analysis using PhyloPhlAn indicated that the four assembled genome bins (Bin87, Bin 55, Bin74, and Bin 09) in the *Bacteroidetes*, among the most assembled genome bins attributing PEG abundances with a high completeness, constituted a deep branch with *Haliscomenobacter hydrossis* DSM 1100 (IMG taxon ID: 2504756004). Bin08 and Bin 55 in *Bactroidetes* constructed a branch with *Chitinophaga pinensis* UQM 2034 (IMG taxon ID: 644736340). In the *betaproteobacteria*, *Dechloromonas aromatica* RCB (IMG taxon ID: 637000088) was closely clustered with Bin78 and Bin80. In the *gammaproteobacteria*, *Geobacter metallireducens* GS-15 (IMG taxon ID: 637000119) was clustered with Bin79.

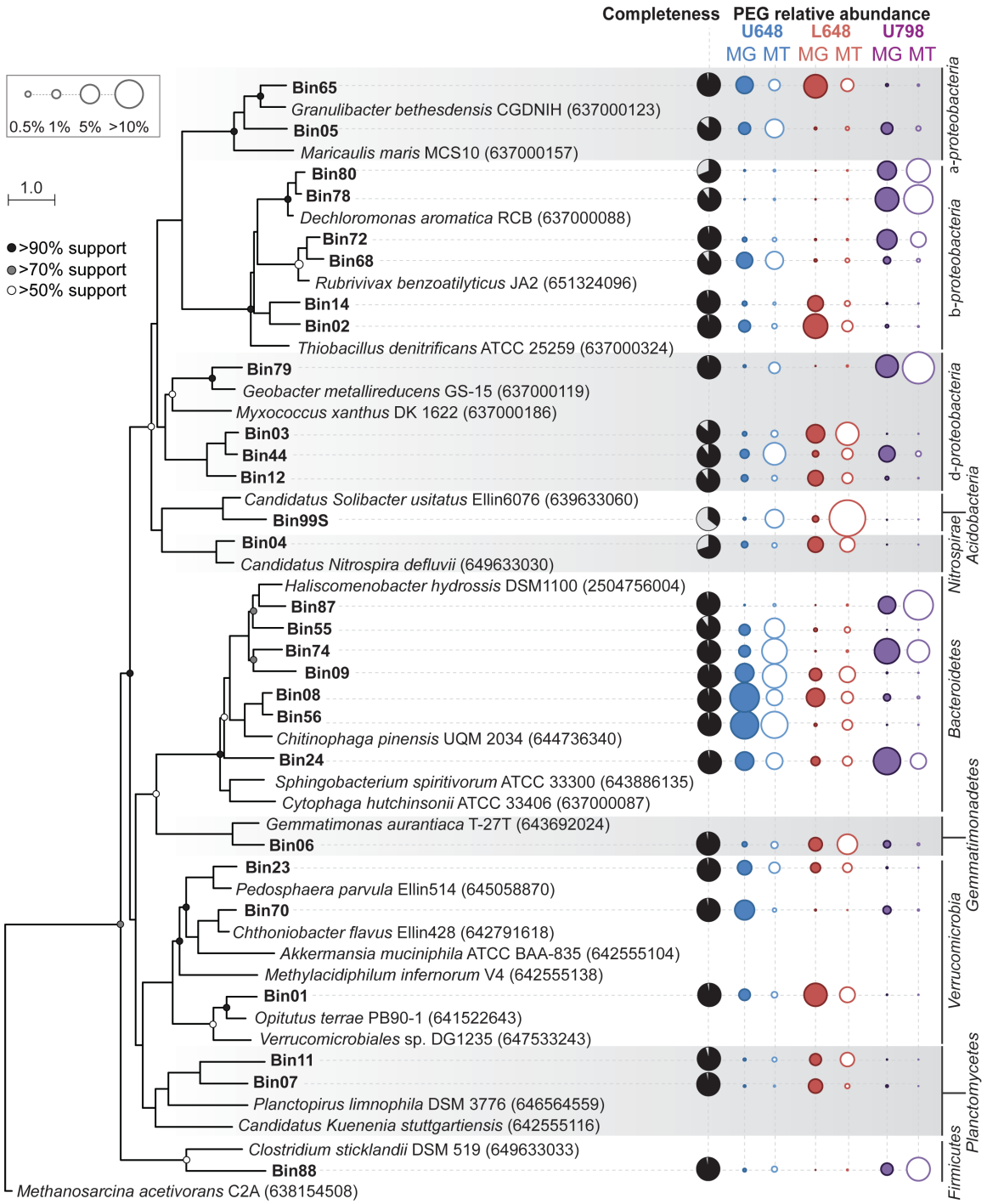



**Figure 4.3** The genome-wide phylogenetic analysis and the abundance profile of the major bins contributing top 50% of protein encoding gene relative abundance for each dataset. The phylogenetic tree was generated by PhyloPhlAn and iTOL from predictied protein sequences of the major bins and 3,171 other reference genomes (bootstrap 1000: >90% black node, >70% gray node, and >50% white node; IMG taxon ID of the reference genomes in parenthesis). MG refers to a metagenomic dataset, and MT refers to an average of the triplicate metatranscriptomic datasets. The genome completeness was shown as black pie charts.

Intrigued by the findings in the carbohydrate related metabolisms of the entire microbial communities, we analyzed the major 28 assembled genome bins with respect to the genomic potential and transcriptomic expression of hydrolytic enzymes involved in polysaccharide and glycan degradation. The profile hidden Markov model specifying CAZy database was used, offering a sequence-based family classification of enzymes involved in degradation and cleavage of various types of polysaccharides.[58] The total number of putative genes and the respective CAZy families from the assembled genome bins were listed, which the assembled genome bins were clustered for glycoside hydrolases based on (Figure C.8). The most genomically predicted enzymes belonged to CBM families 20, 32, 37, 40, 44, 50, 61, GH families 13, 23, 33, 74, 109, PL families 9, 22, CE families 1, 4, 10, and AA family 2, including the separate families of Cohesin and Dockerin (Figure 4.4). Among those, the affiliates of the *Bacteroidetes* were closely clustered and showed the highest expression in the families, Cohesin, Dockerin and CBMs. Among the *Haliscomenobacter*-related assembled genome bins (Bin09, Bin55, Bin 74, and Bin87), scaffolding and binding genes by Bin09 and Bin55 were highly expressed in U648 and L648 whereas the genes by Bin74 and Bin87 became more abundantly expressed in U798. *Chitinophaga*-related assembled genomes bins (Bin08 and Bin56) were highly expressed in U648 except for the families of Cohesin and Dockerin. Bin24 in the *Bacteroidetes* encoded and actively expressed genes in the glucan specific CBM family (CBM44). Expression of peptidoglycan specific binding modules (CBM50) in U798 was also overrepresented by *Geobacter* and *Dechloromonas*-related assembled genome bins, Bin78 and Bin79. The predicted glycoside hydrolytic GH families were mostly endoglucanase (GH74), GalNAc hydrolase (GH109), and peptidoglycan lyase (GH23). As the binding modules were expressed by the microbial population in the *Bacteroidetes*, the GH families were also covered by Bin08, Bin09, and Bin74 in U648 and L648. They were continuously covered by Bin74 and taken over by Bin87, Bin78 and Bin79 in U798. Thiopeptidoglycan lyses (PL9) was uniquely expressed by *Deltaproteobacteria*-related genome (Bin03) in L648. GlcNAc and MurNAc deacetylase (CE1, 4, and 10), enzymes released in the process of endo-utilization of peptidoglycan construction units, were also highly encoded and expressed by the *Bacteroidetes*-related assembled genome bins. Oligogalactouronate lyase (PL22) and peroxidase (AA2) in U798 was again transcribed by Bin74 and Bin78.

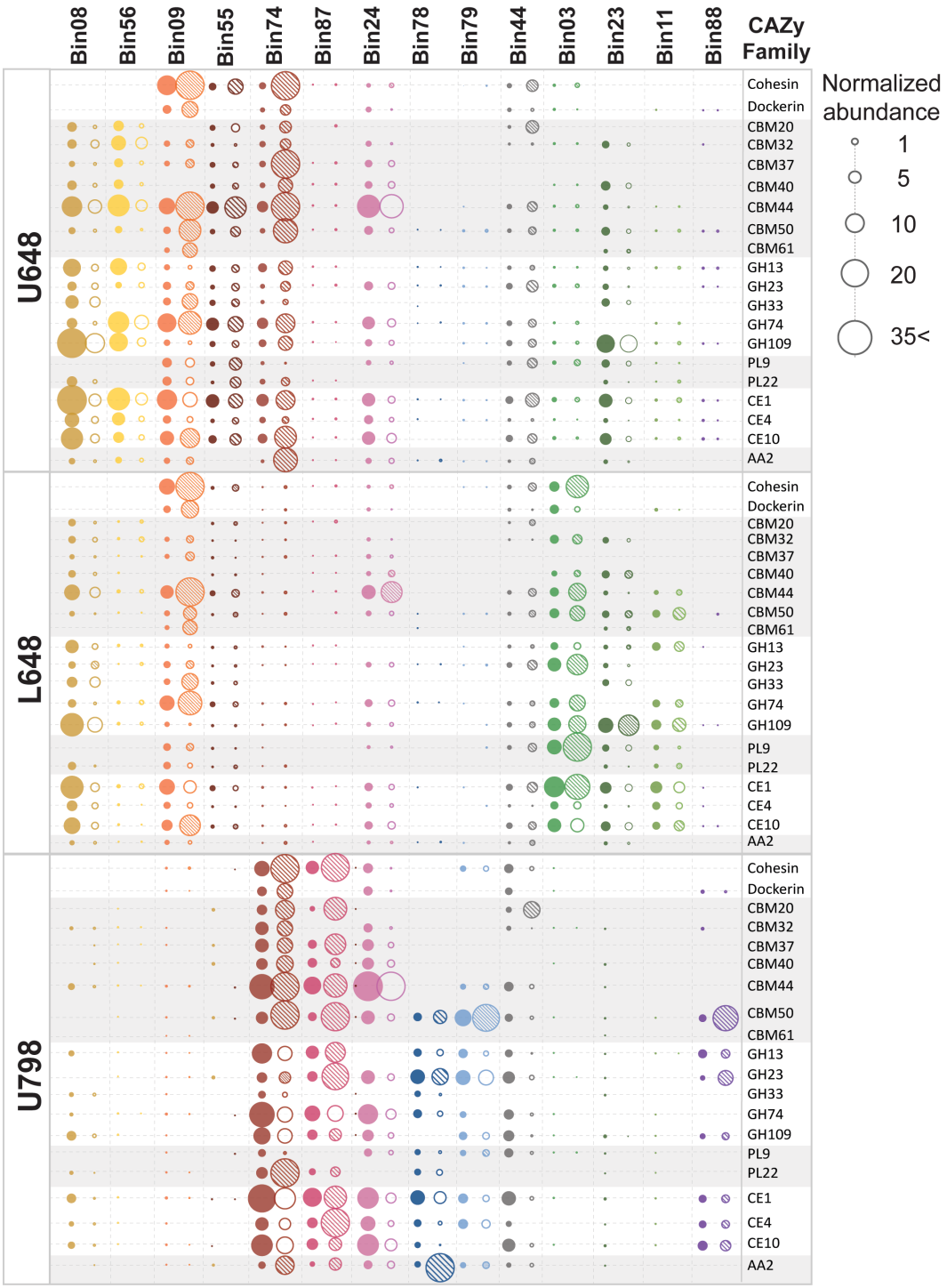



**Figure 4.4** Potential encoding and expression of carbohydrate-active enzymes (CAZy) by the dominant draft genomes. The genomic normalized abundance (closed circle) and the transcriptional normalized abundance (open circle) of the each CAZy family from the draft genomes were plotted, and the transcriptional activities of the CAZy families from the draft genomes in each dataset, which were significantly abundant at the 98% confidence level, were marked with a line pattern in the open circle.

### 4.4.4 Characteristics of predominant BAP of SMP from the AP and HP reactors

Based on the performance of the DHS reactor and SMP degradation during the operation described in the previous study,[29] the samples for the metagenomic and metatranscriptomic analyses were collected at the two time points; one was day 648 in Phase III, which was the last phase when the low and stable SMP loading in terms of SCOD concentration was provided to the DHS reactor, and the other was day 798 in Phase V when the four times higher SMP loading was given to the reactor than that in Phase III. The SMP produced from the AP and HP reactor were characterized as bimodal MW distribution with large compounds, 14-20 kDa, and small compounds, less than 4 kDa.[29] Classified by the unified theory of SMP, the large compounds were contemplated as BAP and the small compounds were considered to represent UAP.[1] In both phases when the samples for the meta-omic analyses were collected, very skewed SMP MW distributions to the large MW were detected, meaning that the majority of the SMP were likely associated with BAP, detritus released from cell lysis and decay. These skewed proportion of BAP between two types of SMP was generally observed in processes with a long SRT under which the AP and HP reactors were operated.[12, 21, 62] The remaining BAP may be related to a low biodegradability caused by their intrinsic complex heteropolymeric structures compared to UAP. The accumulation and preservation of BAP in the system might be because relatively limited microbial organisms can utilize BAP whereas diverse microbial assemblages are able to preferentially uptake UAP as a form of substrates.[25]

### 4.4.5 Degradation of carbohydrate components of SMP in the DHS reactor

One of the most abundant metabolic category in the global functionality analysis of the DHS microbial communities was Carbohydrates. In the sub-categories of Carbohydrates, outer membrane binding proteins for extracellular polysaccharides in the Cellulosome sub-category of the all three datasets were consistently activated together with transporting functions for mono- and di-saccharide uptakes, such as the glucose-specific PTS and the GlcNAc-specific ABC transporter. It led to the speculation that the microbial communities utilized extracellular polysaccharides composed of glucose and GlcNAc by confining the polysaccharides for glycoside hydrolases to cleave the bonds of targeted polymers. To

further investigate the genomic potential and transcriptional expression of the polysaccharide degrading gene families in individual organisms, the metagenomic and metatranscriptomic data were mapped on to the binned contigs that encoded the CAZy genes. As intrigued in the Carbohydrate metabolic gene activities, cohesin, dockerin, and binding module protein families were highly expressed by *Haliscomenobacter*-related assembled genome bins (Bin09, Bin55, Bin74, and Bin87) together with Bin24 in the all three datasets and the part of expression was taken over by *Geobacter* (Bin79) and *Dechloromonas*-related genomes (Bin78). Active gene expressions of these assembled genome bins also showed a similar pattern in the GH families such as endoglucanses, GalNAc hydrolases, and peptidoglycan lyases with expression of GlcNAc and MurNAc deacetylase in the CE families. When hydrolytic enzyme systems act on homogeneous and heterogeneous polysaccharide chains, including glucan, glycan, cellulose, hemicellulose, and pectin, extracellular hydrolases and lyases that are free or cell associated must be produced by microorganisms. Endo- and exo- hydrolases randomly cleave glycosidic bonds at an internal amorphous sites and an end of polysaccharide chains, respectively, into oligosaccharides. The freed oligosaccharides are further hydrolyzed into di- and mono-saccharides by glycosidases.[63] In this process, CBMs facilitate for the glycosidic hydrolases, which are anchored to the cell wall scaffold by the combined unit of cohesin and dockerin, to be close to the concentrated polysaccharides.[63-64] Each CBM binds to its specific target saccharides.[63] The most expressed four CBMs at the 98% confidence level, in this study, had a binding property to the peptidoglycan and its constituents such as GlcNAc. For instance, CBM44, which Bin09, Bin24, Bin55, Bin74, and Bin87 significantly expressed in, facilitates endo-(1,4)-beta-glucanase, which is able to cleave beta-(1,4)-glycosidic bond of a GlcNAc and MurNAc linkage. CBM 32, 37, and 50 bind to the GlcNAc residues in peptidoglycans to effect facilitated lysins, such as muramidase, N-acetylglucosaminidase, muropeptidase, and N-acetylmuramoyl-L-alanine amidase. Considering the recalcitrant property of the peptidoglycan and the outer membrane components among the detritus of microbial cell lysis,[65-7172-73] the highly encoded and expressed binding modules may indicate that the major populations utilized the fragmented polysaccharide cell wall structures, initiating the binds to them. In addition, the facilitated GH families by the CBMs, endoglucanses, GalNAc hydrolase, and peptidoglycan lyases, were represented by the major assembled genome bins.

GlcNAc and MurNAc deacetylases (CE4), one of the enzymes to convert GlcNAc to fructose-6-phosphate entering into gluconeogenesis, were significantly activated by *Haliscomenobacter*-related assembled genome bins (Bin87) in U798.

### 4.4.6 Degradation of protein derivatives of SMP in the DHS reactor

Amino acid and derivatives and Protein metabolism in the global functionality contained by the DHS microbial communities were the most abundantly encoded and expressed functional categories related to utilization of SMP. It was observed that the DHS microbial communities were significantly involved in the degradation of detrital cell wall components including peptidoglycans as a form of SMP by the genes showing high expressional activities at the sub-levels of these categories. First, in the Protein categories ATP dependent proteolysis and proteasome were evenly activated among the three data sets and the major twelve homologs at the order level. Endopeptidase Clp, which hydrolyzes oligopeptidases shorter than five amino acids in the absence of ATP, in the sub-categories, proteolysis and proteasome, were highly activated, suggesting that the overall expression of the genes were to cleave interpeptide bridges crosslinking the peptidoglycan strands and stemmed amino acids.[74] *Bacteroidales* relatives were observed to significantly express Dipeptidases. In the Amino acid derivatives category, metabolisms of amino acids composing the stemmed peptidoglycan-peptides, including alanine, glutamine, lysine, ornithine, and threonine showed outstanding expressions. The expression of aminotransferases converting alanine and glutamate to pyruvate and alpha-ketoglutarate, respectively, that are metabolites entering into the citric acid cycle, significantly activated in the upper and lower datasets in Phase III. Especially these expressions were predominated by *Solibacterales*, which also indicated high expression in proteasome for hydrolysis of oligopeptidases. *Sphingobacteriales* actively expressed glutamate dehydrogenases, deaminating glutamate to alpha-ketoglutarate in all three datasets followed by *Flavobacteriales* in U648, *Bacteroidetes* and *Clostridiales* in U798. Taken together, it was speculated that *Solibacterales* and *Sphingobacteriales* played important roles in fragmentation of peptide fragments released from cell membrane components to alanine and glutamate, as well as their catabolic processes.

### 4.4.7 Comparison of genes related to structural biomass detritus utilization in the major assembled genome bins

Given from the genome-wide phylogenetic analysis, it was revealed that the most dominant assembled genome bins in the three meta-data sets were closely clustered with *Haliscomenobacter hydrossis* and *Chitinophaga pinensis* in the *Bacteroidetes* phylum. These gram negative filamentous bacteria have been detected worldwide in activated sludge, especially known as specialized feeders using a narrow range of substrate groups such as glucose and N-acetylglucosamine under strict aerobic conditions, not various fatty acids.[75-78] The property of these bacteria explained their dominant existence in an activated sludge process, since the specialization on sugar degradation allowed them to convert cell wall structural components, such as lipopolysaccharides and peptidoglycan, liberated by decaying cells in the process.[76] To utilize the specialized substrates, these bacterial groups were known for presence of exo-enzyme activity, such as chitinase, glucuronidase, esterase, and phosphatase. Furthermore, they were identified to be equipped with genes (hex, nagZ, nagK, murQ, nagA, and nagB), which were necessary to catabolize N-substituted polysaccharide components found in lipopolysaccharides and peptidoglycan, such as GlcNAc, GalNAc and MurNAc. Here, gene inventory of the six assembled genome bins (Bin08, Bin09, Bin55, Bin56, Bin74, and Bin87) in *Bacteroidetes* was compared with *H.hydrossis* DSM1100 and *C.pinensis* UQM2034 as references to found out whether genes for the specialized substrate utilization were equipped in those bins (Figure 4.5 and Table C.9). Chitinase, targeting hydrolysis of N-acetylglucosamine polymers, lytic murein transglycosylase-related genes, and peptidoglycan binding lysine motifs were found the most of the assembled genome bins. Consecutive gene sets with TonB-dependent receptors, SusC, and SusD, which are transporters in outer membranes for the N-substituted oligosacchride derivatives from the exo-enzyme activities, were found in the all assembled genomes except for Bin09. Those assembled genome bins also contained nagX and nagP, which are specific transporters for N-acetylglucosamine in cytoplasmic membranes, together with nagA and nagB that are necessary deacetylase and deaminase to convert N-acetylglucosamine to Fructose-6-phposphate to get into the glycolysis pathway. Furthermore, among glycosidases in the assembled genome bins, beta-galactosidase and beta-glucanase, which hydrolyze major components of lipopolysaccharides, found in the

most of the genome bins. Having these findings, it is contemplated that the abundant existence of the filamentous bacteria-related to *Bacteroidetes* in the DHS consortia, compared to the conventional activated sludge process, was caused by the limited carbon and energy source from the influent, and rather they specifically adjusted to rely on utilization of biomass structural detritus released from the decaying cells. Furthermore, the three assembled genome bins (Bin03, Bin78, and Bin79) in Proteobacteria, the abundance of which increased in Phase V, were added to the comparison together with *Dechloromonas aromatica* RCB and *Geobacter metallireducens* GS-15 as references (Figure 4.5 and Table C.9). The most of the outer membrane gene sets were found in these assembled genome bins, whereas the specific transporters for N-substituted oligosaccharides and the gene sets for their catabolic processes were absent in the genomic bins proliferating in Phase V.

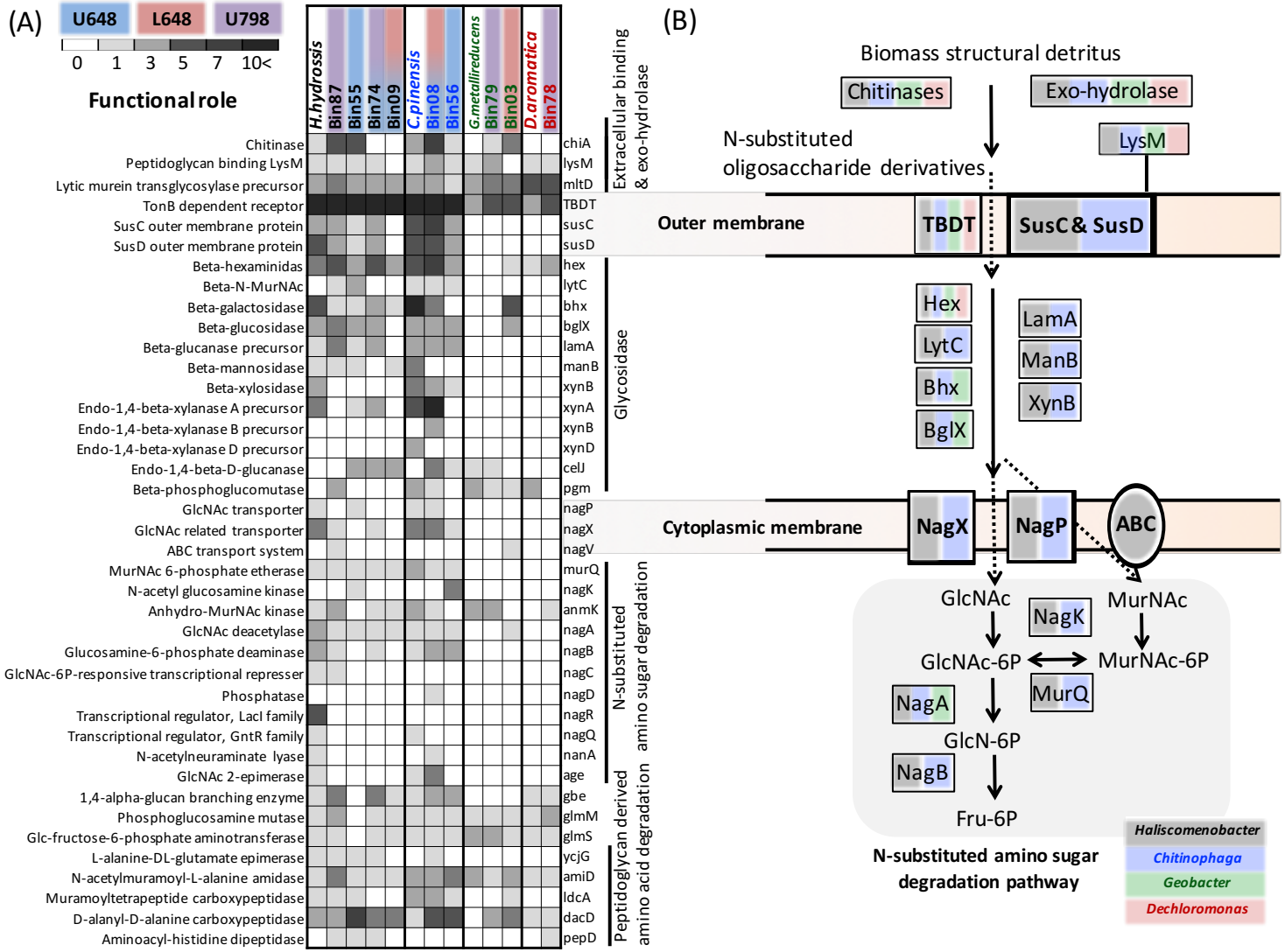



**Figure 4.5** Gene inventory analysis related to N-substituted biomass structural detritus utilization. (A) Related gene content of reference genomes and assembled genome bins. (B) Reconstruction of the N-substituted polysaccharide utilization pathway.

## 4.5 Conclusions

In this study, we investigated how the DHS microbial communities were functionally involved in the degradation of SMP generated from the anaerobic methanogenic reactors using metagenomic and metatranscriptomic approaches. As an increase of the SMP loading, a shift of the dominant microbial populations from *Sphingobacteriales*, *Flavobacteriaceae*, and *Cytophaga* to *Saprospiraceae*, *Dechloromonas*, and *Geobacter* was observed, whereas global functionality of the microbial communities for the SMP degradation was converged into the amino acids and derivative, carbohydrate, and protein-related metabolisms. On the other hand, a different functionality was assessed with high expression of the oligopeptide metabolism by relatively diverse community in the lower part of reactor compared to the upper, indicating a stratified SMP degradation in the reactor. The gene expression of carbohydrate-active enzymes in the dominant assembled genome bins and related gene set comparison with the reference genomes in *Bacteroidetes* showed that *Bacteroidetes*-related assembled genome bins mainly played important roles to specifically bind and utilize polysaccharide fragments derived from lipopolysaccharide and peptidoglycan-like BAP. The findings from the function-driven metagenomic and metatranscriptomic approaches suggested that the microbial communities degrading SMP in the DHS reactor were enriched to metabolize detrital components originated from microbial cell wall structural components. This is considered to be a unique observation, compared to conventionally used activated sludge microbial communities generally representing more central carbohydrate metabolisms by the dominant *Proteobacteria* and *Actinobacter*-related populations.

## 4.6 Acknowledgements

I deeply appreciated Dr. Christopher Fields, Dr. Jenny Drnevich, and Dr. Kathleen Keating at HPCBio for insightful discussion on metagenomic binning and differential expression analyses. I would like to thank Dr. Suengdae Oh for procedure guidance and sharing perl scripts for statistical analyses.

## 4.7 References


1. Laspidou, C. S.; Rittmann, B. E., A unified theory for extracellular polymeric substances, soluble microbial products, and active and inert biomass. *Water Research* **2002,** *36* (11), 2711-2720.

2. Barker, D. J.; Stuckey, D. C., A review of soluble microbial products (SMP) in wastewater treatment systems. *Water Research* **1999,** *33* (14), 3063-3082.

3. Barker, D. J.; Mannucchi, G. A.; Salvi, S. M. L.; Stuckey, D. C., Characterisation of soluble residual chemical oxygen demand (COD) in anaerobic wastewater treatment effluents. *Water Research* **1999,** *33* (11), 2499-2510.

4. Jarusutthirak, C.; Amy, G., Role of soluble microbial products (SMP) in membrane fouling and flux decline. *Environmental Science & Technology* **2006,** *40* (3), 969-974.

5. Liang, S.; Liu, C.; Song, L. F., Soluble microbial products in membrane bioreactor operation: Behaviors, characteristics, and fouling potential. *Water Research* **2007,** *41* (1), 95-101.

6. Magbanua, B. S.; Bowers, A. R., Characterization of soluble microbial products (SMP) derived from glucose and phenol in dual substrate activated sludge bioreactors. *Biotechnology and Bioengineering* **2006,** *93* (5), 862-870.

7. Tang, H. L.; Chen, Y. C.; Regan, J. M.; Xie, Y. F. F., Disinfection by-product formation potentials in wastewater effluents and their reductions in a wastewater treatment plant. *Journal of Environmental Monitoring* **2012,** *14* (6), 1515-1522.

8. Wei, Y. Y.; Liu, Y.; Zhang, Y.; Dai, R. H.; Liu, X. A.; Wu, J. J.; Zhang, Q. A., Influence of soluble microbial products (SMP) on wastewater disinfection byproducts: trihalomethanes and haloacetic acid species from the chlorination of SMP. *Environmental Science and Pollution Research* **2011,** *18* (1), 46-50.

9. Rosenberger, S.; Laabs, C.; Lesjean, B.; Gnirss, R.; Amy, G.; Jekel, M.; Schrotter, J. C., Impact of colloidal and soluble organic material on membrane performance in membrane bioreactors for municipal wastewater treatment. *Water Research* **2006,** *40* (4), 710-720.

10. Wang, X. M.; Waite, T. D., Role of Gelling Soluble and Colloidal Microbial Products in Membrane Fouling. *Environmental Science & Technology* **2009,** *43* (24), 9341-9347.

11. Fan, F.; Zhou, H., Interrelated effects of aeration and mixed liquor fractions on membrane fouling for submerged membrane bioreactor processes in wastewater treatment. *Environmental science & technology* **2007,** *41* (7), 2523-2528.

12. Jiang, T.; Myngheer, S.; De Pauw, D. J. W.; Spanjers, H.; Nopens, I.; Kennedy, M. D.; Amy, G.; Vanrolleghem, P. A., Modelling the production and degradation of soluble

microbial products (SMP) in membrane bioreactors (MBR). *Water Research* **2008,** *42* (20), 4955-4964.

13. Barker, D. J.; Salvi, S. M. L.; Langenhoff, A. A. M.; Stuckey, D. C., Soluble microbial products in ABR treating low-strength wastewater. *J. Environ. Eng.-ASCE* **2000,** *126* (3), 239-249.

14. Rittmann, B. E. a. M., P.L. , *Environmental Biotechnology: Principles and Applications*. McGraw-Hill: 2001.

15. Akram, A.; Stuckey, D. C., Flux and performance improvement in a submerged anaerobic membrane bioreactor (SAMBR) using powdered activated carbon (PAC). *Process Biochemistry* **2008,** *43* (1), 93-102.

16. Hu, A. Y.; Stuckey, D. C., Activated carbon addition to a submerged anaerobic membrane bioreactor: Effect on performance, transmembrane pressure, and flux. *J. Environ. Eng.-ASCE* **2007,** *133* (1), 73-80.

17. Wu, B.; An, Y.; Li, Y.; Wong, F. S., Effect of adsorption/coagulation on membrane fouling in microfiltration process post-treating anaerobic digestion effluent. *Desalination* **2009,** *242* (1-3), 183-192.

18. Aquino, S. F.; Stuckey, D. C., Soluble microbial products formation in anaerobic chemostats in the presense of toxic compounds. *Water Research* **2004,** *38* (2), 255-266.

19. Huang, G. T.; Jin, G.; Wu, J. H.; Liu, Y. D., Effects of glucose and phenol on soluble microbial products (SMP) in sequencing batch reactor systems. *International Biodeterioration & Biodegradation* **2008,** *62* (2), 104-108.

20. Kuo, W. C.; Parkin, G. F., Characterization of soluble microbial products from anaerobic treatment by molecular weight distribution and nickel-chelating properties. *Water Research* **1996,** *30* (4), 915-922.

21. Ni, B. J.; Rittmann, B. E.; Fang, F.; Xu, J. A.; Yu, H. Q., Long-term formation of microbial products in a sequencing batch reactor. *Water Research* **2010,** *44* (13), 3787-3796.

22. Cottrell, M. T.; Kirchman, D. L., Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacter cluster consuming low-and high-molecular-weight dissolved organic matter. *Applied and Environmental Microbiology* **2000,** *66* (4), 1692-1697.

23. Kindaichi, T.; Ito, T.; Okabe, S., Ecophysiological interaction between nitrifying bacteria and heterotrophic bacteria in autotrophic nitrifying biofilms as determined by microautoradiography-fluorescence in situ hybridization. *Applied and Environmental Microbiology* **2004,** *70* (3), 1641-1650.

24. O'Sullivan, L. A.; Weightman, A. J.; Fry, J. C., New degenerate Cytophaga-Flexibacter-Bacteroides-specific 16S ribosomal DNA-targeted oligonucleotide probes reveal

high bacterial diversity in River Taff epilithon. *Applied and environmental microbiology* **2002,** *68* (1), 201-210.

25. Okabe, S.; Kindaichi, T.; Ito, T., Fate of C-14-labeled microbial products derived from nitrifying bacteria in autotrophic nitrifying biofilms. *Applied and Environmental Microbiology* **2005,** *71* (7), 3987-3994.

26. Chon, K.; Lee, K.; Kim, I.-S.; Jang, A., Performance assessment of a submerged membrane bioreactor using a novel microbial consortium. *Bioresource technology* **2016,** *210*, 2-10.

27. Cai, L.; Yu, K.; Yang, Y.; Chen, B.-w.; Li, X.-d.; Zhang, T., Metagenomic exploration reveals high levels of microbial arsenic metabolism genes in activated sludge and coastal sediments. *Applied Microbiology and Biotechnology* **2013,** *97* (21), 9579-9588.

28. Miura, Y.; Okabe, S., Quantification of cell specific uptake activity of microbial products by uncultured Chloroflexi by microautoradiography combined with fluorescence in situ hybridization. *Environmental Science & Technology* **2008,** *42* (19), 7380-7386.

29. Kim, N. K.; Oh, S.; Liu, W. T., Enrichment and characterization of microbial consortia degrading soluble microbial products discharged from anaerobic methanogenic bioreactors. *Water Research* **2016,** *90*, 395-404.

30. Albertsen, M.; Hansen, L. B. S.; Saunders, A. M.; Nielsen, P. H.; Nielsen, K. L., A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. *Isme Journal* **2012,** *6* (6), 1094-1106.

31. Mason, O. U.; Hazen, T. C.; Borglin, S.; Chain, P. S. G.; Dubinsky, E. A.; Fortney, J. L.; Han, J.; Holman, H. Y. N.; Hultman, J.; Lamendella, R.; Mackelprang, R.; Malfatti, S.; Tom, L. M.; Tringe, S. G.; Woyke, T.; Zhou, J. H.; Rubin, E. M.; Jansson, J. K., Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *Isme Journal* **2012,** *6* (9), 1715-1727.

32. Ye, L.; Zhang, T.; Wang, T. T.; Fang, Z. W., Microbial Structures, Functions, and Metabolic Pathways in Wastewater Treatment Bioreactors Revealed Using High-Throughput Sequencing. *Environmental Science & Technology* **2012,** *46* (24), 13244-13252.

33. Yu, K.; Zhang, T., Metagenomic and Metatranscriptomic Analysis of Microbial Community Structure and Gene Expression of Activated Sludge. *Plos One* **2012,** *7* (5).

34. Cox, M.; Peterson, D.; Biggs, P., SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **2010,** *11* (1), 485.

35. Kopylova, E.; Noe, L.; Touzet, H., SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **2012,** *28* (24), 3211-3217.

36. 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* (D1), D590-D596.

37. Miller, C. S.; Baker, B. J.; Thomas, B. C.; Singer, S. W.; Banfield, J. F., EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biology* **2011,** *12* (5).

38. Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Pena, A. G.; Goodrich, J. K.; Gordon, J. I.; Huttley, G. A.; Kelley, S. T.; Knights, D.; Koenig, J. E.; Ley, R. E.; Lozupone, C. A.; McDonald, D.; Muegge, B. D.; Pirrung, M.; Reeder, J.; Sevinsky, J. R.; Tumbaugh, P. J.; Walters, W. A.; Widmann, J.; Yatsunenko, T.; Zaneveld, J.; Knight, R., QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **2010,** *7* (5), 335-336.

39. Haas, B. J.; Gevers, D.; Earl, A. M.; Feldgarden, M.; Ward, D. V.; Giannoukos, G.; Ciulla, D.; Tabbaa, D.; Highlander, S. K.; Sodergren, E., Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome research* **2011,** *21* (3), 494-504.

40. Caporaso, J. G.; Bittinger, K.; Bushman, F. D.; DeSantis, T. Z.; Andersen, G. L.; Knight, R., PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **2010,** *26* (2), 266-267.

41. Zerbino, D. R.; Birney, E., Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **2008,** *18* (5), 821-829.

42. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y.; Tang, J.; Wu, G.; Zhang, H.; Shi, Y.; Liu, Y.; Yu, C.; Wang, B.; Lu, Y.; Han, C.; Cheung, D.; Yiu, S.-M.; Peng, S.; Xiaoqian, Z.; Liu, G.; Liao, X.; Li, Y.; Yang, H.; Wang, J.; Lam, T.-W.; Wang, J., SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **2012,** *1* (1), 18.

43. Peng, Y.; Leung, H. C. M.; Yiu, S. M.; Chin, F. Y. L., IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **2012,** *28* (11), 1420-1428.

44. Luo, C. W.; Tsementzi, D.; Kyrpides, N. C.; Konstantinidis, K. T., Individual genome assembly from complex community short-read metagenomic datasets. *Isme Journal* **2012,** *6* (4), 898-901.

45. Meyer, F.; Paarmann, D.; D'Souza, M.; Olson, R.; Glass, E. M.; Kubal, M.; Paczian, T.; Rodriguez, A.; Stevens, R.; Wilke, A.; Wilkening, J.; Edwards, R. A., The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *Bmc Bioinformatics* **2008,** *9*.

46. Wilke, A.; Bischof, J.; Harrison, T.; Brettin, T.; D'Souza, M.; Gerlach, W.; Matthews, H.; Paczian, T.; Wilkening, J.; Glass, E. M., A RESTful API for accessing microbial community data for MG-RAST. *Plos Comput Biol* **2015,** *11* (1), e1004008.

47. Wu, Y.-W.; Tang, Y.-H.; Tringe, S. G.; Simmons, B. A.; Singer, S. W., MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2014,** *2*, 26-26.

48. Kang, D. D.; Froula, J.; Egan, R.; Wang, Z., *A robust statistical framework for reconstructing genomes from metagenomic data*. 2014.

49. Strous, M.; Kraft, B.; Bisdorf, R.; Tegetmeyer, H. E., The binning of metagenomic contigs for microbial physiology of mixed cultures. *Frontiers in microbiology* **2012,** *3* (December), 410.

50. Parks, D. H.; Imelfort, M.; Skennerton, C. T.; Hugenholtz, P.; Tyson, G. W., CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **2015**.

51. Wu, M.; Scott, A. J., Phylogenomic Analysis of Bacterial and Archaeal Sequences with AMPHORA2. *Bioinformatics* **2012**.

52. Hultman, J.; Waldrop, M. P.; Mackelprang, R.; David, M. M.; McFarland, J.; Blazewicz, S. J.; Harden, J.; Turetsky, M. R.; McGuire, A. D.; Shah, M. B., Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* **2015**.

53. Segata, N.; Börnigen, D.; Morgan, X. C.; Huttenhower, C., PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature communications* **2013,** *4*.

54. Overbeek, R.; Olson, R.; Pusch, G. D.; Olsen, G. J.; Davis, J. J.; Disz, T.; Edwards, R. A.; Gerdes, S.; Parrello, B.; Shukla, M.; Vonstein, V.; Wattam, A. R.; Xia, F.; Stevens, R., The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research* **2014,** *42* (Database issue), D206-D214.

55. Wattam, A. R.; Abraham, D.; Dalay, O.; Disz, T. L.; Driscoll, T.; Gabbard, J. L.; Gillespie, J. J.; Gough, R.; Hix, D.; Kenyon, R., PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic acids research* **2013**, gkt1099.

56. Rho, M. N.; Tang, H. X.; Ye, Y. Z., FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Research* **2010,** *38* (20).

57. Yin, Y. B.; Mao, X. Z.; Yang, J. C.; Chen, X.; Mao, F. L.; Xu, Y., dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **2012,** *40* (W1), W445-W451.

58. Lombard, V.; Ramulu, H. G.; Drula, E.; Coutinho, P. M.; Henrissat, B., The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **2014,** *42* (D1), D490-D495.

59. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Gloeckner, F. O., The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **2013,** *41* (D1), D590-D596.

60. Shi, Y.; Tyson, G. W.; Eppley, J. M.; DeLong, E. F., Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J* **2011,** *5* (6), 999-1013.

61. Huson, D. H.; Auch, A. F.; Qi, J.; Schuster, S. C., MEGAN analysis of metagenomic data. *Genome Research* **2007,** *17* (3), 377-386.

62. Xie, W.-M.; Ni, B.-J.; Sheng, G.-P.; Seviour, T.; Yu, H.-Q., Quantification and kinetic characterization of soluble microbial products from municipal wastewater treatment plants. *Water Research* **2016,** *88*, 703-710.

63. Lynd, L. R.; Weimer, P. J.; Van Zyl, W. H.; Pretorius, I. S., Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and molecular biology reviews* **2002,** *66* (3), 506-577.

64. Gilkes, N. R.; Henrissat, B.; Kilburn, D. G.; Miller, R. C.; Warren, R. A., Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families. *Microbiological Reviews* **1991,** *55* (2), 303-315.

65. Hansell, D. A.; Carlson, C. A., *Biogeochemistry of marine dissolved organic matter*. Academic Press: 2014.

66. Kirchman, D. L., *Microbial ecology of the oceans*. John Wiley & Sons: 2010; Vol. 36.

67. McCarthy, M. D.; Hedges, J. I.; Benner, R., Major bacterial contribution to marine dissolved organic nitrogen. *Science* **1998,** *281* (5374), 231-234.

68. Tanoue, E.; Nishiyama, S.; Kamo, M.; Tsugita, A., Bacterial membranes: possible source of a major dissolved protein in seawater. *Geochimica et Cosmochimica Acta* **1995,** *59* (12), 2643-2648.

69. Aquino, S. F.; Hu, A. Y.; Akram, A.; Stuckey, D. C., Characterization of dissolved compounds in submerged anaerobic membrane bioreactors (SAMBRs). *Journal of Chemical Technology and Biotechnology* **2006,** *81* (12), 1894-1904.

70. Le-Clech, P.; Chen, V.; Fane, T., Fouling in membrane bioreactors used in wastewater treatment. *J Membr Sci* **2006,** *284*, 17 - 53.

71. Dignac, M.-F.; Ginestet, P.; Rybacki, D.; Bruchet, A.; Urbain, V.; Scribe, P., Fate of wastewater organic pollution during activated sludge treatment: nature of residual organic matter. *Water Research* **2000,** *34* (17), 4185-4194.

72. Nagata, T.; Meon, B.; Kirchman, D., Microbial degradation of peptidoglycan in seawater. *Limnology and Oceanography* **2003,** *48* (2), 745-754.

73. JÖrgensen, N. O.; Stepanaukas, R.; Pedersen, A.-G. U.; Hansen, M.; Nybroe, O., Occurrence and degradation of peptidoglycan in aquatic environments. *FEMS microbiology ecology* **2003,** *46* (3), 269-280.

74. Maurizi, M. R.; Clark, W. P.; Katayama, Y.; Rudikoff, S.; Pumphrey, J.; Bowers, B.; Gottesman, S., Sequence and structure of Clp P, the proteolytic component of the ATP-dependent Clp protease of Escherichia coli. *J Biol Chem* **1990,** *265* (21), 12536-12545.

75. Eikelboom, D. H., *Identification and control of filamentous micro-organisms in industrial wastewater treatment plants*. IWA Publ.: 2006.

76. Kragelund, C.; Levantesi, C.; Borger, A.; Thelen, K.; Eikelboom, D.; Tandoi, V.; Kong, Y.; Krooneman, J.; Larsen, P.; Thomsen, T. R.; Nielsen, P. H., Identity, abundance and ecophysiology of filamentous bacteria belonging to the Bacteroidetes present in activated sludge plants. *Microbiol-Sgm* **2008,** *154*, 886-894.

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

78. Gutierrez-Zamora, M.-L.; Manefield, M., An appraisal of methods for linking environmental processes to specific microbial taxa. *Reviews in Environmental Science and Bio/Technology* **2010,** *9* (2), 153-185.

# CHAPTER 5: CONCLUSION

## 5.1 Conclusion

The AP and HP methanogenic reactors were successfully operated to treat high-strength synthetic soft drink wastewater, providing stable and high SCOD removal efficiency (>95%) over 800 days. Based on 16S rRNA gene pyrosequencing analyses, the predominant microbial populations in the AP and HP reactors were identified. As regards the bacterial classification, *Bacteroidetes*, *Chloroflexi*, *Firmicutes*, and KSB3 were the most dominant populations, which may primarily degrade organic constituents, such as glucose, fructose, and PEG. *Syntroph*-related populations, such as *Syntrophomonas*, *Syntrophobacter*, and *Smithella*, may support the degradation of VFAs, which are derived from the organic compounds. As for the archaeal classification, *Methanosaeta*, *Methanosarcia*, and *Methanobacterium* were detected as major methanogenic populations, oxidizing $H_2$ and acetate in the reactors. While the ecological role involved in the treatment of soft drink wastewater was not defined clearly, *Geobacter*, *Spirochaetes*, and GN04 were also detected as prevalent microbial groups in the anaeorbic reactors. The RDA analysis indicated that *Bacteroidetes*, *Chloroflexi*, KSB3, and GN04 were strongly influenced by changes in the OLR. This finding suggested that specific microorganisms in the microbial community, which are responsible for the sugar/PEG degradation, might be adapted to changes in the operational conditions.

The effluent produced from the AP and HP reactors were further treated in the following DHS reactor to improve the effluent quality by reducing the SMP in it. During the long-term and stable operation, the microbial consortia in the DHS reactor were selectively enriched to utilize SMP. The SMP contained in the effluent from the AP and HP reactors exhibited a bimodal MW distribution with 14-20 kDa and <4 kDa. About 70% of SMP in terms of SCOD removed by the enriched microbial consortia in the DHS reactor. Using 16S rRNA gene pyrosequencing analyses, the microbial community structure was characterized, and the spatial and temporal variability was correlated with operational factors by performing network and redundancy analyses. The results revealed that *Flavobacteriales*, *Saprospiraceae*, *Cytophaga*, and *Chloroflexi* were the predominant bacterial populations and significantly involved in SMP degradation. In particular, the *Saprospiraceae*-related

population was strongly correlated to the increasing SMP loading condition, indicating positive co-occurrences with neighboring bacterial populations. The abundance of these microbial groups was significantly affected by HRT, ORL, and SCOD removal. Moreover, the microbial diversity was influenced by reactor depth, implying adaptation of the microbial communities for an increased SMP loading and stratified degradation in the DHS reactor.

Besides an identification of the microbial communities, degrading SMP in the DHS reactor, in order to understand the functional mechanisms that are activated for the SMP degradation by the microbial community, it was important to address the role of individual populations and their interactions. Employing metagenomic and metatranscriptomic approaches, functional profiles, as well as the phylogenetic profiles of the DHS microbial communities, were assessed. The microbial composition was shifted, as with the increasing SMP loading; the dominant populations changed from *Sphingobacteriales*, *Flavobacteriaceae*, and *Cytophaga* to *Saprospiraceae*, *Dechloromonas*, and *Geobacter*. Nevertheless, the disparate microbial communities indicated a functional convergence in the annotation analyses of gene encodings and expressions based on a SEED subsystem. Composition and functionality of the microbial community in the lower part of the DHS reactor differed from those in the upper part, suggesting that stratified SMP degradation occurred. Results of the active gene expression in the global functionality, and the CAZy families, demonstrated that the microbial community significantly represented genes related to the metabolism of oligopeptides and polysaccharide constituents of peptidoglycan. Observations from the function-driven metagenomic and metatranscriptomic approaches reveal how microbial communities in the DHS reactor were to utilize detrital cell structural components released from peptidoglycan. These components may compose the majority of the SMP produced from the AP and HP reactors.

## 5.2 Contribution

This research demonstrated a promising alternative process, offering a combined process of anaerobic packed-bed reactors and a DHS reactor, to treat high-strength industrial wastewater. The process maximizes the advantages of an anaerobic reactor by retaining a high concentration of biomass in the system. Also, it successfully minimizes side effects

caused by the high concentration of biomass, such as increasing residual COD derived from SMP, by employing a DHS reactor as post-treatment. Biological SMP degradation, using a DHS reactor, herein, may resolve long-term application of SMP reduction, which remained limited by conventionally used chemical and physical methods. Additionally, this work studied an origin of the residual organic compounds in effluent from the anaerobic process and characterized their properties, which had been still unclear in previous reports that described the combined system of UASB and DHS reactors. The findings proved that the majority of SCOD in the anaerobic effluent originated from anaerobic biomass metabolisms, rather than yet-untreated raw wastewater. Upon verifying the feasibility of a long-term reduction of SMP, via selectively enriched microbial consortia in the DHS reactor, this study also scrutinized the phylogenetic characteristic and metabolic functionality of the DHS microbial community involved in the SMP degradation, using the NGS technology. The findings, which resulted from an observation of the overrepresented genes by the DHS microbial community, also led to the speculation that SMP might be derived from the detrital materials of the cell structures of the anaerobic biomass, such as peptidoglycan. These findings provide a possible application for the biological degradation of SMP using a DHS reactor, as well as broaden knowledge of SMP produced from mixed culture biotechnology.

### 5.3 Future prospects

Future research, in the context of the findings in this study, may take two directions: (1) the practical applications of biological SMP degradation; and (2) a fundamental understanding of commensal interactions among the microorganisms involved in the SMP degradation. SMP, ubiquitously present in bioprocesses, are often found to negatively impact the processes. Their compositions and properties vary, depending on system configurations, operational parameters, and substrates, among others. As regards a broad application of the biological process for SMP reduction, the utility of the SMP-reducing process needs to be investigated for improving the efficiency of conventional wastewater treatment systems and water reclamation. This study identified the microbial community involved in the SMP degradation and investigated their metabolic roles related to the de-polymerization of SMP. Possible commensal interactions of the dominant microbial

populations were interpreted using statistic-networking analysis. In addition to these findings, genomic aspects of the commensalisms in the microbial community for the SMP degradation need to further studied further.

## APPENDIX A: SUPPLEMENTAL MATERIALS IN CHAPTER 2

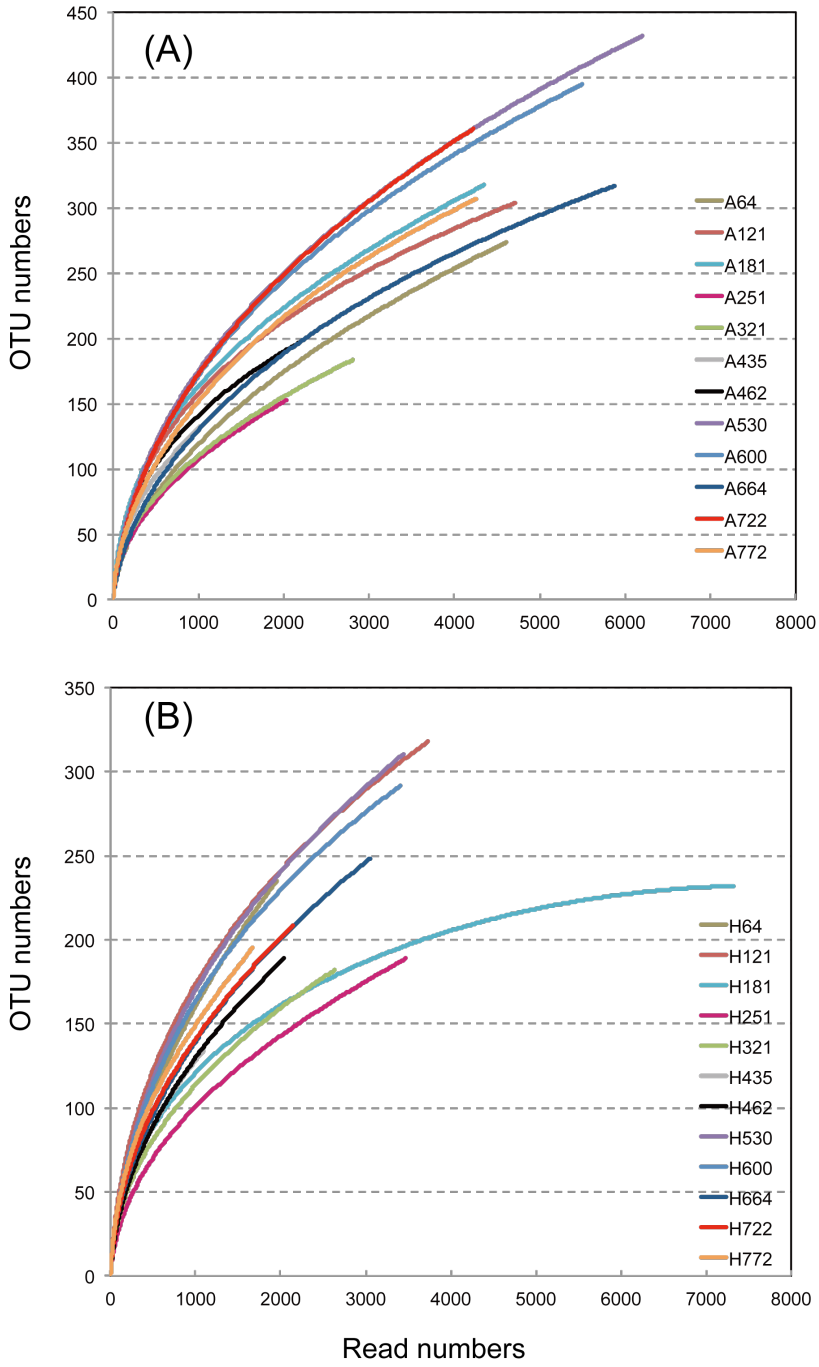



**Figure A.1** Rarefaction curves of 16S rRNA gene sequences of (A) anaerobic packed-bed (AP) and (B) hybrid packed-bed (HP) reactors.

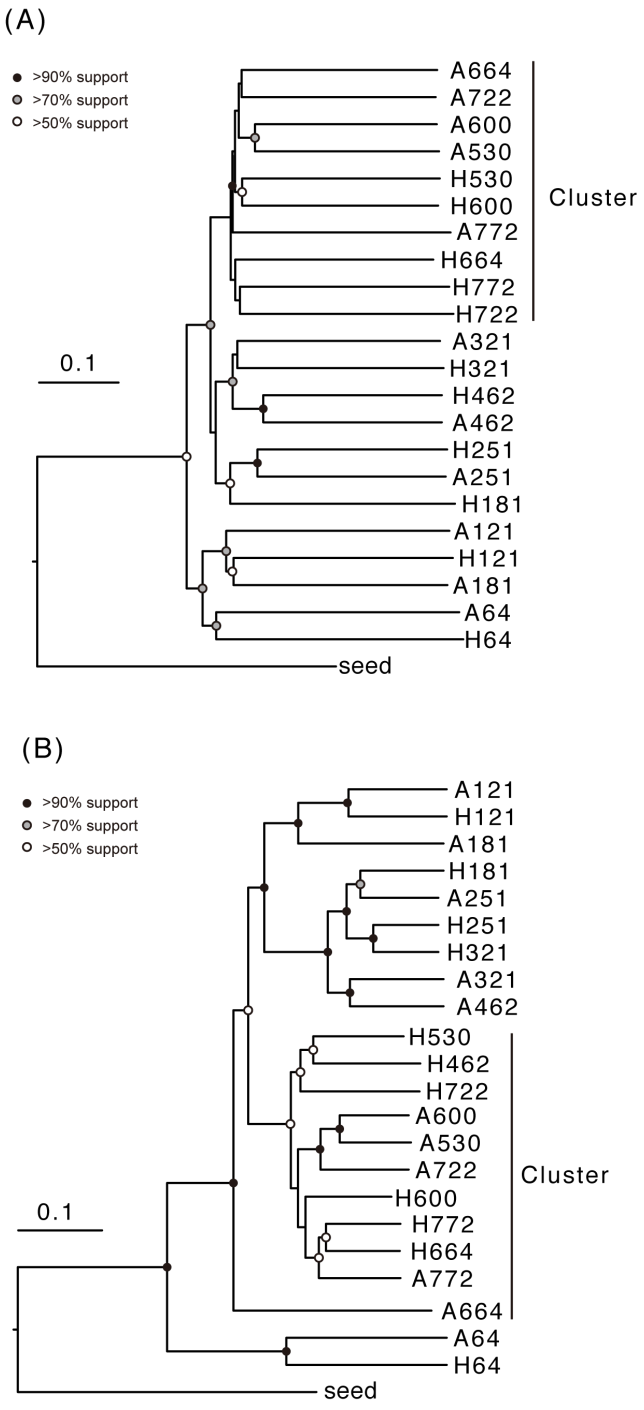



**Figure A.2** Jackknife clustering of 16S rRNA gene pyrotag libraries from anaerobic packed-bed (AP) and hybrid packed-bed (HP) reactors based on (A) unweighted and (B) weighted Uni- Frac normalized to 1,400 reads per sample. "Cluster" indicates the grouped samples showed in Fig. 3 (unweighted) and S3 Fig. (weighted).

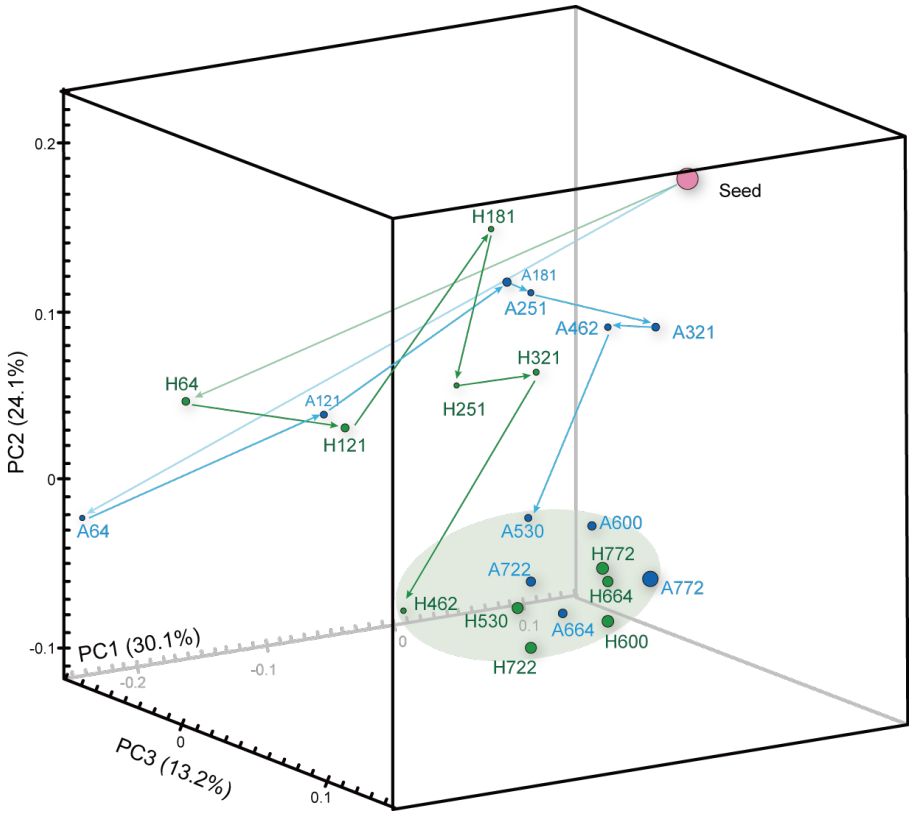



**Figure A.3** PCoA based on the abundances of 16S rRNA gene OTUs (weighted UniFrac). For this analysis, observed 16S rRNA gene OTUs were normalized to 1,400 reads per sample. A and H indicate the samples taken from the anaerobic packed-bed (AP) and hybrid packed-bed (HP) reactors. The numbers following A and H indicate days of the operation for biomass sampling.

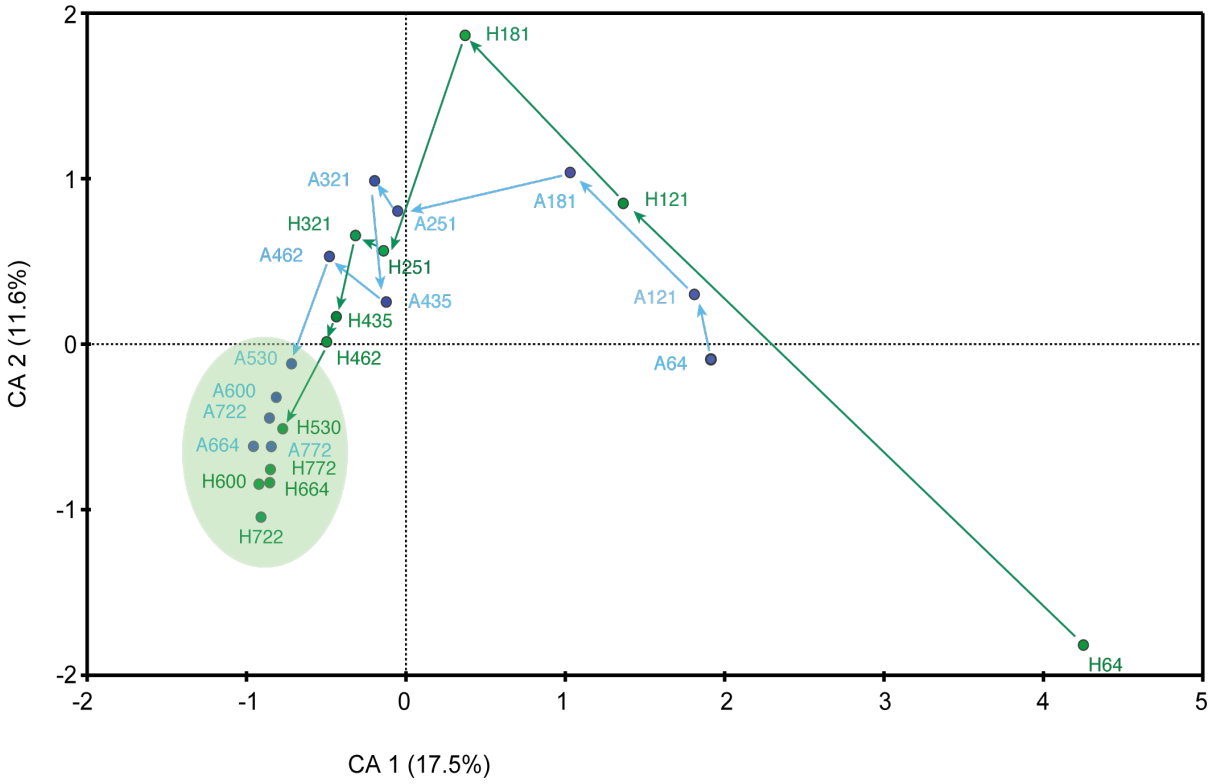



**Figure A.4** Correspondence analysis (CA) based on the abundances of 16S rRNA gene OTUs. A an H indicate the samples taken from the anaerobic packed-bed (AP) and hybrid packed-bed (HP) reactors. The numbers following A and H indicate days of the operation for biomass sampling.

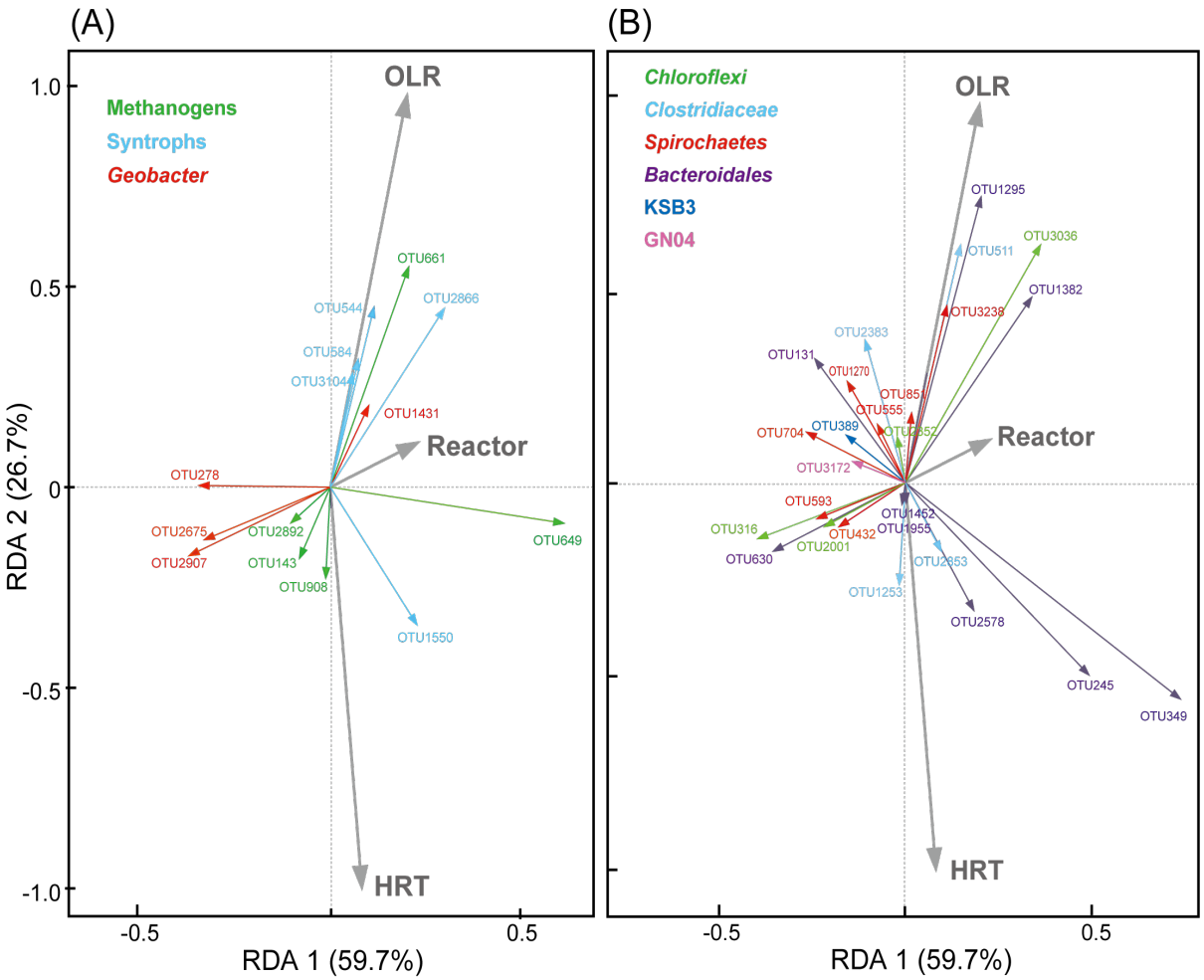



**Figure A.5** Redundancy analysis (RDA) based on the abundances of 16S rRNA gene OTUs of (A) known methanogens, syntrophs and Geobacter populations and (B) the phyla Bacteroi- detes, Chloroflexi, Firmicutes, and Spirochaetes and candidate phyla KSB3 and GN04 popu- lations.

**Table A.1** Pyrosequencing results of 16S rRNA genes amplicon reads from anaerobic packed-bed (AP) and hybrid packed-bed (HP) reactors.

| | seed | Operation days at sampling on AP reactor | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 64 | 121 | 181 | 251 | 321 | 435 | 462 | 530 | 600 | 664 | 722 | 772 |
| Total 16S pyrotag reads | 14,090 | 4,622 | 4,717 | 4,363 | 2,035 | 2,823 | 1,031 | 2,213 | 6,215 | 5,510 | 5,881 | 4,229 | 4,262 |
| Total OTU number (>97% identity) | 1,008 | 274 | 304 | 318 | 153 | 184 | 133 | 198 | 432 | 395 | 317 | 361 | 307 |
| Good's coverage | 96.8 | 96.8 | 97.4 | 96.7 | 96.2 | 96.8 | 93.8 | 96.3 | 96.8 | 96.7 | 97.6 | 95.8 | 96.8 |
| Chao1 | 1,680 | 552 | 525 | 612 | 341 | 338 | 259 | 302 | 716 | 688 | 533 | 682 | 522 |

| | Operation days at sampling on HP reactor | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 64 | 121 | 181 | 251 | 321 | 435 | 462 | 530 | 600 | 664 | 722 | 772 |
| Total 16S pyrotag reads | 1,955 | 3,737 | 7,336 | 3,478 | 2,637 | 1,105 | 2,045 | 3,458 | 3,415 | 3,069 | 2,147 | 1,684 |
| Total OTU number (>97% identity) | 235 | 318 | 232 | 189 | 182 | 135 | 189 | 311 | 292 | 249 | 209 | 196 |
| Good's coverage | 93.3 | 96.4 | 99.9 | 97.2 | 96.7 | 93.9 | 95.3 | 95.8 | 96.3 | 96.0 | 95.0 | 93.7 |
| Chao1 | 485 | 506 | 232 | 401 | 332 | 220 | 355 | 622 | 487 | 429 | 392 | 505 |

AP, anaerobic packed-bed reactor; HP, hybrid packed-bed reactor; OTU,operational taxonomic unit.

**Table A.2** Microbial community composition of anaerobic packed-bed (AP) and hybrid packed-bed (HP) reactors.

| | Seed | Sampling date on AP reactor (population %) | | | | | | | | | | | | Sampling date on HP reactor (population %) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | | 64 | 121 | 181 | 251 | 321 | 435 | 462 | 530 | 600 | 664 | 722 | 772 | 64 | 121 | 181 | 251 | 321 | 435 | 462 | 530 | 600 | 664 | 722 | 772 |
| *Bacteria* | | | | | | | | | | | | | | | | | | | | | | | | | |
| *Deltaproteobacteria* | 4.4 | 6.9 | 24.0 | 22.9 | 43.7 | 35.8 | 18.3 | 41.3 | 33.6 | 28.6 | 16.4 | 23.8 | 13.3 | 8.7 | 15.4 | 43.6 | 44.7 | 43.0 | 24.5 | 24.8 | 24.4 | 23.7 | 28.7 | 14.5 | 22.2 |
| *Bacteroidetes* | 18.2 | 43.1 | 20.9 | 14.6 | 10.6 | 8.5 | 22.4 | 8.2 | 6.4 | 5.1 | 2.9 | 11.9 | 13.7 | 40.8 | 20.9 | 13.9 | 8.3 | 7.2 | 7.8 | 4.4 | 8.5 | 6.4 | 8.0 | 12.9 | 11.4 |
| *Chloroflexi* | 8.6 | 5.0 | 10.7 | 12.0 | 10.7 | 19.4 | 9.6 | 10.8 | 9.2 | 7.0 | 3.9 | 6.2 | 13.4 | 3.7 | 10.2 | 7.3 | 11.5 | 12.5 | 18.5 | 11.2 | 10.3 | 12.4 | 12.4 | 12.8 | 10.9 |
| *Firmicutes* | 35.6 | 9.0 | 12.5 | 15.4 | 9.7 | 11.2 | 9.2 | 5.9 | 4.9 | 3.5 | 2.4 | 3.5 | 12.8 | 16.6 | 13.0 | 8.5 | 6.6 | 7.4 | 3.9 | 4.6 | 5.1 | 3.6 | 5.2 | 3.9 | 8.7 |
| *Spirochaetes* | 5.8 | 0.6 | 2.8 | 7.1 | 7.4 | 13.6 | 20.6 | 14.9 | 11.8 | 10.9 | 9.0 | 7.8 | 15.8 | 1.6 | 3.5 | 11.2 | 5.8 | 11.6 | 20.6 | 17.8 | 11.7 | 16.3 | 16.5 | 14.0 | 20.8 |
| *Nitrospirae* | 0.0 | 0.0 | 0.0 | 0.5 | 1.8 | 2.7 | 1.9 | 2.7 | 3.9 | 2.5 | 1.0 | 2.2 | 1.8 | 0.0 | 0.0 | 1.7 | 1.0 | 1.1 | 2.0 | 0.9 | 1.6 | 1.4 | 2.1 | 1.4 | 2.0 |
| *Planctomycetes* | 0.6 | 0.2 | 0.3 | 0.2 | 0.4 | 0.7 | 0.6 | 2.9 | 1.2 | 1.5 | 1.2 | 1.4 | 0.4 | 0.4 | 0.3 | 0.2 | 0.3 | 0.3 | 0.9 | 1.1 | 0.5 | 0.4 | 0.3 | 0.4 | 0.2 |
| *Chlorobi* | 0.1 | 0.3 | 0.9 | 2.8 | 1.6 | 2.4 | 1.6 | 3.2 | 4.5 | 3.4 | 2.6 | 3.4 | 1.9 | 0.1 | 0.6 | 1.0 | 0.5 | 1.4 | 0.5 | 0.6 | 3.3 | 1.1 | 0.6 | 1.0 | 0.4 |
| *Acidobacteria* | 0.1 | 0.1 | 0.0 | 0.1 | 0.1 | 0.6 | 0.9 | 0.6 | 0.4 | 0.6 | 0.2 | 0.9 | 0.2 | 0.1 | 0.3 | 0.2 | 0.1 | 0.4 | 1.5 | 0.5 | 0.4 | 0.3 | 0.2 | 0.3 | 0.5 |
| *Alphaproteobacteria* | 2.9 | 0.2 | 0.4 | 0.5 | 0.0 | 0.3 | 0.7 | 0.7 | 0.2 | 0.3 | 0.1 | 0.1 | 1.1 | 0.3 | 1.0 | 0.1 | 0.2 | 0.4 | 0.5 | 0.1 | 0.1 | 0.2 | 0.0 | 0.3 | 0.0 |
| *Caldiserica* | 0.2 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.2 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 |
| *Verrucomicrobia* | 1.1 | 0.0 | 0.1 | 1.0 | 0.0 | 0.1 | 0.0 | 0.1 | 0.2 | 0.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.0 | 0.2 | 0.1 | 0.0 | 0.0 | 0.1 |
| *Cyanobacteria* | 0.1 | 0.0 | 0.9 | 3.5 | 0.1 | 0.1 | 1.7 | 1.6 | 0.1 | 0.1 | 0.1 | 0.2 | 0.0 | 0.0 | 3.3 | 1.0 | 0.2 | 0.8 | 0.3 | 0.1 | 0.1 | 0.1 | 0.0 | 0.4 | 0.1 |
| *Armatimonadetes* | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.0 | 0.2 | 0.1 | 0.0 | 0.1 | 0.0 | 0.3 | 0.1 | 0.0 | 0.2 | 0.1 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| *Gemmatimonadetes* | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.4 | 0.1 | 0.2 | 0.1 | 0.2 | 0.1 | 0.0 | 0.1 | 0.1 | 0.0 | 0.0 | 0.1 | 0.4 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| *Betaproteobacteria* | 3.6 | 0.7 | 0.1 | 0.3 | 0.0 | 0.1 | 0.1 | 0.3 | 0.0 | 0.1 | 0.0 | 0.5 | 0.4 | 0.5 | 0.1 | 1.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.0 | 0.0 | 0.0 | 3.6 | 0.1 |
| *Actinobacteria* | 3.2 | 0.1 | 0.2 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.2 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 |
| *Synergistetes* | 2.4 | 0.6 | 1.2 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 3.0 | 0.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| *Gammaproteobacteria* | 0.9 | 0.1 | 0.2 | 0.0 | 0.0 | 0.4 | 0.1 | 0.4 | 0.2 | 0.1 | 0.1 | 0.0 | 0.2 | 0.2 | 0.1 | 0.0 | 0.1 | 0.3 | 0.6 | 0.1 | 0.1 | 0.0 | 0.0 | 0.4 | 0.2 |
| *Thermotogae* | 0.6 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| *Tenericutes* | 0.4 | 0.2 | 0.7 | 0.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.5 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| *Fusobacteria* | 0.2 | 0.0 | 0.0 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| *Fibrobacteres* | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| *Lentisphaerae* | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| *Epsilonproteobacteria* | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 |
| *Chlamydiae* | 0.0 | 0.0 | 0.0 | 0.1 | 0.2 | 0.0 | 0.2 | 0.1 | 0.2 | 0.1 | 0.2 | 0.0 | 0.4 | 0.0 | 0.3 | 0.1 | 0.0 | 0.3 | 0.0 | 0.1 | 0.2 | 0.1 | 0.1 | 0.1 | 0.4 |
| *Elusimicrobia* | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| *Thermi* | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

**Table A.2 (cont.)**

| | Seed | Sampling date on AP reactor (population %) | | | | | | | | | | | | Sampling date on HP reactor (population %) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | | 64 | 121 | 181 | 251 | 321 | 435 | 462 | 530 | 600 | 664 | 722 | 772 | 64 | 121 | 181 | 251 | 321 | 435 | 462 | 530 | 600 | 664 | 722 | 772 |
| Candidate phyla | | | | | | | | | | | | | | | | | | | | | | | | | |
| GN04 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.9 | 0.5 | 4.7 | 5.0 | 2.7 | 6.1 | 0.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 1.4 | 1.1 | 5.3 | 12.0 | 3.8 | 1.7 | 0.7 |
| KSB3 | 0.0 | 0.0 | 0.0 | 0.3 | 4.0 | 0.0 | 0.0 | 0.0 | 0.0 | 12.2 | 38.6 | 12.5 | 5.4 | 0.0 | 0.0 | 1.8 | 2.7 | 0.0 | 0.0 | 0.0 | 0.4 | 1.0 | 5.0 | 4.3 | 3.1 |
| FCPU426 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.4 | 2.0 | 0.6 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.8 | 4.1 | 2.8 | 1.0 | 1.1 | 0.6 | 0.2 |
| OP3 | 0.3 | 0.0 | 0.5 | 3.8 | 0.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.3 | 0.7 | 0.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 |
| WPS-2 | 0.0 | 0.0 | 0.1 | 0.1 | 0.2 | 1.2 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.1 | 0.3 | 0.1 | 0.2 | 0.0 | 0.0 | 0.3 | 0.1 | 0.0 | 0.0 | 0.0 | 0.1 | 0.2 | 0.2 |
| NKB19 | 0.4 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.3 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.3 | 0.1 | 0.2 | 0.0 | 0.0 | 0.1 | 0.5 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 |
| WSA2 | 0.3 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.4 | 0.1 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 |
| BRC1 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 | 0.5 | 0.1 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| OP11 | 0.0 | 0.0 | 0.7 | 1.4 | 0.2 | 0.1 | 0.0 | 0.2 | 0.2 | 0.0 | 0.1 | 0.0 | 0.1 | 0.1 | 0.3 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| WWE1 | 2.9 | 0.6 | 1.4 | 0.9 | 0.0 | 0.3 | 0.0 | 0.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.1 | 2.5 | 0.1 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Hyd24-12 | 0.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| OP9 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| OD1 | 0.1 | 0.0 | 0.1 | 0.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.1 | 0.5 | 0.0 | 0.4 | 0.7 | 0.6 | 0.0 | 0.3 | 0.2 | 0.1 | 0.0 | 0.3 | 0.1 | 0.0 | 0.0 | 0.7 |
| WS1 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.0 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| OP8 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 |
| TM6 | 0.0 | 0.2 | 0.0 | 0.1 | 0.3 | 0.1 | 0.9 | 0.2 | 0.3 | 0.2 | 0.1 | 0.3 | 0.1 | 0.0 | 0.2 | 0.1 | 0.1 | 0.4 | 0.2 | 0.5 | 0.3 | 0.1 | 0.1 | 0.1 | 0.2 |
| SAR406 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| GN02 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.1 | 0.0 | 0.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 |
| WS4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| WS3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.2 | 0.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| TM7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| SR1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| TA18 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| NC10 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 |
| LD1 | 0.0 | 0.0 | 0.0 | 0.6 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.3 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| FBP | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

**Table A.2 (cont.)**

| | Seed | Sampling date on AP reactor (population %) | | | | | | | | | | | | Sampling date on HP reactor (population %) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | | 64 | 121 | 181 | 251 | 321 | 435 | 462 | 530 | 600 | 664 | 722 | 772 | 64 | 121 | 181 | 251 | 321 | 435 | 462 | 530 | 600 | 664 | 722 | 772 |
| *Archaea* | | | | | | | | | | | | | | | | | | | | | | | | | |
| *Methanosaeta* | 3.1 | 27.1 | 10.3 | 5.2 | 3.3 | 1.0 | 6.2 | 3.4 | 14.9 | 14.7 | 15.5 | 16.5 | 16.1 | 2.7 | 15.4 | 2.6 | 8.7 | 8.4 | 13.3 | 24.9 | 22.4 | 17.7 | 14.7 | 26.1 | 16.2 |
| *Methanosarcina* | 0.0 | 0.0 | 5.5 | 1.2 | 0.0 | 0.1 | 1.3 | 0.1 | 0.1 | 0.1 | 0.1 | 0.0 | 0.0 | 12.4 | 0.0 | 0.0 | 0.4 | 0.8 | 0.1 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 |
| *Methanospirillum* | 0.4 | 0.0 | 1.9 | 0.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.4 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| *Methanolinea* | 0.6 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.1 | 0.0 | 0.2 | 0.1 | 0.1 | 0.1 | 0.0 | 0.2 | 0.2 | 0.1 | 0.2 | 0.0 | 0.1 | 0.0 |
| *Methanoculleus* | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Other *Methanomicrobiales* | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| *Methanobacterium* | 0.0 | 1.3 | 1.4 | 1.3 | 4.2 | 0.0 | 1.1 | 0.0 | 0.2 | 0.1 | 0.0 | 0.2 | 0.1 | 1.0 | 3.2 | 2.7 | 6.5 | 2.2 | 0.0 | 0.0 | 0.5 | 0.4 | 0.1 | 0.2 | 0.1 |
| *Methanobrevibacter* | 0.0 | 0.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| *Other Methanobacteriaceae* | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| *Methanomassiliicoccaceae* | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 0.1 | 0.5 | 0.1 | 0.2 | 0.1 | 0.0 | 0.1 | 0.0 | 0.4 | 0.1 | 0.0 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.1 |
| *Crenarchaeota* | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.0 | 0.1 | 0.1 | 0.0 | 0.0 | 0.2 | 0.3 | 0.2 | 0.3 | 0.1 | 0.1 | 0.0 |
| *Parvarchaeota* | 0.2 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Unassigned | 0.6 | 2.1 | 1.7 | 2.0 | 0.3 | 0.5 | 0.3 | 0.2 | 0.8 | 0.7 | 0.8 | 0.9 | 0.7 | 0.3 | 3.6 | 0.7 | 0.3 | 0.2 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.2 | 0.2 |

## APPENDIX B: SUPPLEMENTAL MATERIALS IN CHAPTER 3

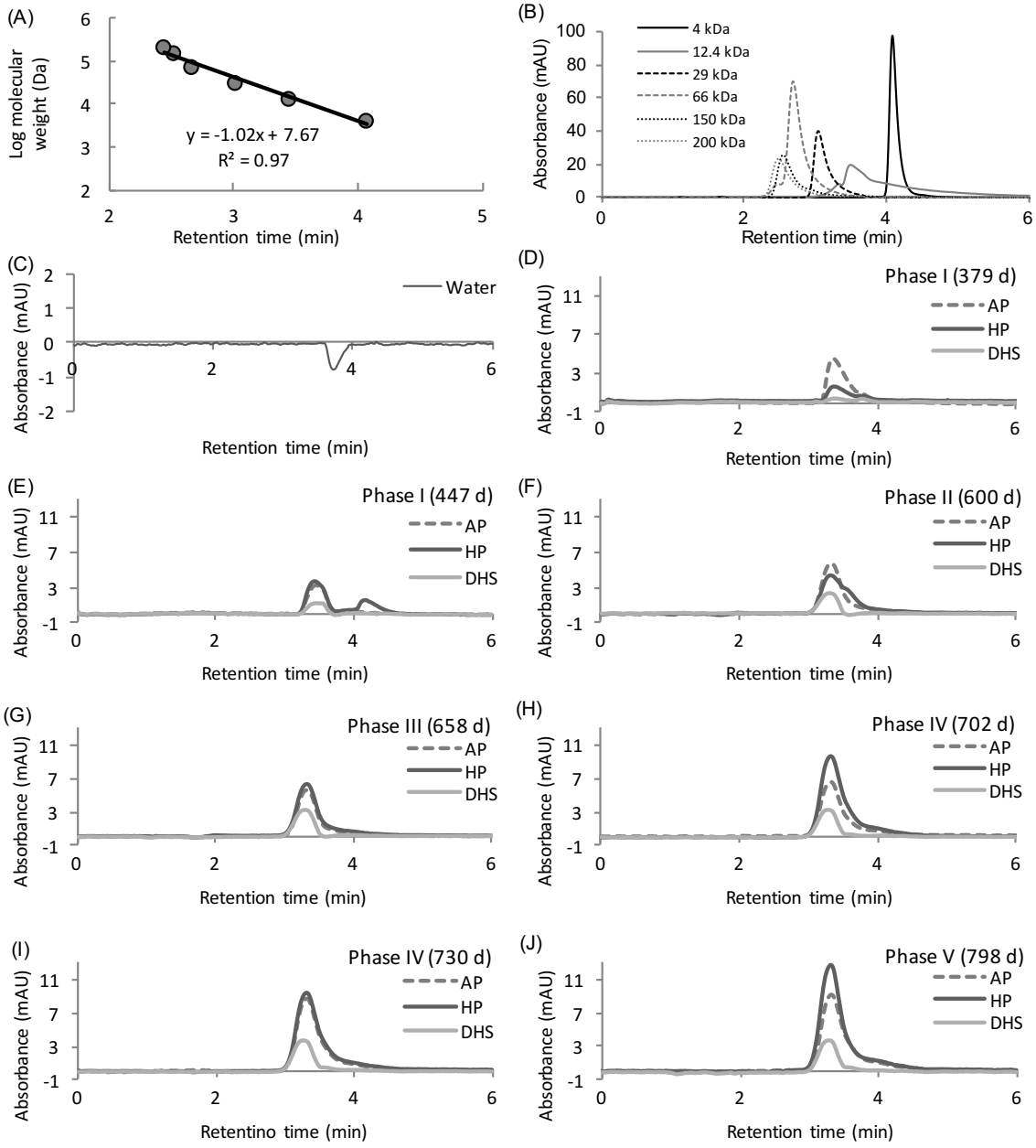



**Figure B.1** High performance liquid chromatography-size exclusion chromatography (HPLC-SEC) analyses of the effluent SMP from the AP and HP reactors and the effluent from the DHS reactor in Phases I-V: (A) the standard curve, (B) the chromatograms of the standards, (C) the chromatograms of a water sample, and (D-J) the chromatograms of the samples. The number in parentheses indicates the days when the samples were collected.

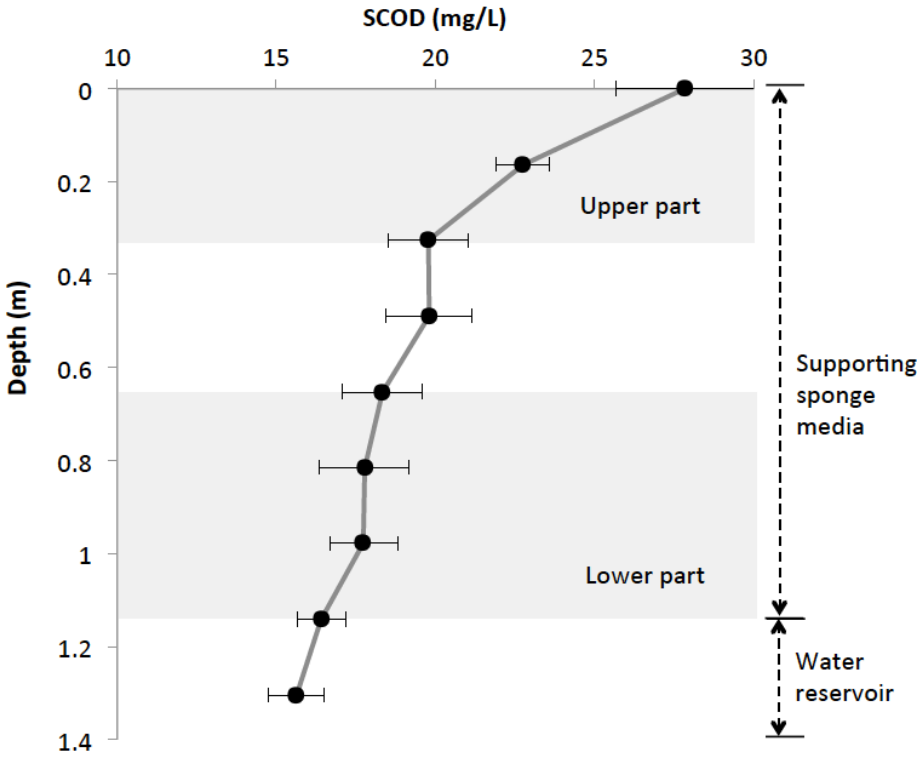



**Figure B.2** SMP degradation profiles in terms of SCOD removal along with the DHS reactor depth (n=4). Eight samples in a depth between 0.0-1.2 m were collected from the supporting sponge media, and a sample in a depth between 1.2-1.4 m was collected from the water reservoir.

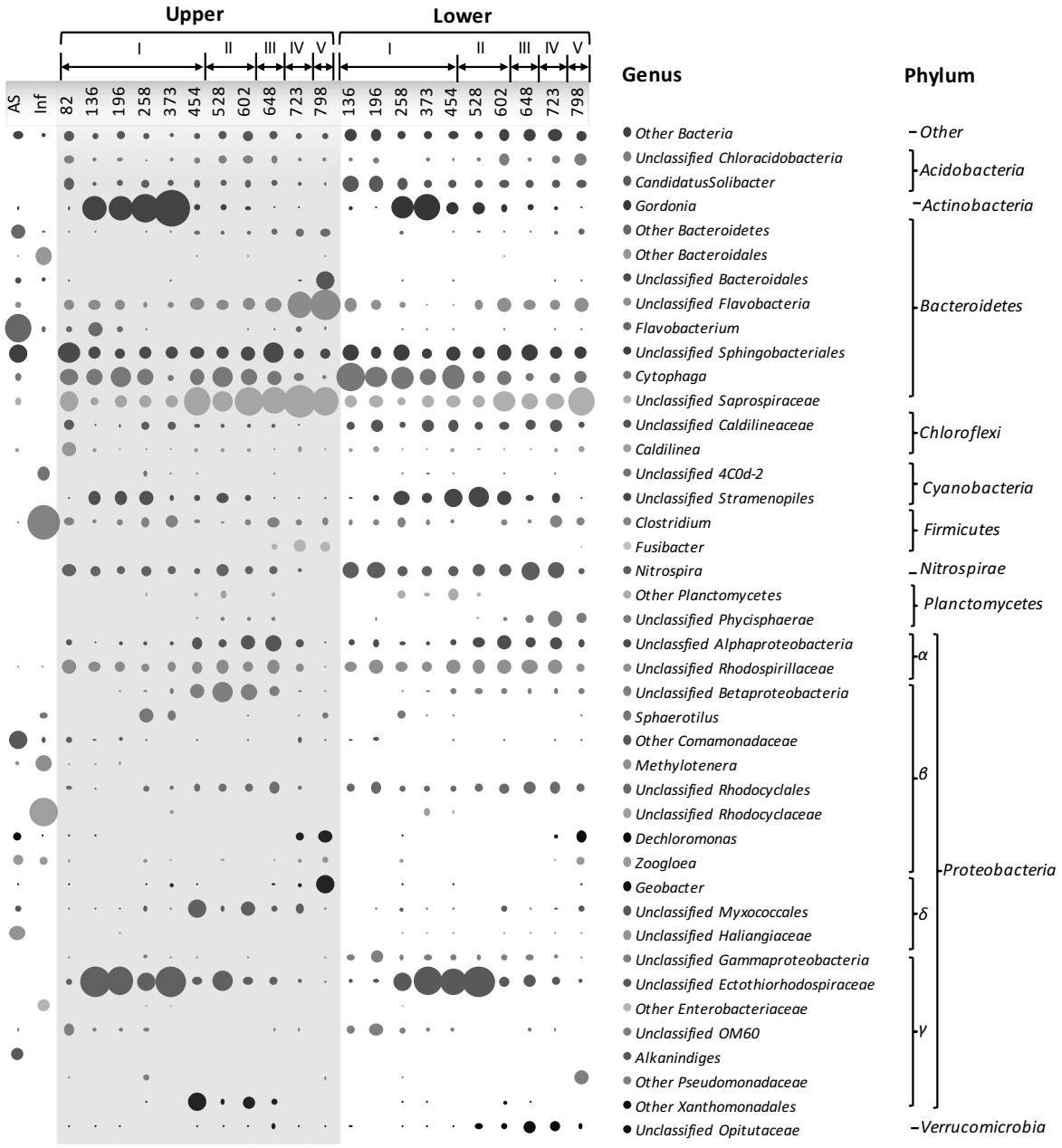



**Figure B.3** Pyrosequencing profiles showing the relative abundance of the microbial communities in the DHS reactor at the genus level (abundance >3% in any sample). The roman numerals indicate the five phases. 'AS' stands for the inoculated activated sludge, 'Inf' stands for influent, and the numbers indicate the days when the biomasses were collected.

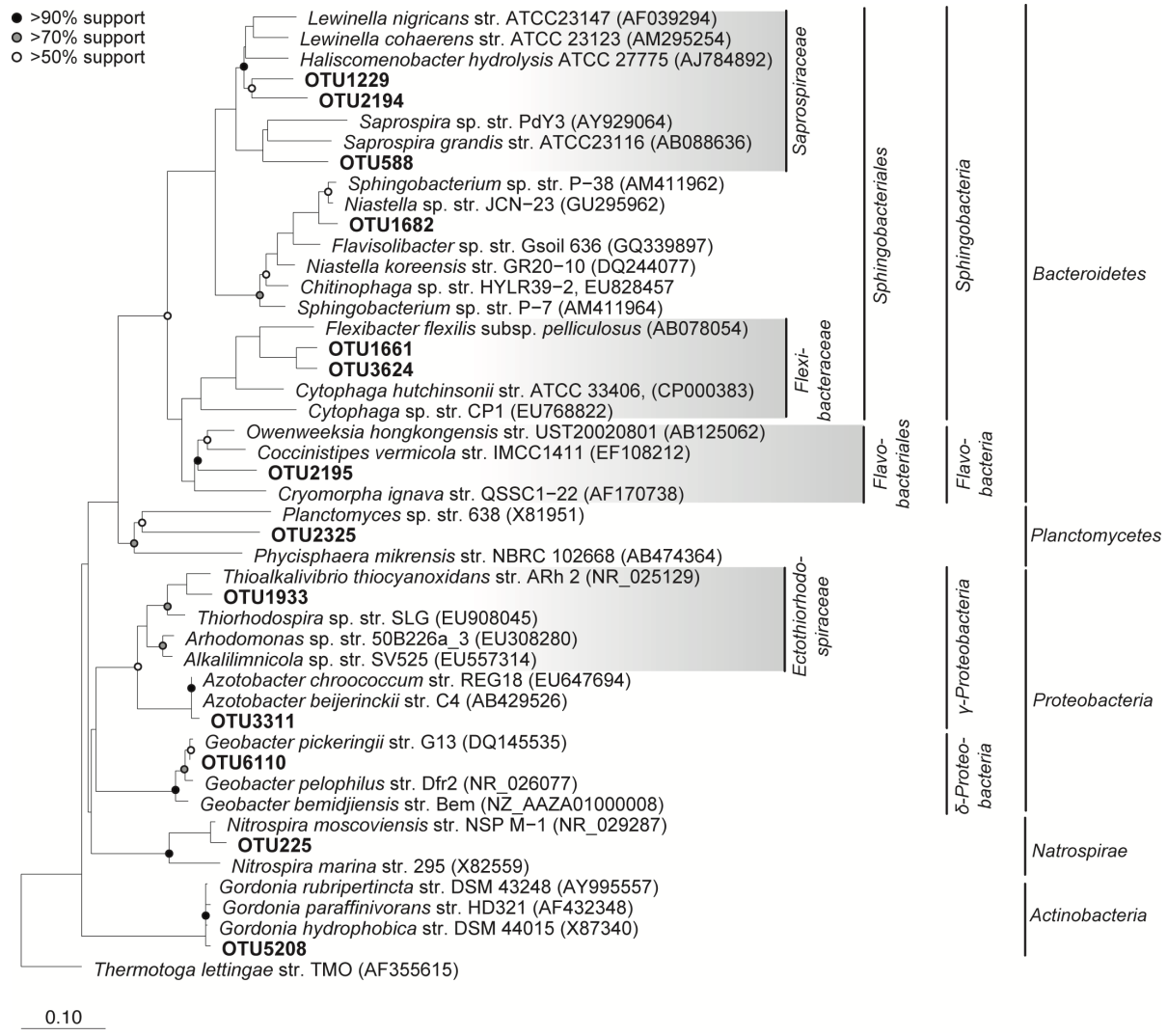



**Figure B.4** Phylogenetic tree based on the abundant OTUs (>4%) that had direct correlations with the operational factors in the network. Boldface indicates the sequences obtained in this study. The tree was constructed using the neighbor-joining algorithm with Jukes-Cantor correction and out-grouped with *Thermotoga lettingae* TMO strain (AF355615). The bar indicates 10% base substitution. Bootstrap values were calculated based on 1000 replications. >90%, >70%, and >50% of bootstrap values are indicated by black, gray, and white circles, respectively.

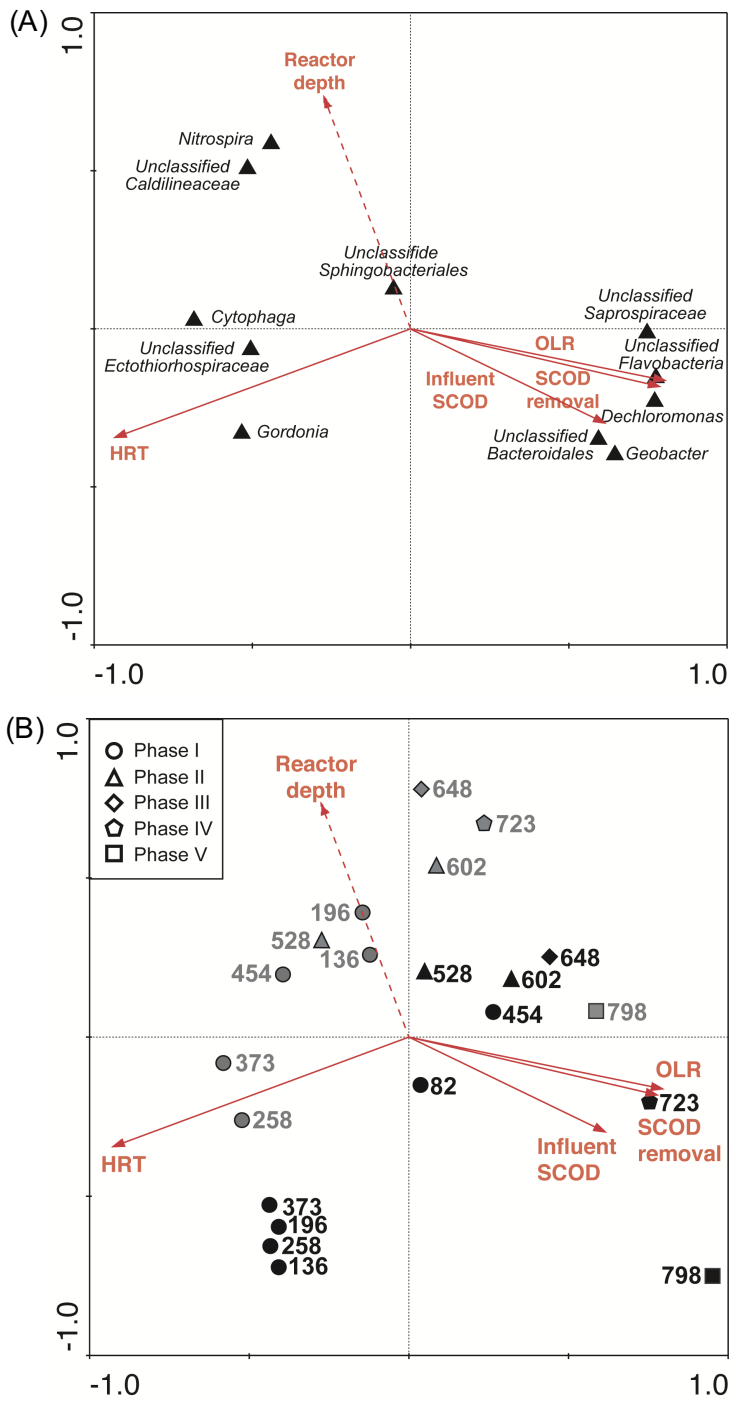



**Figure B.5** RDA ordination of the microbial community by genus (A) and the samples from the DHS reactor (B). Correspondence of the 414 genera and the 23 samples with the operational variables, HRT, OLR, SCOD removal, influent SCOD and reactor depth, were analyzed. For (A), eleven dominant genera are selectively shown in the ordination by black triangles. HRT, OLR, SCOD removal, and influent SCOD are indicated by red arrows due to their statistical significance ($P < 0.05$), and reactor depth is indicated by a red dotted arrow ($P > 0.05$).

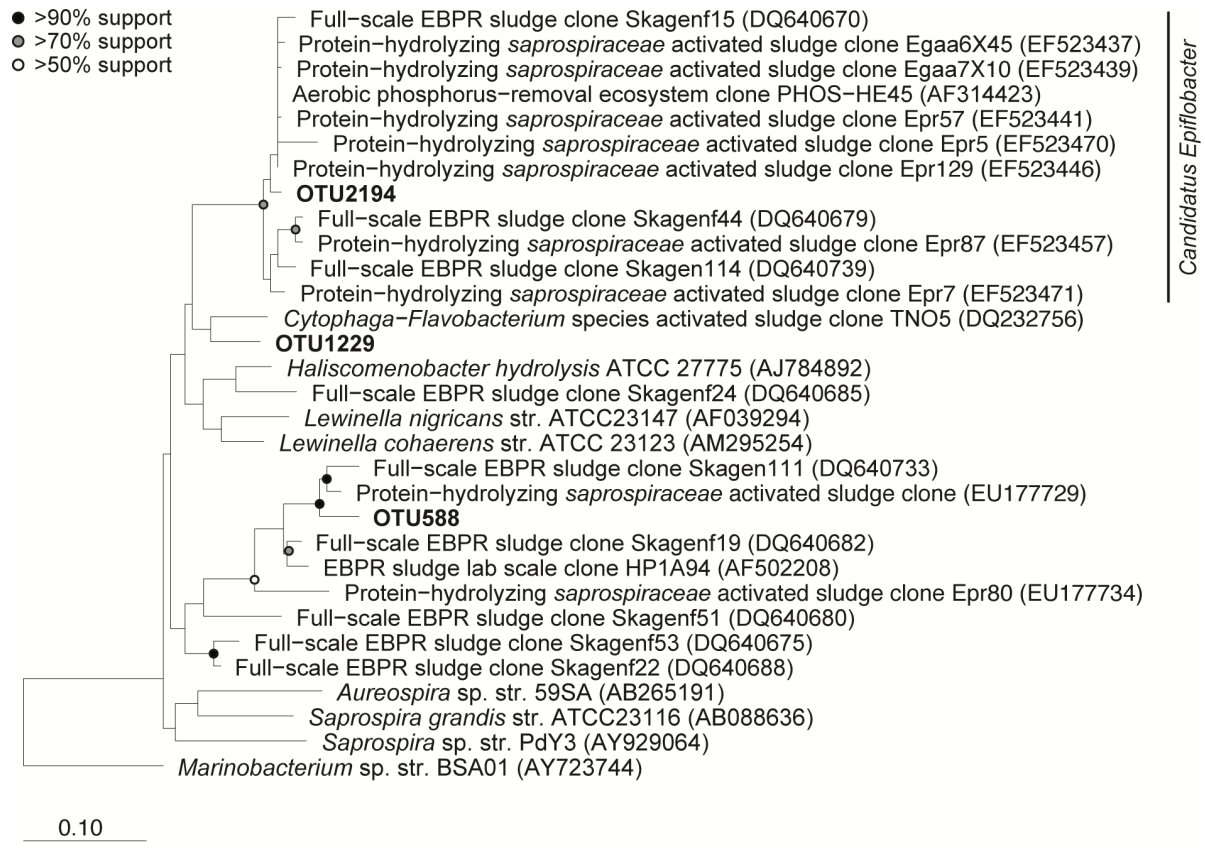



**Figure B.6** Phylogenetic tree based on *Saprospiraceae*-related OTUs with *Candidatus Epiflobacter* spp. in the family *Saprospiraceae*. The tree was constructed using the neighbor-joining algorithm with Jukes-Cantor correction and out-grouped with *Marinobacterium* sp. strain. BSA01 (AY723744). The bar indicates 10% base substitution. Bootstrap values were calculated based on 1000 replications. >90%, >70%, and >50% of bootstrap values are indicated by black, gray, and white circles, respectively.

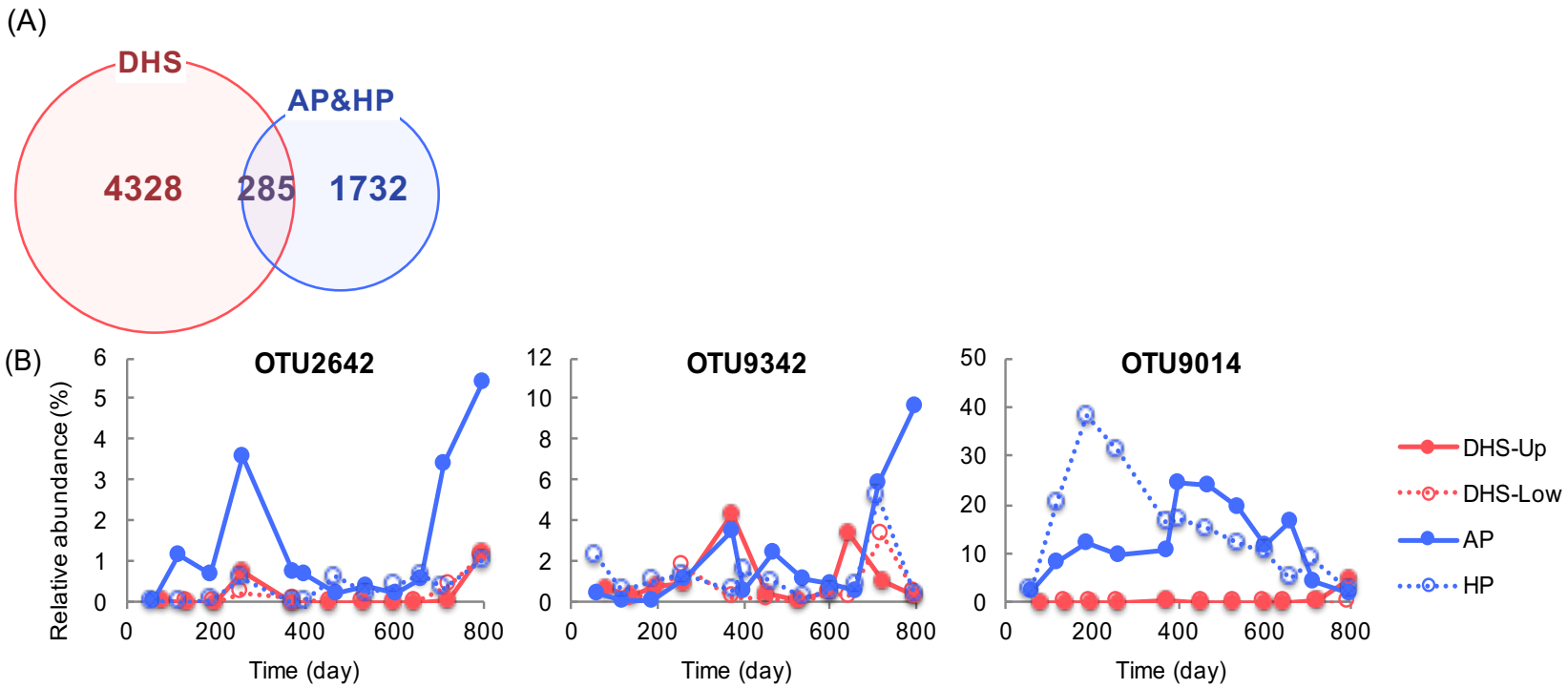



**Figure B.7** Unique and shared OTUs between the microbial communities in the DHS reactor and the AP and HP reactors: (A) venn diagram of the shared and uniqeu OTUs and (B) abundance profiles of the three dominant OTUs that were commonly detected in the DHS reactor and the AP and HP reactors.

**Table B.1** Synthetic wastewater composition and characteristics.

| Components | Concentration ($mg\ L^{-1}$) |
|---|---|
| Synthetic wastewater composition | |
| High fructose corn syrup (CornSweet® High Fructose 55, ADM)<br>Carbohydrate composition (dry weight basis)<br>Fructose 55%<br>Glucose 41%<br>Polysaccharides 4% | 1500.0 |
| Polyethylene glycol 200 | 1100.0 |
| Acetone | 30.0 |
| Ethanol | 30.0 |
| Potassium phosphate ($K_2HPO_4$) | 16.0 |
| Ferrous sulfate ($FeSO_4 \cdot 7H_2O$) | 19.0 |
| Sodium bicarbonate ($NaHCO_3$) | 366.0 |
| Sodium fluoride (NaF) | 2.0 |
| Sodium hypochlorite (NaOCl) | 2.5 |
| Ammonium bicarbonate ($NH_4HCO_3$) [a] | 28.0 |
| Characteristics | |
| Total COD | 3000.0 |
| Soluble COD | 2939.0 |

a Ammonium biocarbonate was added as a component of synthetic wastewater at day 84.

**Table B.2** Summary of the relative abundance of OTUs (at least >4%) applied to the network analysis.

Unit (%)

| Phase | I | | | | | | | | | | | II | | | | III | | IV | | V | | Average relative abundance |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Day | 82 | 136 | | 196 | | 258 | | 373 | | 454 | | 528 | | 602 | | 648 | | 723 | | 798 | | |
| Location | Up | Up | Low | Up | Low | Up | Low | Up | Low | Up | Low | Up | Low | Up | Low | Up | Low | Up | Low | Up | Low | |
| OTU1229 | 1.9 | 0.3 | 0.2 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 2.1 | 0.1 | 9.8 | 1.5 | 0.8 |
| OTU1255 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.2 | 0.3 | 0.1 | 0.0 | 8.2 | 0.0 | 0.1 | 0.0 | 4.4 | 0.8 | 0.8 | 0.0 | 2.1 | 0.2 | 0.0 | 0.3 | 0.8 |
| OTU1650 | 2.9 | 0.2 | 3.4 | 2.0 | 3.7 | 1.2 | 0.5 | 2.7 | 3.6 | 9.1 | 4.1 | 2.3 | 3.9 | 9.2 | 9.0 | 7.8 | 5.8 | 0.4 | 3.7 | 0.0 | 0.0 | 3.6 |
| OTU1661 | 0.2 | 0.3 | 12.3 | 6.9 | 6.7 | 5.2 | 8.0 | 0.4 | 4.6 | 1.3 | 7.6 | 3.3 | 1.8 | 1.5 | 1.1 | 1.9 | 0.8 | 0.1 | 1.3 | 0.0 | 1.1 | 3.2 |
| OTU1682 | 7.6 | 1.7 | 2.9 | 1.0 | 1.1 | 1.7 | 4.1 | 0.7 | 0.7 | 0.7 | 0.8 | 0.2 | 0.4 | 1.3 | 2.8 | 4.1 | 2.1 | 0.5 | 1.6 | 0.1 | 0.4 | 1.7 |
| OTU1933 | 1.1 | 25.2 | 0.0 | 19.0 | 0.7 | 8.8 | 8.6 | 22.4 | 18.7 | 1.7 | 12.9 | 8.1 | 19.0 | 1.5 | 2.3 | 1.5 | 3.1 | 0.5 | 1.5 | 0.1 | 0.4 | 7.5 |
| OTU2101 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.0 | 8.8 | 0.0 | 0.9 | 0.1 | 3.8 | 0.4 | 1.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.7 |
| OTU2194 | 1.2 | 0.0 | 0.0 | 0.1 | 0.3 | 0.0 | 0.0 | 0.0 | 0.0 | 1.4 | 0.0 | 2.6 | 0.0 | 1.1 | 0.4 | 4.3 | 0.0 | 24.9 | 0.9 | 9.6 | 5.8 | 2.5 |
| OTU2195 | 2.5 | 3.1 | 5.0 | 3.4 | 3.2 | 0.4 | 0.9 | 0.6 | 0.2 | 4.2 | 0.2 | 2.9 | 2.3 | 4.1 | 5.3 | 5.9 | 3.2 | 16.0 | 2.5 | 23.6 | 5.4 | 4.5 |
| OTU225 | 4.1 | 2.6 | 8.1 | 2.0 | 6.8 | 3.1 | 2.7 | 2.0 | 2.3 | 0.5 | 1.4 | 2.0 | 0.7 | 0.8 | 0.5 | 1.0 | 0.9 | 0.2 | 0.9 | 0.0 | 0.4 | 2.0 |
| OTU2325 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.2 | 0.1 | 2.2 | 0.0 | 5.9 | 0.0 | 2.6 | 0.5 |
| OTU3099 | 3.1 | 0.2 | 1.7 | 0.4 | 5.0 | 0.8 | 1.0 | 0.0 | 0.2 | 0.0 | 0.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.3 | 0.1 | 0.1 | 0.0 | 0.0 | 0.6 |
| OTU3237 | 0.0 | 0.0 | 4.3 | 0.2 | 2.7 | 0.2 | 1.0 | 0.0 | 0.9 | 0.1 | 1.2 | 0.1 | 0.3 | 0.1 | 0.6 | 0.1 | 1.0 | 0.0 | 0.4 | 0.0 | 0.0 | 0.6 |
| OTU3311 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.3 | 5.7 | 0.3 |
| OTU3624 | 0.8 | 1.2 | 8.9 | 1.6 | 3.8 | 1.2 | 4.1 | 0.2 | 1.4 | 0.3 | 1.3 | 0.5 | 0.8 | 0.4 | 0.8 | 0.1 | 0.5 | 0.0 | 0.3 | 0.0 | 0.0 | 1.3 |
| OTU3987 | 5.8 | 4.1 | 0.1 | 1.1 | 0.0 | 0.3 | 0.0 | 0.0 | 0.1 | 2.4 | 0.3 | 2.7 | 0.4 | 3.3 | 1.6 | 1.7 | 0.2 | 1.8 | 0.1 | 0.3 | 2.0 | 1.3 |
| OTU4556 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.3 | 0.2 | 3.0 | 0.5 | 9.8 | 1.1 | 4.0 | 0.7 | 2.4 | 0.2 | 0.1 | 0.1 | 0.0 | 0.4 | 1.1 |
| OTU4826 | 0.1 | 5.1 | 0.3 | 4.5 | 0.9 | 6.1 | 6.4 | 0.8 | 3.2 | 1.4 | 6.8 | 3.0 | 9.9 | 0.5 | 3.8 | 0.0 | 1.1 | 0.1 | 1.7 | 0.1 | 0.1 | 2.7 |
| OTU5084 | 2.5 | 0.1 | 1.8 | 0.3 | 4.0 | 2.0 | 1.2 | 1.6 | 4.6 | 0.6 | 3.4 | 0.4 | 1.3 | 0.5 | 2.0 | 0.4 | 2.6 | 0.1 | 4.5 | 0.0 | 1.0 | 1.7 |
| OTU5208 | 0.4 | 15.0 | 0.2 | 15.9 | 0.1 | 22.0 | 14.0 | 33.7 | 19.5 | 0.6 | 2.5 | 0.5 | 2.1 | 0.4 | 0.9 | 0.0 | 0.5 | 0.0 | 0.2 | 0.0 | 0.1 | 6.1 |
| OTU5665 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 6.4 | 0.0 | 4.3 | 0.5 | 4.1 | 1.0 | 1.2 | 0.1 | 0.0 | 0.2 | 0.0 | 0.2 | 0.9 |
| OTU5677 | 0.0 | 0.3 | 0.3 | 1.5 | 0.0 | 1.1 | 0.9 | 2.7 | 2.5 | 0.5 | 4.9 | 3.1 | 6.7 | 0.6 | 0.8 | 0.4 | 1.3 | 0.2 | 0.6 | 0.0 | 0.1 | 1.4 |
| OTU5687 | 0.7 | 0.4 | 0.2 | 0.8 | 0.5 | 0.9 | 1.8 | 4.4 | 0.2 | 0.5 | 0.2 | 0.2 | 0.0 | 0.7 | 0.5 | 3.5 | 0.4 | 1.1 | 3.5 | 0.4 | 0.5 | 1.0 |
| OTU588 | 0.0 | 0.0 | 0.0 | 0.1 | 0.5 | 0.5 | 0.3 | 0.1 | 0.3 | 1.2 | 0.1 | 1.0 | 0.1 | 0.6 | 0.2 | 0.3 | 0.1 | 1.4 | 1.6 | 0.1 | 7.3 | 0.8 |
| OTU6110 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.3 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 | 0.3 | 0.0 | 4.7 | 0.1 | 0.3 |
| OTU6134 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.5 | 0.2 | 3.3 | 0.9 | 2.0 | 3.0 | 4.9 | 5.2 | 5.8 | 2.9 | 0.2 | 2.4 | 0.0 | 1.1 | 1.5 |
| OTU851 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.7 | 0.1 | 0.8 | 0.5 | 2.2 | 0.3 | 2.8 | 0.4 | 6.9 | 0.1 | 3.8 | 0.0 | 0.1 | 0.9 |
| OTU964 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 5.0 | 1.8 | 1.9 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.7 | 0.0 | 0.5 |

**Table B.3** Information of the data used in the principal component analysis of the different ecosystems.

| Sample ID | Accession info. | | Sample name | Sample description | Reactor type | Method | Country | No. of bacterial 16S rRNA gene sequences | Reference |
|---|---|---|---|---|---|---|---|---|---|
| 1 | NCBI | SRA026842 | Suo-Jin-Cun (Nanjing, PRC) | Full-scale activated sludge treating domestic wastewater | Anoxic/aerobic | Pyrosequencing | China | 20964 | Zhang et al., 2012 |
| 2 | | | Tuan-Dao (Qingdao, PRC) | | Anaerobic/anoxic/aerobic | | China | 24456 | |
| 3 | | | Ha-Er-Bin (Haerbin, PRC) | | Anoxic/aerobic | | China | 22603 | |
| 4 | | | Min-Hang (Shanghai, PRC) | | Anoxic/aerobic | | China | 21412 | |
| 5 | | | Bei-Xiao-He (Beijing, PRC) | | Anaerobic/anoxic/aerobic+MBR | | China | 23621 | |
| 6 | | | Long-Wang-Zui (Wuhan, PRC) | | Anaerobic/anoxic/aerobic | | China | 22497 | |
| 7 | | | Da-Tan-Sha (Guangzhou, PRC) | | Anaerobic/anoxic/aerobic | | China | 22098 | |
| 8 | | | Ulu Pandan (Singapore) | | Conventional activated sludge+MBR | | China | 23967 | |
| 9 | | | Columbia Regional (Columbia,USA) | | Conventional activated sludge | | USA | 25500 | |
| 10 | | | Potato Creek (Griffin, USA) | | Oxidation ditch | | USA | 26383 | |
| 11 | | | Guelph (Guelph, Canada) | | Conventional activated sludge | | Canada | 22098 | |
| 12 | | | Sha-Tin 1 (Hong Kong, PRC) | | Anoxic/aerobic | | China | 28260 | |
| 13 | | | Sha-Tin 2 (Hong Kong, PRC) | | Anoxic/aerobic | | China | 26900 | |
| 14 | | | Stanley (Hong Kong, PRC) | | Anoxic/aerobic | | China | 24796 | |
| 15 | Genbank | EF222481–EF248596 | Brazil | Soil | - | Pyrosequencing | Brazil | 26140 | Roesch et al., 2007 |
| 16 | | EF276845–EF308590 | Illinois | | - | | USA, Illinois | 31818 | |
| 17 | | EF308591–EF361836 | Canada | | - | | Canada | 53533 | |
| 18 | Genbank | GU481685-GU549391 | Suspended sample | Full-scale fixed-film activated | Fixed-film activated sludge system (IFAS) | Sanger | Korea | 86 | Kwon et al., 2010 |
| 19 | | | Attached samples | | | | | 82 | |

**Table B.3 (cont.)**

| Sample ID | Accession info. | | Sample name | Sample description | Reactor type | Method | Country | No. of bacterial 16S rRNA gene sequences | Reference |
|---|---|---|---|---|---|---|---|---|---|
| 20 | | | Suspended sample | sludge treating domestic wastewater | | Pyrosequencing | | 23536 | |
| 21 | | | Attached samples | | | | | 44003 | |
| 22 | Genbank | SRX005900-SRX005907 | Apr-05-JI | Sewage | Conventional activated sludge | Pyrosequencing | USA, Wisconsin | 17338 | McLellan et al., 2010 |
| 23 | | | Apr-07-JI | | | | | 21352 | |
| 24 | | | Aug-07-JI | | | | | 31877 | |
| 25 | | | Dec-07-JI | | | | | 26503 | |
| 26 | | | Apr-05-SS | | | | | 28684 | |
| 27 | | | Apr-07-SS | | | | | 34080 | |
| 28 | | | Aug-07-SS | | | | | 24463 | |
| 29 | | | Dec-07-SS | | | | | 30793 | |
| 30 | NCBI | AB479546–AB479708 AB618290–AB618481 AB818471–AB818472 | 238B1 | Full-scale UASB-DHS treating domestic wastewater | DHS | Sanger | Japan | 93 | Kubota et al., 2014 |
| 31 | | | 238B4 | | | | | 99 | |
| 32 | | | 238B8 | | | | | 96 | |
| 33 | | | 441B1 | | | | | 127 | |
| 34 | | | 441B2 | | | | | 165 | |
| 35 | | | 441B3 | | | | | 120 | |

**Table B.4** Coverage and diversity of the microbial communities of the samples collected from the DHS reactor.

| Sample ID | | | No. of reads | No. of OTUs[a] | Chao1 richness estimator[b] | Good's coverage[c] | Equitability[b, d] | PD[b] | Shannon[b] |
|---|---|---|---|---|---|---|---|---|---|
| Phase | Day | Location | | | | | | | |
| I | 82 | Up | 2296 | 325 | 577 | 0.93 | 0.80 | 36.3 | 6.59 |
| | 136 | Up | 1070 | 133 | 203 | 0.93 | 0.71 | 55.1 | 6.53 |
| | | Low | 6357 | 585 | 768 | 0.95 | 0.69 | 24.8 | 5.37 |
| | 196 | Up | 1570 | 203 | 377 | 0.94 | 0.71 | 54.0 | 6.74 |
| | | Low | 1054 | 209 | 332 | 0.89 | 0.85 | 15.6 | 4.77 |
| | 258 | Up | 3165 | 285 | 466 | 0.96 | 0.69 | 55.5 | 6.84 |
| | | Low | 1573 | 241 | 510 | 0.91 | 0.75 | 52.3 | 5.74 |
| | 373 | Up | 2846 | 196 | 327 | 0.97 | 0.55 | 65.8 | 7.51 |
| | | Low | 2003 | 256 | 439 | 0.94 | 0.67 | 59.1 | 6.59 |
| | 454 | Up | 9509 | 849 | 934 | 0.94 | 0.69 | 58.8 | 7.19 |
| | | Low | 11351 | 967 | 939 | 0.94 | 0.69 | 56.1 | 6.33 |
| II | 528 | Up | 8516 | 839 | 968 | 0.95 | 0.73 | 53.6 | 6.22 |
| | | Low | 7272 | 743 | 967 | 0.94 | 0.66 | 55.2 | 7.28 |
| | 602 | Up | 9145 | 792 | 887 | 0.95 | 0.72 | 32.6 | 5.51 |
| | | Low | 9338 | 870 | 950 | 0.95 | 0.76 | 39.1 | 5.29 |
| III | 648 | Up | 8168 | 785 | 933 | 0.94 | 0.73 | 29.8 | 5.84 |
| | | Low | 5601 | 709 | 975 | 0.94 | 0.77 | 23.6 | 6.30 |
| IV | 723 | Up | 7351 | 661 | 822 | 0.95 | 0.62 | 25.0 | 4.16 |
| | | Low | 6681 | 866 | 1125 | 0.93 | 0.77 | 29.4 | 6.14 |
| V | 798 | Up | 7765 | 451 | 540 | 0.97 | 0.61 | 33.5 | 5.36 |
| | | Low | 8320 | 810 | 933 | 0.95 | 0.76 | 14.0 | 3.90 |

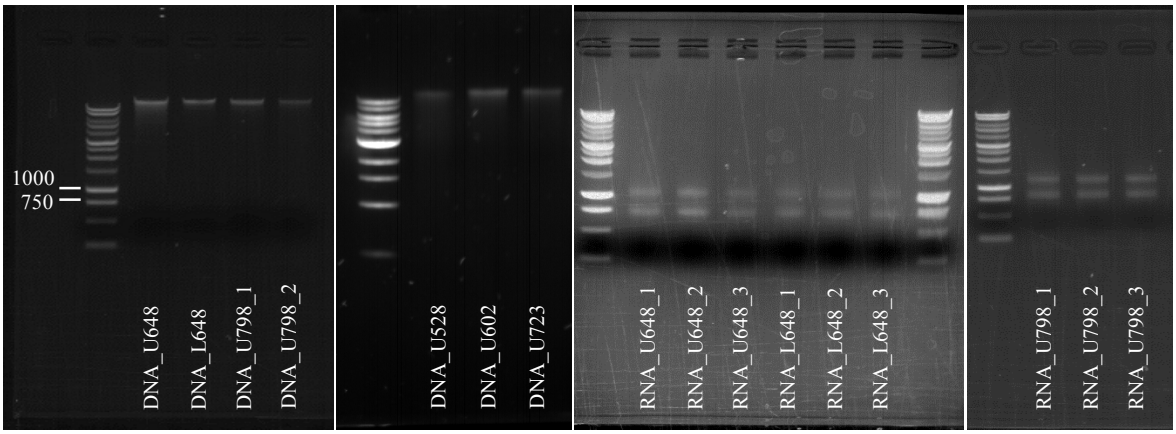
[a] Operational taxonomic units (OTU) were defined at a 97% similarity threshold.

[b] Chao1 richness estimators at 95% confidence interval, Equitability, PD, and Shannon diversity indices were calculated using QIIME pipeline.

[c] Good's coverage was calculated using the equation: [1-(n/N)], where n is the number of singleton reads and N is the total number of reads.

[d] Equitability index was a measure of evenness.

## APPENDIX C: SUPPLEMENTAL MATERIALS IN CHAPTER 4




* Ladder: 1 kbp ladder

**Figure C.1** Electrophoresis gel of extracted genomic DNA (left) and the triplicates of total RNA (right) with a 1 kb DNA ladder.

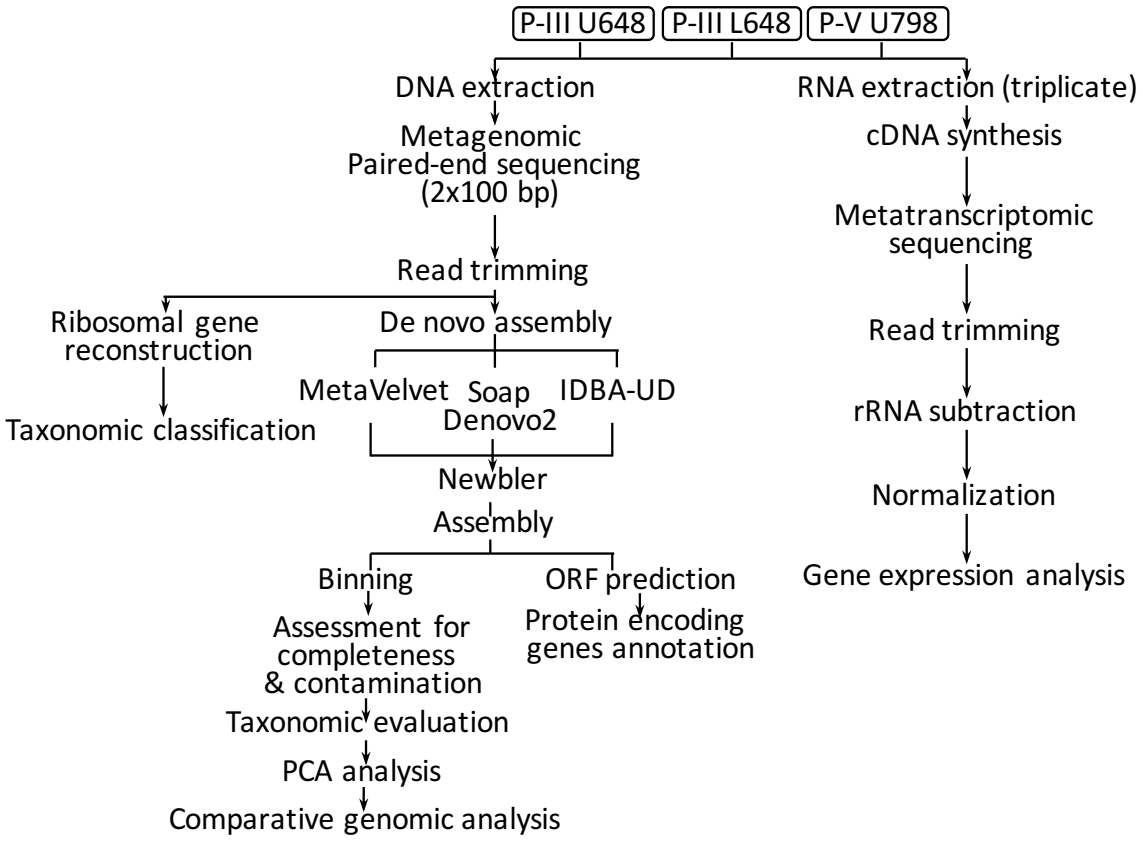



**Figure C.2** Overview of the bioinformatic analytical workflow.

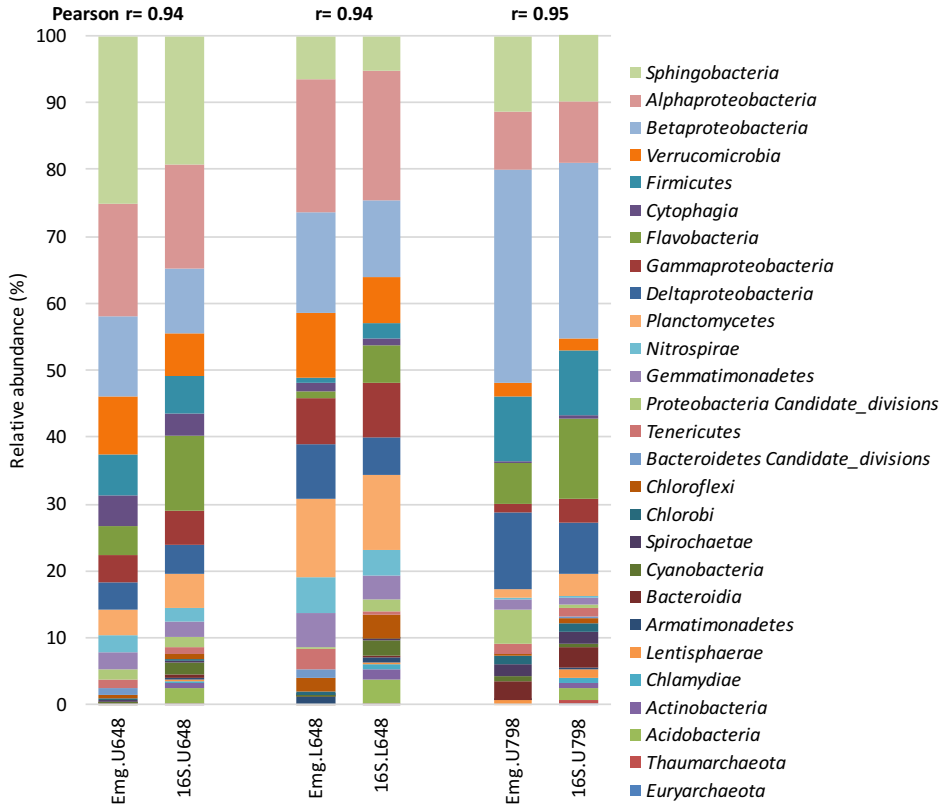



**Figure C.3** Microbial community composition at the phylum/class level in U648, L648, and U798 metagenomic datasets. Relative abundances of reads blasted to the reconstructed 16S rRNA genes using EMIRGE (Emg) and reads blasted to the SILVA rRNA gene database (release 119) are shown with Pearson correlation coefficient.

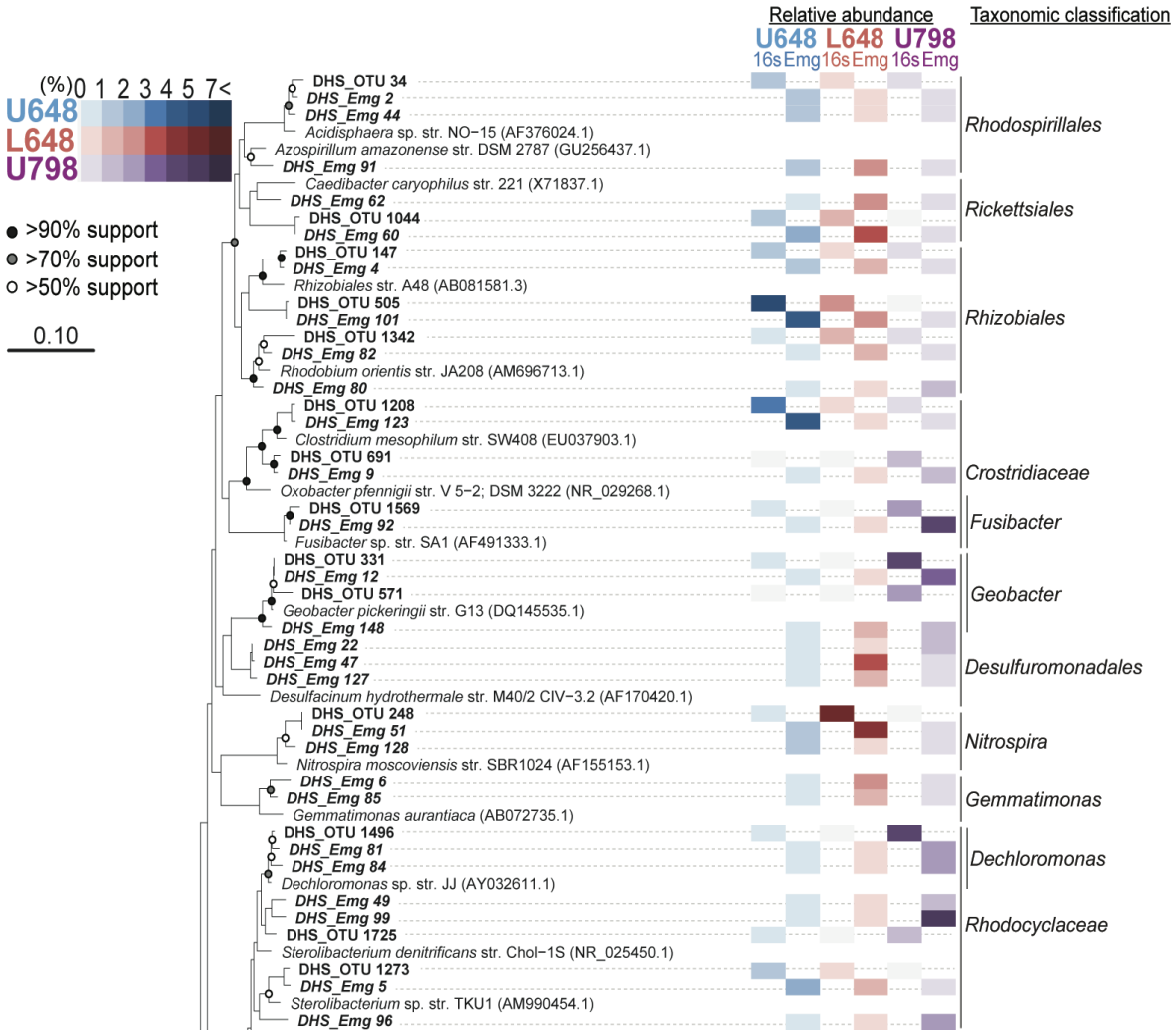



**Figure C.4** Microbial phylogenetic composition in the DHS reactor. In the 16S rRNA gene-based phylogenetic tree (bootstrap 1000: >90% black node, >70% gray node, and >50% white node), DHS_Emg refers to reconstructed ribosomal sequences using EMIRGE, and DHS_OTU refers to representative operational taxonomic units (OTU) from amplified 16S rRNA gene analysis by pyrosequencing. The relative abundance of the pyrosequencing OTUs (16S) and reads from the metagenome datasets blasted to the reconstructed 16S rRNA genes using EMIRGE (Emg) is indicated with color codes. The relative abundance is normalized to total number of bacterial 16s rRNA gene sequences in each dataset.

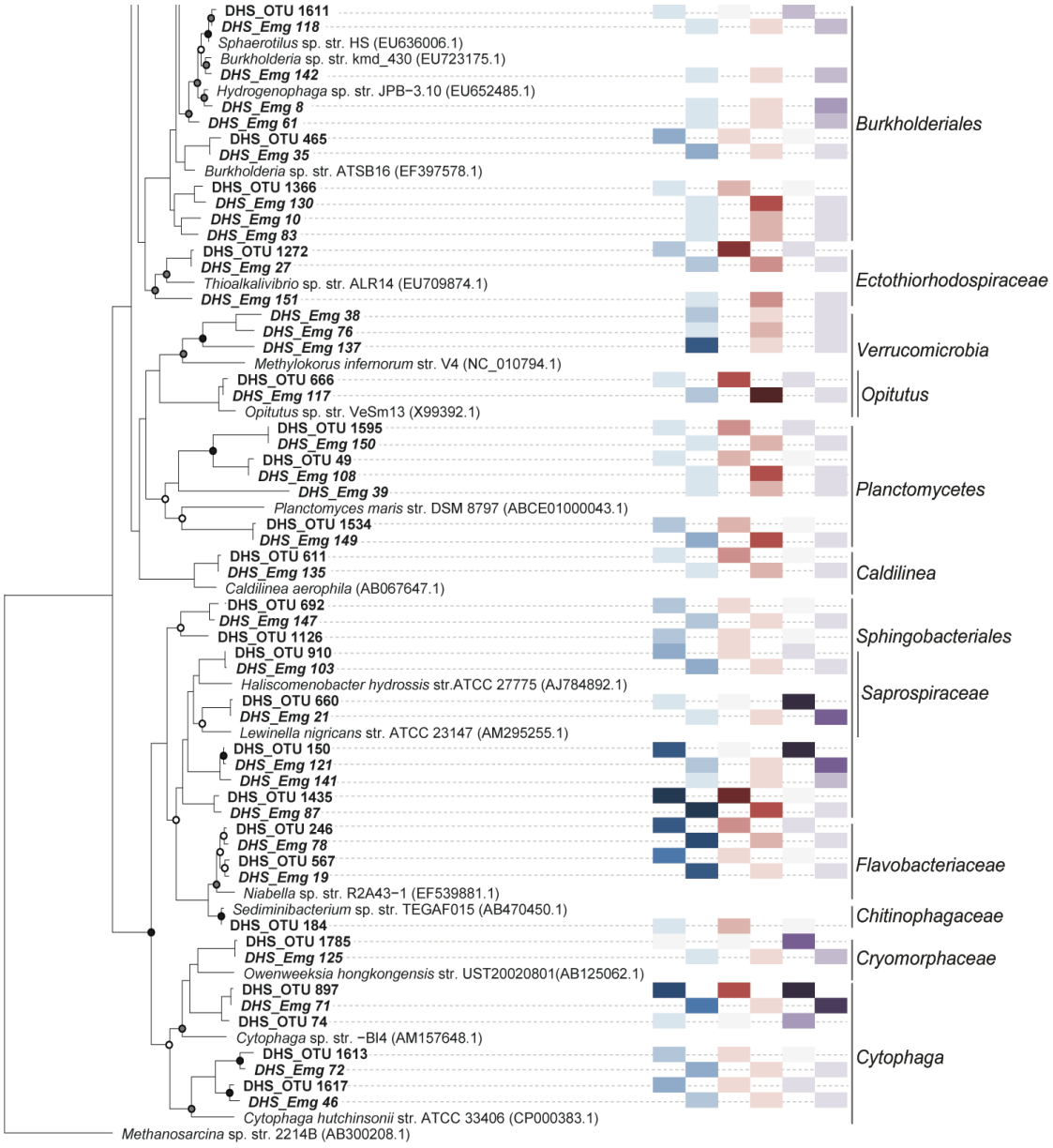



**Figure C.4 (cont.)**

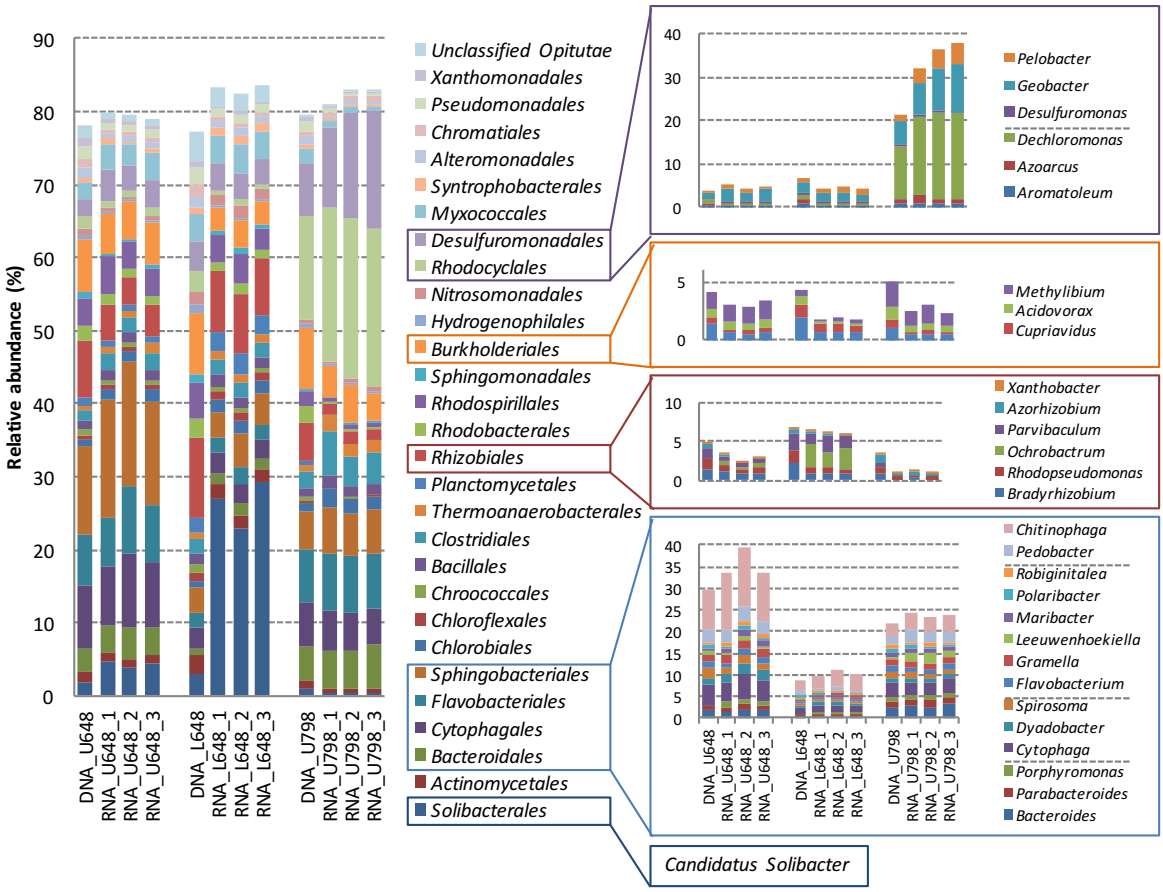



**Figure C.5** Microbial community compositions determined by protein encoding gene-based analyses in the metagenomes and metatranscriptomes. Taxonomic classification was assigned at the order level for the entire datasets (on left), and the genus level classifications were further indicated in the dominant order groups (on right).

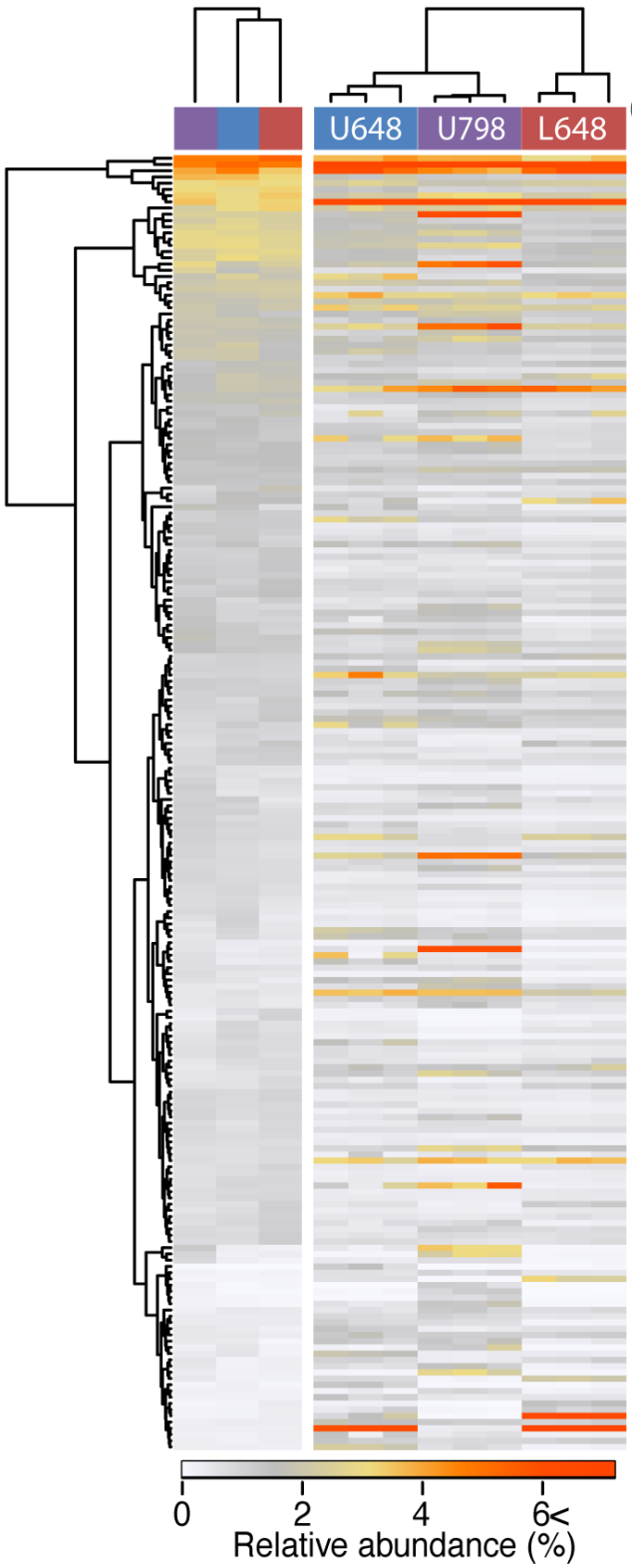



**Figure C.6** SEED subsystem level 3 that is significantly abundant at the 98% confidence level. Three columns on the left indicated metagenomic library, and the three triplicate columns on the right indicated each metatranscriptional library that is relative to the corresponding the metagenomic libraries.

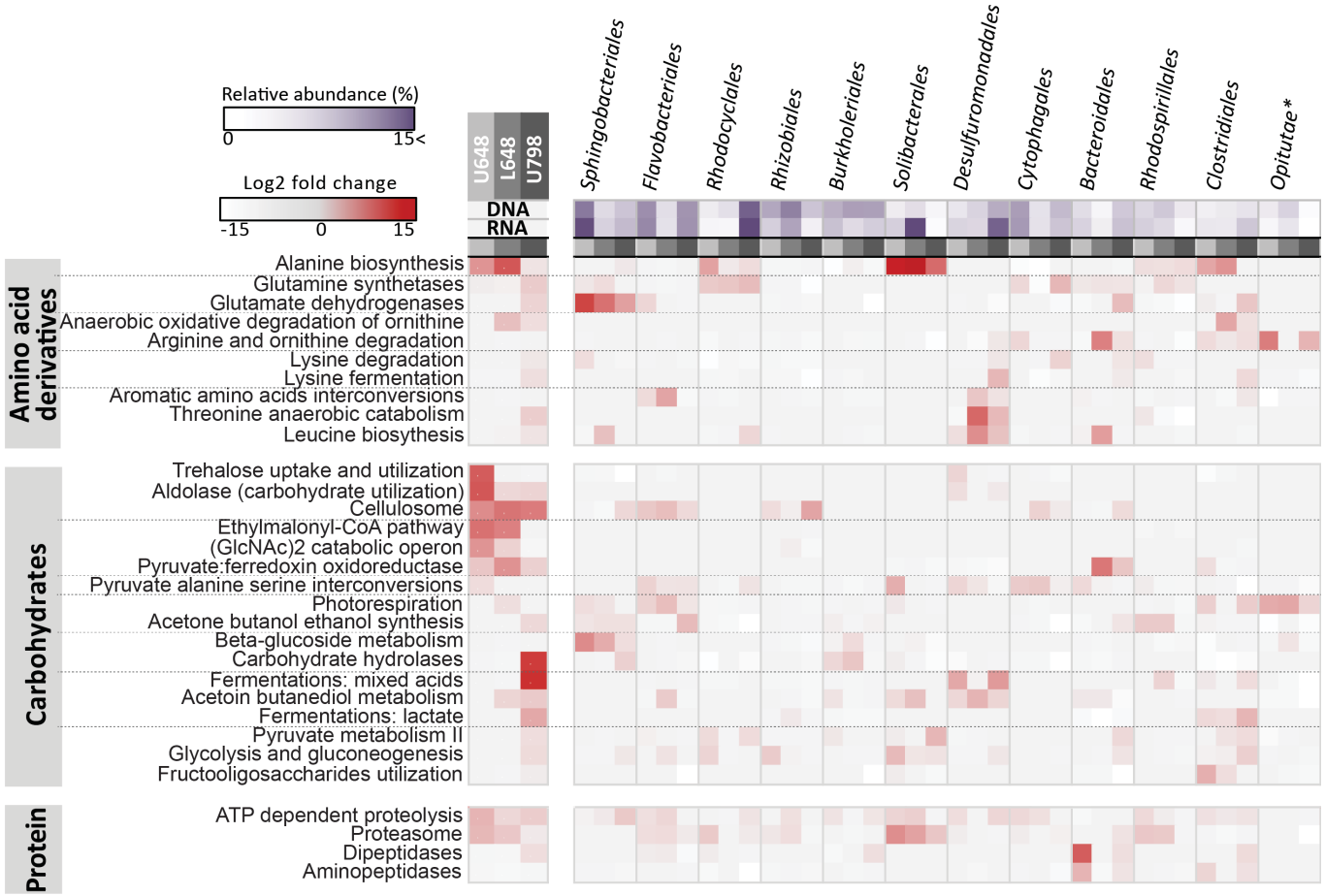



**Figure C.7** Global analysis of metabolic potential and functional activities in the DHS communities. Genomic relative abundance and expression profile of dominant orders at the SEED subsystem level 3. The genomic and transcriptomic relative abundance in percentage was indicated by white-purple scale color codes for the ten most dominant orders in sequence from left to right, and the transcriptional activity in terms of ratio of the transcriptional relative abundance to the genomic relative abundance was represented in log 2-fold change by white-red scale color codes. Light gray indicates U648, gray indicates L648, and dark gray indicates U798. The tree columns on the left side showed the entire transcriptional activity of each datasets, and the rest of the columns on the right side represented the transcriptional activity by each dominant order. * Unclassified *Opitutae* at the order level.

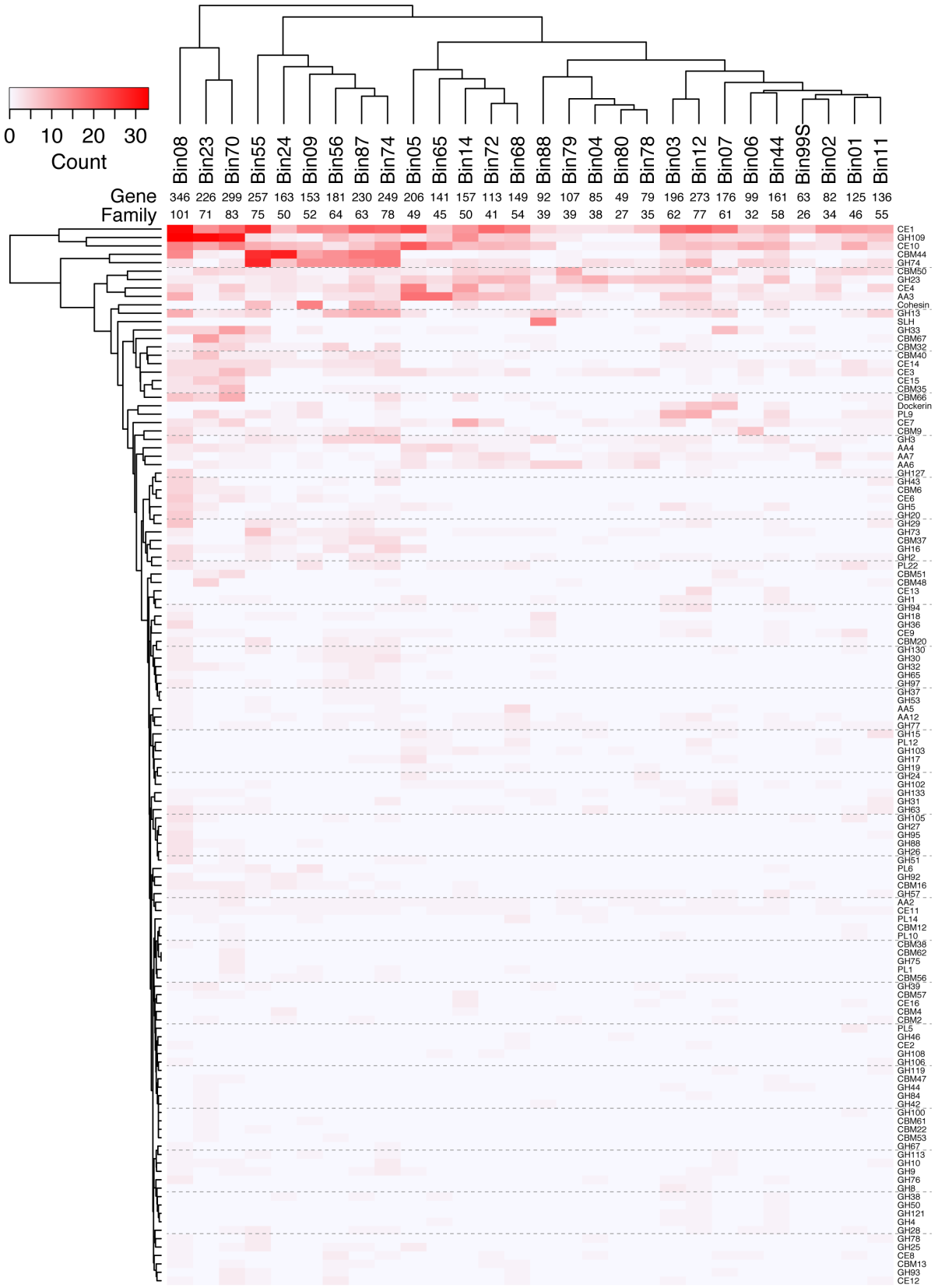



**Figure C.8** Heatmap reflecting the putative genes of carbohydrate-active enzyme families in the dominant draft genomes.

**Table C.1** Information of genomic and transcriptomic datasets and subtraction of rDNA and rRNA.

| Sample | Library name | Pre-QC no. of reads[a] | Post-QC no. of reads[a] | rRNA reads | Coding DNA reads and non-rRNA reads | Coding DNA reads and non-rRNA reads aligned to the assembly (>95% similarity) | | MG-RAST ID[b] |
|---|---|---|---|---|---|---|---|---|
| Genome | U648 | 89,172,658 | 66,692,412 (74.8%) | 119,188 | 66,573,224 (99.8%) | 38,902,409 | 58.40% | 4623852.3 |
| | L648 | 106,079,512 | 77,214,480 (72.8%) | 146,348 | 77,068,132 (99.8%) | 39,653,970 | 51.50% | 4623716.3 |
| | U798 | 195,173,938 | 143,790,358 (73.7%) | 320,343 | 143,470,015 (99.8%) | 83,954,088 | 58.50% | 4623717.3 |
| Transcriptome | U648_1 | 14,559,832 | 13,283,368 (91.2%) | 9,919,717 | 3,363,651 (25.3%) | 980,798 | 29.20% | 4622228.3 |
| | U648_2 | 14,362,838 | 13,130,084 (91.4%) | 9,218,856 | 3,911,228 (29.8%) | 1,336,392 | 34.20% | 4622229.3 |
| | U648_3 | 15,113,751 | 13,773,593 (91.1%) | 7,251,092 | 6,522,501 (47.4%) | 2,433,721 | 37.30% | 4622230.3 |
| | L648_1 | 15,177,201 | 13,784,821 (90.8%) | 9,375,257 | 4,409,564 (32.0%) | 1,003,397 | 22.80% | 4622226.3 |
| | L648_2 | 13,890,712 | 12,646,690 (91.0%) | 3,623,167 | 9,023,523 (71.4%) | 2,241,401 | 24.80% | 4622351.3 |
| | L648_3 | 14,599,754 | 13,217,552 (90.5%) | 2,983,808 | 10,233,744 (77.4%) | 2,606,952 | 25.50% | 4622227.3 |
| | U798_1 | 18,183,052 | 16,651,704 (91.6%) | 13,998,258 | 2,653,446 (15.9%) | 965,784 | 36.40% | 4622352.3 |
| | U798_2 | 18,297,640 | 16,744,277 (91.5%) | 12,864,623 | 3,879,654 (23.2%) | 1,495,042 | 38.50% | 4622234.3 |
| | U798_3 | 15,842,203 | 14,589,031 (92.1%) | 9,605,762 | 4,983,269 (34.2%) | 1,987,216 | 39.90% | 4622235.3 |

a. QC, quality control

**b.** The listed libraries were submitted under the MG-RAST project (ID: mgp9993).

**Table C.2** Pre- and final- assemblies and their statistics.

| Dataset | Assembler | k-value | Post-QC no. of reads | Assembled reads | Total contig size | No. of contig | Max. contig size | N50 | N90 |
|---|---|---|---|---|---|---|---|---|---|
| U798-1 | Velvet | 61 | 72,824,136 | 38,228,542 | 141,585,064 | 226,253 | 295,973 | 1,521 | 272 |
| | SOAPdenovo2 | 75 | 72,824,136 | 30,399,419 | 104,645,387 | 180,721 | 296,063 | 703 | 257 |
| | IDBA-UD | 55-95 | 72,824,136 | 45,412,756 | 235,275,495 | 114,100 | 770,539 | 7,347 | 679 |
| U798-2 | Velvet | 65 | 70,966,222 | 25,032,687 | 102,094,822 | 160,681 | 295,977 | 1,624 | 272 |
| | SOAPdenovo2 | 65 | 70,966,222 | 27,416,319 | 151,940,000 | 307,211 | 279,983 | 606 | 218 |
| | SOAPdenovo2 | 73 | 70,966,222 | 22,103,187 | 96,872,247 | 165,165 | 224,654 | 799 | 250 |
| | IDBA-UD | 59-99 | 70,966,222 | 31,570,350 | 178,946,071 | 91,167 | 474,299 | 9,572 | 611 |
| U648 | Velvet | 49 | 66,692,412 | 40,703,566 | 276,956,125 | 784,337 | 384,151 | 607 | 176 |
| | SOAPdenovo2 | 69 | 66,692,412 | 26,384,373 | 136,996,898 | 306,880 | 415,117 | 427 | 230 |
| | SOAPdenovo2 | 83 | 66,692,412 | 14,799,218 | 35,239,331 | 42,786 | 249,487 | 1,968 | 284 |
| | IDBA-UD | 49-79 | 66,692,412 | 39,109,187 | 264,379,966 | 161,141 | 750,416 | 6,449 | 441 |
| L648 | Velvet | 51 | 77,214,480 | 43,954,847 | 314,226,368 | 942,534 | 426,095 | 486 | 178 |
| | Velvet | 63 | 77,214,480 | 29,572,859 | 141,284,737 | 237,453 | 136,179 | 1,204 | 257 |
| | SOAPdenovo2 | 49 | 77,214,480 | 46,542,214 | 406,086,569 | 1,135,112 | 414,414 | 384 | 168 |
| | SOAPdenovo2 | 73 | 77,214,480 | 24,096,435 | 123,339,653 | 236,266 | 65,682 | 586 | 246 |
| | IDBA-UD | 45-95 | 77,214,480 | 43,754,598 | 341,435,213 | 226,361 | 1,008,602 | 3,579 | 475 |
| Final assembly | Newbler | | | | 440,563,666 | 45,392 | 1,008,159 | 25,560 | 3,186 |

**Table C.3** Assembly statistic of metagenomic datasets.

| | Total scaffolds | Contigs > 300 bp | Contigs > 1 kb | Contigs > 50 kb | Contigs > 100 kb |
|---|---|---|---|---|---|
| Total Base (Mbp) | 440.6 | 440.5 | 440.2 | 165.9 | 110.9 |
| Number of contigs | 45,392 | 45,037 | 44,592 | 1,354 | 551 |
| Mean length (bp) | 9,705 | 9,781 | 9,872 | 122,523 | 201,273 |
| N50 (bp) | 25,560 | 25,560 | 25,560 | 143,676 | 217,873 |
| N90 (bp) | 3,186 | 3,186 | 3,186 | 60,804 | 114,858 |
| Largest Scaffold (bp) | 1,008,159 | 1,008,159 | 1,008,159 | 1,008,159 | 1,008,159 |

**Table C.4** MG-RAST annotation of Assembly (contigs > 300 bp).

| Item | Statistics |
|---|---|
| Contigs | 45,037 |
| Average length (bp) | 9,780 ± 27,941 |
| Total length (bp) | 440,496,337 |
| Predicted ORFs | 272,083 |
| Annotated | 200,515 |
| rRNAs | 511 |
| Functional category | 166,555 |
| Unrecognized | 71,568 |

**Table C.5** Summary of protein encoding genes annotated by SEED subsystem.

| Total number of protein encoding genes | | 95,002 |
|---|---|---|
| Summary of protein encoding genes (length) | Minimum | 300 |
| | 1st Quantile | 907 |
| | Median | 1,458 |
| | Mean | 1,875 |
| | 3rd Quantile | 2,389 |
| | Maximum | 29,070 |

**Table C.6** Protein encoding genes aligned with coding-DNA and non-rRNA sequences.

| Genomic sample | Present protein encoding genes | | Transcriptomic sample | Expressed protein encoding genes | |
|---|---|---|---|---|---|
| U648 | 81,025 | 85.3% | U648_1 | 34,291 | 36.1% |
| | | | U648_2 | 32,684 | 34.4% |
| | | | U648_3 | 47,054 | 49.5% |
| L648 | 76,121 | 80.1% | L648_1 | 27,442 | 28.9% |
| | | | L648_2 | 38,546 | 40.6% |
| | | | L648_3 | 38,605 | 40.6% |
| U798 | 86,674 | 91.2% | U798_1 | 18,647 | 19.6% |
| | | | U798_2 | 25,886 | 27.2% |
| | | | U798_3 | 27,926 | 29.4% |

**Table C.7** Metagenomic and metatranscriptomic statistics of the ten most dominant homologous orders at the subsystem level 3.

| | U648 | | L648 | | U798 | |
|---|---|---|---|---|---|---|
| Order | MG[a] | MT[a] | MG[a] | MT[a] | MG[a] | MT[a] |
| *Sphingobacteriales* | 11.9 | 15.7 | 3.3 | 4.2 | 5.1 | 6.1 |
| *Flavobacteriales* | 10.0 | 10.5 | 2.9 | 2.7 | 9.4 | 9.6 |
| *Rhodocyclales* | 1.7 | 0.9 | 2.8 | 0.8 | 14.1 | 21.6 |
| *Rhizobiales* | 7.8 | 4.3 | 11.2 | 8.1 | 5.2 | 1.4 |
| *Burkholderiales* | 7.1 | 5.5 | 8.4 | 3.3 | 8.3 | 4.5 |
| *Solibacterales* | 1.7 | 4.3 | 3.0 | 26.3 | 1.0 | 0.4 |
| *Desulfuromonadales* | 2.3 | 3.9 | 4.0 | 3.6 | 7.4 | 13.9 |
| *Cytophagales* | 8.6 | 9.0 | 2.6 | 2.7 | 6.1 | 5.2 |
| *Bacteroidales* | 3.1 | 4.1 | 1.0 | 1.6 | 4.8 | 5.5 |
| *Rhodospirillales* | 3.7 | 4.3 | 4.9 | 3.5 | 2.0 | 0.6 |
| *Clostridiales* | 1.4 | 2.1 | 1.9 | 2.1 | 2.5 | 4.9 |
| *Optitutae** | 1.9 | 0.8 | 4.1 | 2.3 | 0.4 | 0.1 |

a. Percentage based on each dataset

**Table C.8** Metagenomic bins that contributes top 50 % of abundance in genomic presence and transcriptomic expression.

| Bin_ID | U648 | | | | L648 | | | | U798 | | | | Marker lineage | Marker gene copies | | | | Complete-ness[c] | Contami-nation[c] | Size (Mb) | Contig count | ORF[e] | PEG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MG[a] | MT1[ab] | MT2[ab] | MT3[ab] | MG[a] | MT1[ab] | MT2[ab] | MT3[ab] | MG[a] | MT1[ab] | MT2[ab] | MT3[ab] | | 0 | 1 | 2 | 3 | | | | | | |
| Bin01 | 9.3 | 2.8 | 2.6 | 3.7 | 38.4 | 22.2 | 19.7 | 18.0 | 0.1 | 0.1 | 0.1 | 0.0 | *g_Opitutus* | 1 | 226 | 3 | 0 | 99.3 | 1.4 | 3.6 | 34 | 2294 | 872 |
| Bin02 | 9.6 | 2.4 | 1.2 | 2.5 | 41.7 | 8.6 | 12.1 | 7.9 | 0.7 | 0.2 | 0.2 | 0.1 | *o_Burkholderiales* | 19 | 403 | 2 | 1 | 98.6 | 1.0 | 3.8 | 22 | 2448 | 958 |
| Bin03 | 1.3 | 5.1 | 2.0 | 4.2 | 24.4 | 46.1 | 45.0 | 41.4 | 0.0 | 0.1 | 0.1 | 0.0 | *o_Desulfuromonadales* | 37 | 148 | 5 | 0 | 86.3 | 3.8 | 5.2 | 372 | 3351 | 877 |
| Bin04 | 2.9 | 2.0 | 1.5 | 2.4 | 17.3 | 17.7 | 19.5 | 18.3 | 0.0 | 0.0 | 0.0 | 0.0 | *g_Nitrospira* | 7 | 170 | 4 | 0 | 95.9 | 2.8 | 3.1 | 30 | 2002 | 612 |
| Bin05 | 21.1 | 14.7 | 9.5 | 7.7 | 39.5 | 15.7 | 13.5 | 11.4 | 0.5 | 0.4 | 0.5 | 0.2 | *o_Rhodospirillales* | 7 | 325 | 4 | 0 | 97.5 | 1.7 | 7.8 | 103 | 5701 | 1716 |
| Bin06 | 1.7 | 4.4 | 2.2 | 5.0 | 12.8 | 38.4 | 26.7 | 36.3 | 3.5 | 0.1 | 0.9 | 0.6 | *g_Gemmatimonas* | 3 | 143 | 1 | 0 | 96.7 | 1.1 | 3.2 | 12 | 1715 | 523 |
| Bin07 | 0.4 | 0.4 | 0.1 | 0.3 | 14.3 | 2.1 | 1.6 | 1.7 | 0.4 | 0.0 | 0.0 | 0.0 | *f_Planctomycetaceae* | 0 | 141 | 2 | 0 | 100.0 | 2.3 | 4.9 | 99 | 3205 | 693 |
| Bin08 | 61.9 | 24.0 | 18.4 | 23.7 | 24.1 | 10.8 | 13.9 | 10.0 | 3.5 | 0.3 | 1.0 | 0.3 | *g_Chitinophaga* | 1 | 297 | 2 | 1 | 99.5 | 1.5 | 6.7 | 82 | 4610 | 1203 |
| Bin09 | 25.1 | 47.6 | 57.7 | 40.8 | 11.2 | 21.4 | 22.6 | 19.7 | 0.1 | 0.1 | 0.1 | 0.0 | *g_Haliscomenobacter* | 2 | 300 | 0 | 0 | 99.0 | 0.0 | 4.2 | 72 | 2809 | 636 |
| Bin11 | 0.5 | 0.7 | 0.7 | 3.6 | 9.8 | 14.1 | 19.0 | 15.8 | 0.1 | 0.0 | 0.0 | 0.0 | *f_Planctomycetaceae* | 14 | 128 | 1 | 0 | 95.5 | 1.1 | 4.0 | 106 | 2546 | 687 |
| Bin12 | 3.3 | 2.3 | 1.7 | 3.4 | 17.1 | 9.1 | 10.4 | 9.2 | 0.9 | 0.1 | 0.1 | 0.1 | *o_Desulfuromonadales* | 20 | 219 | 8 | 0 | 90.8 | 2.6 | 6.4 | 335 | 4534 | 1190 |
| Bin14 | 1.5 | 0.9 | 0.5 | 0.6 | 17.8 | 2.3 | 2.9 | 2.1 | 0.1 | 0.0 | 0.1 | 0.1 | *o_Burkholderiales* | 6 | 411 | 2 | 0 | 98.4 | 0.4 | 6.5 | 204 | 4327 | 1392 |
| Bin23 | 14.7 | 10.6 | 10.9 | 8.8 | 7.0 | 5.4 | 9.0 | 6.9 | 0.2 | 0.1 | 0.0 | 0.0 | *f_Verrucomicrobiaceae* | 2 | 223 | 5 | 0 | 98.7 | 3.4 | 5.5 | 143 | 3225 | 833 |
| Bin24 | 24.5 | 22.8 | 23.4 | 22.5 | 5.7 | 6.8 | 8.0 | 7.7 | 53.4 | 23.7 | 28.5 | 17.6 | *g_Fluviicola* | 1 | 274 | 3 | 0 | 99.5 | 1.6 | 4.4 | 71 | 2449 | 704 |
| Bin44 | 5.6 | 56.7 | 28.3 | 39.2 | 2.8 | 9.6 | 10.0 | 10.9 | 18.6 | 1.4 | 3.0 | 3.6 | *o_Desulfuromonadales* | 17 | 227 | 3 | 0 | 93.7 | 1.9 | 5.4 | 223 | 3319 | 792 |
| Bin55 | 8.7 | 23.3 | 40.6 | 36.9 | 0.9 | 2.8 | 2.4 | 3.2 | 0.0 | 0.1 | 0.1 | 0.0 | *g_Haliscomenobacter* | 20 | 274 | 8 | 0 | 91.8 | 2.0 | 6.1 | 236 | 3409 | 924 |
| Bin56 | 56.7 | 62.7 | 66.5 | 51.6 | 0.6 | 3.5 | 11.0 | 11.5 | 0.1 | 0.1 | 0.0 | 0.0 | *g_Chitinophaga* | 2 | 299 | 1 | 0 | 99.0 | 0.5 | 4.7 | 71 | 3310 | 905 |
| Bin65 | 10.3 | 36.9 | 28.0 | 22.2 | 0.4 | 0.7 | 1.8 | 1.0 | 9.8 | 0.6 | 3.3 | 2.2 | *f_Acetobacteraceae* | 38 | 292 | 4 | 2 | 86.4 | 1.3 | 4.6 | 544 | 3603 | 1337 |
| Bin68 | 18.8 | 23.6 | 25.7 | 30.1 | 0.6 | 0.7 | 2.1 | 1.5 | 3.7 | 1.0 | 1.3 | 0.8 | *o_Burkholderiales* | 56 | 342 | 28 | 0 | 90.7 | 8.5 | 6.2 | 760 | 4564 | 1852 |
| Bin70 | 28.1 | 1.3 | 1.0 | 1.4 | 0.2 | 0.0 | 0.0 | 0.0 | 4.6 | 0.3 | 0.3 | 0.1 | *f_Verrucomicrobiaceae* | 1 | 222 | 6 | 0 | 99.3 | 4.5 | 6.0 | 57 | 3641 | 931 |
| Bin72 | 1.4 | 1.5 | 1.3 | 1.4 | 0.2 | 0.2 | 0.3 | 0.3 | 28.8 | 21.4 | 28.2 | 16.0 | *o_Burkholderiales* | 15 | 408 | 3 | 0 | 98.3 | 0.8 | 6.1 | 97 | 3935 | 1763 |
| Bin74 | 9.6 | 44.2 | 64.7 | 51.2 | 0.1 | 0.4 | 0.3 | 0.4 | 46.2 | 44.9 | 43.9 | 50.1 | *g_Haliscomenobacter* | 1 | 299 | 2 | 0 | 99.5 | 0.5 | 5.7 | 109 | 3944 | 899 |
| Bin78 | 0.1 | 0.4 | 0.2 | 0.3 | 0.1 | 0.2 | 0.1 | 0.1 | 40.6 | 77.6 | 78.5 | 78.6 | *g_Dechlomonas* | 59 | 343 | 23 | 0 | 90.8 | 3.9 | 3.9 | 122 | 2444 | 1226 |
| Bin79 | 0.5 | 1.2 | 2.6 | 0.8 | 0.0 | 0.5 | 0.2 | 0.1 | 36.6 | 91.1 | 100.5 | 110.8 | *g_Geobacter* | 2 | 230 | 15 | 0 | 99.3 | 3.5 | 4.7 | 93 | 2685 | 1121 |
| Bin80 | 0.2 | 0.4 | 0.2 | 0.3 | 0.0 | 0.1 | 0.1 | 0.1 | 22.3 | 34.0 | 38.2 | 38.0 | *g_Dechlomonas* | 181 | 241 | 3 | 0 | 62.6 | 1.0 | 2.8 | 219 | 1878 | 927 |
| Bin87 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 21.9 | 79.8 | 73.1 | 70.9 | *g_Haliscomenobacter* | 2 | 284 | 11 | 1 | 99.0 | 5.7 | 6.0 | 364 | 3194 | 1021 |
| Bin88 | 1.5 | 0.3 | 0.7 | 0.4 | 0.0 | 0.0 | 0.0 | 0.0 | 10.5 | 65.1 | 42.1 | 45.0 | *o_Clostridiales* | 5 | 232 | 10 | 0 | 96.9 | 4.7 | 3.4 | 304 | 2204 | 736 |
| Bin99S | 0.6 | 12.6 | 8.3 | 11.3 | 2.6 | 147.6 | 113.8 | 158.8 | 0.1 | 0.1 | 0.2 | 0.0 | *s_Ca.Solibacter* | 113 | 72 | 0 | 0 | 35.7 | 0.0 | 1.6 | 213 | 1221 | 381 |

**Table C.8 (cont.)**

a. Normalized abundance of genomic (MG) and transcriptomic (MT) datasets aligned to the protein-coding genes. The bins contributes top 50% of abundance for each dataset highlighted in bold.
b. Average values of the triplicates.
c. Completeness and contamination of genome bins were assessed using CheckM. Bins that were less than 60% complete or with greater than 10% contamination were discarded.
d. Marker lineage was analyzed using AMPHORA2 and reported if 75% of the classifications were in agreement at a particular taxonomic level.
e. Open reading frames (ORFs) were predicted using FragGeneScan.
f. The listed bins were submitted under the MG-RAST project (ID: mgp9993).

**Table C.9** Gene inventory analysis related to N-substituted biomass structural detritus utilization.

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Haliscomenobacter hydrossis* (NC_015510) | fig\|760192.3.peg.1023 | Halhy_0946 CDS_1166632_1169307_- (YP_004445720.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.1033 | Halhy_0956 CDS_1176470_1177717_- (YP_004445730.1) | N-acetylglucosamine related transporter, NagX |
| | fig\|760192.3.peg.1058 | Halhy_0978 CDS_1204720_1207575_- (YP_004445750.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.1133 | Halhy_1044 CDS_1278061_1280481_- (YP_004445816.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.1134 | Halhy_1045 CDS_1281067_1283394_- (YP_004445817.1) | beta-glucosidase (EC 3.2.1.21) |
| | fig\|760192.3.peg.1212 | Halhy_1118 CDS_1378600_1381563_+ (YP_004445889.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.1289 | Halhy_1187 CDS_1480711_1481445_+ (YP_004445957.1) | Glucosamine-6-phosphate deaminase (EC 3.5.99.6) |
| | fig\|760192.3.peg.1465 | Halhy_1353 CDS_1686331_1688622_+ (YP_004446121.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.1496 | Halhy_1383 CDS_1738919_1741129_+ (YP_004446151.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.1552 | Halhy_1436 CDS_1798408_1801230_+ (YP_004446203.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.1593 | Halhy_1472 CDS_1852891_1856088_+ (YP_004446239.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.1642 | Halhy_1513 CDS_1907633_1910611_+ (YP_004446278.1) | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Haliscomenobacter hydrossis* (NC_015510) | fig\|760192.3.peg.1665 | Halhy_1539 CDS_1944077_1944943_+ (YP_004446304.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|760192.3.peg.1729 | Halhy_1597 CDS_2010973_2013093_+ (YP_004446361.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|760192.3.peg.1767 | Halhy_1633 CDS_2054708_2056765_+ (YP_004446396.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.1848 | Halhy_1706 CDS_2144811_2147420_- (YP_004446468.1) | Beta-mannosidase (EC 3.2.1.25) |
| | fig\|760192.3.peg.1886 | Halhy_1742 CDS_2185567_2187918_- (YP_004446503.1) | TonB-dependent siderophore receptor |
| | fig\|760192.3.peg.1900 | Halhy_1755 CDS_2201227_2203455_- (YP_004446516.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.1957 | Halhy_1809 CDS_2261545_2264412_- (YP_004446568.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.1959 | Halhy_1819 CDS_2265706_2268000_- (YP_004446519.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.1960 | Halhy_1820 CDS_2268043_2268228_- (YP_004446520.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.1965 | Halhy_1816 CDS_2274491_2277526_- (YP_004446574.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.1977 | Halhy_1827 CDS_2287667_2290417_+ (YP_004446585.1) | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Haliscomenobacter hydrossis* (NC_015510) | fig\|760192.3.peg.1989 | Halhy_1838 CDS_2317816_2319801_+ (YP_004446596.1) | TonB-dependent receptor plug |
| | fig\|760192.3.peg.2097 | Halhy_1939 CDS_2441702_2444239_- (YP_004446697.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.2218 | Halhy_2055 CDS_2607019_2607414_- (YP_004446813.1) | peptidoglycan-binding lysin domain-containing protein |
| | fig\|760192.3.peg.2254 | Halhy_2087 CDS_2649328_2651532_- (YP_004446845.1) | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|760192.3.peg.2315 | Halhy_2144 CDS_2775542_2778781_+ (YP_004446902.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.2467 | Halhy_2286 CDS_2962182_2964080_+ (YP_004447038.1) | 1,4-alpha-glucan (glycogen) branching enzyme, GH-13-type (EC 2.4.1.18) |
| | fig\|760192.3.peg.2538 | Halhy_2353 CDS_3040668_3042125_- (YP_004447102.1) | Transcriptional regulator, GntR family domain / Aspartate aminotransferase (EC 2.6.1.1) |
| | fig\|760192.3.peg.2563 | Halhy_2378 CDS_3068670_3071645_+ (YP_004447127.1) | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|760192.3.peg.2581 | Halhy_2395 CDS_3092399_3093715_+ (YP_004447143.1) | N-acetylglucosamine related transporter, NagX |
| | fig\|760192.3.peg.26 | Halhy_0027 CDS_25892_28654_- (YP_004444814.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.2776 | Halhy_2578 CDS_3294435_3296876_- (YP_004447321.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.2818 | Halhy_2615 CDS_3339688_3340860_+ (YP_004447357.1) | N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Haliscomenobacter hydrossis* (NC_015510) | fig\|760192.3.peg.2828 | Halhy_2625 CDS_3351328_3354237_- (YP_004447367.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.2834 | Halhy_2631 CDS_3360307_3362145_- (YP_004447373.1) | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |
| | fig\|760192.3.peg.2853 | Halhy_2649 CDS_3394286_3395506_- (YP_004447391.1) | Anhydro-N-acetylmuramic acid kinase (EC 2.7.1.170) |
| | fig\|760192.3.peg.2865 | Halhy_2661 CDS_3408415_3409761_+ (YP_004447403.1) | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|760192.3.peg.315 | Halhy_0307 CDS_377670_378965_- (YP_004445092.1) | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|760192.3.peg.3260 | Halhy_3045 CDS_3864273_3865313_- (YP_004447781.1) | Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13) |
| | fig\|760192.3.peg.3297 | Halhy_3084 CDS_3903942_3905336_+ (YP_004447820.1) | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|760192.3.peg.3343 | Halhy_3128 CDS_3951630_3955061_+ (YP_004447863.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.3449 | Halhy_3235 CDS_4083931_4084224_- (YP_004447968.1) | TonB family protein |
| | fig\|760192.3.peg.3506 | Halhy_3293 CDS_4151350_4152246_- (YP_004448025.1) | N-acetylneuraminate lyase (EC 4.1.3.3) |
| | fig\|760192.3.peg.3509 | Halhy_3295 CDS_4153900_4157217_- (YP_004448027.1) | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Haliscomenobacter hydrossis* (NC_015510) | fig\|760192.3.peg.354 | Halhy_0342 CDS_433938_437171_+ (YP_004445127.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.3733 | Halhy_3512 CDS_4417333_4420569_- (YP_004448240.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.3794 | Halhy_3571 CDS_4488951_4492403_+ (YP_004448298.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.3884 | Halhy_3658 CDS_4609360_4612752_- (YP_004448383.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.3948 | Halhy_3721 CDS_4689540_4690634_- (YP_004448446.1) | Transcriptional regulator, LacI family |
| | fig\|760192.3.peg.3949 | Halhy_3722 CDS_4690985_4693975_+ (YP_004448447.1) | TonB family protein / TonB-dependent receptor |
| | fig\|760192.3.peg.3950 | Halhy_3723 CDS_4693988_4695487_+ (YP_004448448.1) | RagB/SusD domain-containing protein |
| | fig\|760192.3.peg.3954 | Halhy_3727 CDS_4702060_4704096_+ (YP_004448452.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|760192.3.peg.3970 | Halhy_3743 CDS_4726420_4727325_- (YP_004448468.1) | ROK family sugar kinase or transcriptional regulator |
| | fig\|760192.3.peg.406 | Halhy_0388 CDS_493981_496365_+ (YP_004445172.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.408 | Halhy_0390 CDS_499750_501342_+ (YP_004445174.1) | SusD, outer membrane protein |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Haliscomenobacter hydrossis* (NC_015510) | fig\|760192.3.peg.4108 | Halhy_3874 CDS_4895183_4897519_+ (YP_004448598.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.4165 | Halhy_3927 CDS_4973819_4976314_+ (YP_004448648.1) | beta-glucosidase (EC 3.2.1.21) |
| | fig\|760192.3.peg.4302 | Halhy_4064 CDS_5148897_5151257_- (YP_004448785.1) | ABC transporter, permease protein |
| | fig\|760192.3.peg.4340 | Halhy_4100 CDS_5185215_5187605_- (YP_004448821.1) | ABC transporter, permease protein |
| | fig\|760192.3.peg.4343 | Halhy_4103 CDS_5189959_5192340_- (YP_004448823.1) | ABC transporter, permease protein |
| | fig\|760192.3.peg.4349 | Halhy_4110 CDS_5195442_5197880_- (YP_004448829.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.4418 | Halhy_4177 CDS_5274650_5277139_+ (YP_004448896.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.4549 | Halhy_4304 CDS_5416504_5417517_+ (YP_004449021.1) | Transcriptional regulator, LacI family |
| | fig\|760192.3.peg.4553 | Halhy_4308 CDS_5421555_5422463_- (YP_004449025.1) | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|760192.3.peg.4606 | Halhy_4356 CDS_5485251_5486033_- (YP_004449073.1) | N-acetylmuramic acid 6-phosphate etherase (EC 4.2.1.126) |
| | fig\|760192.3.peg.462 | Halhy_0443 CDS_566774_568831_+ (YP_004445227.1) | Beta-hexosaminidase (EC 3.2.1.52) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Haliscomenobacter hydrossis* (NC_015510) | fig\|760192.3.peg.4628 | Halhy_4375 CDS_5518070_5519110_- (YP_004449092.1) | Transcriptional regulator, LacI family |
| | fig\|760192.3.peg.4629 | Halhy_4376 CDS_5519378_5522680_+ (YP_004449093.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.4632 | Halhy_4379 CDS_5526554_5528164_+ (YP_004449096.1) | Beta-xylosidase (EC 3.2.1.37) |
| | fig\|760192.3.peg.4690 | Halhy_4439 CDS_5608743_5611940_- (YP_004449154.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.4726 | Halhy_4473 CDS_5659154_5660263_+ (YP_004449187.1) | N-acetylglucosamine related transporter, NagX |
| | fig\|760192.3.peg.4812 | Halhy_4557 CDS_5756371_5759478_+ (YP_004449270.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.4815 | Halhy_4537 CDS_5762400_5763131_+ (YP_004449237.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.4816 | Halhy_4538 CDS_5763082_5765511_+ (YP_004449238.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.4820 | Halhy_4563 CDS_5768102_5771152_+ (YP_004449275.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.4821 | Halhy_4564 CDS_5771192_5772544_+ (YP_004449276.1) | RagB/SusD domain-containing protein |
| | fig\|760192.3.peg.4887 | Halhy_4627 CDS_5852532_5855558_+ (YP_004449336.1) | SusC, outer membrane protein involved in starch binding |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Haliscomenobacter hydrossis* (NC_015510) | fig\|760192.3.peg.4888 | Halhy_4628 CDS_5855572_5857257_+ (YP_004449337.1) | SusD, outer membrane protein |
| | fig\|760192.3.peg.4898 | Halhy_4636 CDS_5867463_5868827_- (YP_004449345.1) | beta-glucosidase (EC 3.2.1.21) |
| | fig\|760192.3.peg.4926 | Halhy_4665 CDS_5903035_5906115_- (YP_004449373.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.4972 | Halhy_4709 CDS_5964404_5965795_- (YP_004449416.1) | N-acetylglucosamine deacetylase (EC 3.5.1.-) / 3-hydroxyacyl-[acyl-carrier-protein] dehydratase, FabZ form (EC 4.2.1.59) |
| | fig\|760192.3.peg.5071 | Halhy_4804 CDS_6066985_6069942_- (YP_004449510.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.5073 | Halhy_4807 CDS_6071732_6074854_- (YP_004449513.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.5088 | Halhy_4821 CDS_6089501_6092962_+ (YP_004449527.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.5115 | Halhy_4847 CDS_6129987_6130994_+ (YP_004449552.1) | Transcriptional regulator, LacI family |
| | fig\|760192.3.peg.5178 | Halhy_4898 CDS_6219711_6223118_- (YP_004449600.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|760192.3.peg.5181 | Halhy_4901 CDS_6226482_6228047_- (YP_004449603.1) | SusD, outer membrane protein |
| | fig\|760192.3.peg.5182 | Halhy_4902 CDS_6228071_6231016_- (YP_004449604.1) | SusC, outer membrane protein involved in starch binding |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Haliscomenobacter hydrossis* (NC_015510) | fig\|760192.3.peg.5303 | Halhy_5020 CDS_6369366_6370739_- (YP_004449721.1) | RagB/SusD domain-containing protein |
| | fig\|760192.3.peg.5304 | Halhy_5021 CDS_6370758_6373727_- (YP_004449722.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.5358 | Halhy_5071 CDS_6439605_6441962_- (YP_004449770.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.5396 | Halhy_5108 CDS_6489645_6492017_- (YP_004449807.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.542 | Halhy_0488 CDS_629171_630199_+ (YP_004445271.1) | Transcriptional regulator, LacI family |
| | fig\|760192.3.peg.5425 | Halhy_5134 CDS_6517834_6521067_- (YP_004449833.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.545 | Halhy_0491 CDS_634311_635978_+ (YP_004445274.1) | Beta-xylosidase (EC 3.2.1.37) |
| | fig\|760192.3.peg.549 | Halhy_0494 CDS_641209_642360_+ (YP_004445277.1) | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|760192.3.peg.551 | Halhy_0496 CDS_643339_644415_+ (YP_004445279.1) | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|760192.3.peg.553 | Halhy_0497 CDS_644673_646598_+ (YP_004445280.1) | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|760192.3.peg.5552 | Halhy_5252 CDS_6687891_6689564_- (YP_004449951.1) | Beta-xylosidase (EC 3.2.1.37) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Haliscomenobacter hydrossis* (NC_015510) | fig\|760192.3.peg.5555 | Halhy_5255 CDS_6691694_6692899_- (YP_004449953.1) | Chitinase (EC 3.2.1.14) |
| | fig\|760192.3.peg.5579 | Halhy_5280 CDS_6723818_6727108_- (YP_004449978.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.5607 | Halhy_5305 CDS_6754885_6757863_+ (YP_004450003.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.5610 | Halhy_5308 CDS_6761094_6762257_+ (YP_004450006.1) | N-acylglucosamine 2-epimerase (EC 5.1.3.8) |
| | fig\|760192.3.peg.5692 | Halhy_5388 CDS_6860833_6861264_- (YP_004450086.1) | TonB family protein |
| | fig\|760192.3.peg.5708 | Halhy_5404 CDS_6886087_6889209_- (YP_004450102.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|760192.3.peg.5722 | Halhy_5419 CDS_6905991_6909203_- (YP_004450117.1) | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|760192.3.peg.5776 | Halhy_5474 CDS_6969873_6973112_- (YP_004450172.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.5863 | Halhy_5557 CDS_7087100_7090264_+ (YP_004450254.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.5896 | Halhy_5587 CDS_7130831_7133218_+ (YP_004450284.1) | TonB-dependent siderophore receptor |
| | fig\|760192.3.peg.6003 | Halhy_5689 CDS_7251939_7254377_+ (YP_004450385.1) | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Haliscomenobacter hydrossis* (NC_015510) | fig\|760192.3.peg.6010 | Halhy_5695 CDS_7264146_7267073_+ (YP_004450391.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.6042 | Halhy_5727 CDS_7315830_7319018_+ (YP_004450423.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.607 | Halhy_0556 CDS_693473_694411_+ (YP_004445339.1) | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|760192.3.peg.6124 | Halhy_5803 CDS_7431751_7434561_- (YP_004450499.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.6127 | Halhy_5806 CDS_7436355_7437734_+ (YP_004450502.1) | Phosphomannomutase (EC 5.4.2.8) / Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|760192.3.peg.6157 | Halhy_5837 CDS_7479923_7483069_+ (YP_004450533.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.6161 | Halhy_5841 CDS_7492360_7493403_+ (YP_004450537.1) | Transcriptional regulator, LacI family |
| | fig\|760192.3.peg.6191 | Halhy_5871 CDS_7542221_7545163_- (YP_004450567.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.6192 | Halhy_5872 CDS_7545452_7547287_- (YP_004450568.1) | RagB/SusD domain-containing protein |
| | fig\|760192.3.peg.6193 | Halhy_5873 CDS_7547321_7550467_- (YP_004450569.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.6305 | Halhy_5983 CDS_7644137_7645636_+ (YP_004450679.1) | Beta-hexosaminidase (EC 3.2.1.52) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Haliscomenobacter hydrossis* (NC_015510) | fig\|760192.3.peg.6445 | Halhy_6113 CDS_7807143_7810139_+ (YP_004450808.1) | TonB-dependent receptor plug |
| | fig\|760192.3.peg.6518 | Halhy_4544 CDS_7890942_7891583_- (YP_004450844.1) | ABC transporter, permease protein |
| | fig\|760192.3.peg.656 | Halhy_0602 CDS_749944_751011_- (YP_004445384.1) | L-alanine-DL-glutamate epimerase (EC 5.1.1.n1) |
| | fig\|760192.3.peg.6564 | Halhy_6222 CDS_7942864_7945611_+ (YP_004450915.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.6601 | Halhy_6252 CDS_7982182_7983345_+ (YP_004450945.1) | N-acetylglucosamine related transporter, NagX |
| | fig\|760192.3.peg.6813 | Halhy_6457 CDS_8245857_8246969_- (YP_004451147.1) | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|760192.3.peg.6857 | Halhy_6499 CDS_8302128_8303399_+ (YP_004451188.1) | N-acetyl glucosamine transporter, NagP |
| | fig\|760192.3.peg.6883 | Halhy_6528 CDS_8336343_8338724_+ (YP_004451217.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.6885 | Halhy_6530 CDS_8339641_8340393_- (YP_004451219.1) | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|760192.3.peg.6903 | Halhy_6546 CDS_8364784_8367066_- (YP_004451235.1) | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|760192.3.peg.6904 | Halhy_6547 CDS_8367204_8369123_- (YP_004451236.1) | Glucosamine-6-phosphate deaminase (EC 3.5.99.6) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Haliscomenobacter hydrossis* (NC_015510) | fig\|760192.3.peg.6906 | Halhy_6549 CDS_8370365_8371435_+ (YP_004451238.1) | Transcriptional regulator, LacI family |
| | fig\|760192.3.peg.6965 | Halhy_6608 CDS_71656_73884_+ (YP_004451297.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.6966 | Halhy_6609 CDS_74405_76576_+ (YP_004451298.1) | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|760192.3.peg.6967 | Halhy_4547 CDS_76727_76921_+ (YP_004450847.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.712 | Halhy_0652 CDS_806997_810143_- (YP_004445433.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.783 | Halhy_0721 CDS_897452_900592_+ (YP_004445502.1) | TonB-dependent receptor |
| | fig\|760192.3.peg.793 | Halhy_0730 CDS_911316_912437_- (YP_004445511.1) | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|760192.3.peg.802 | Halhy_0738 CDS_921624_924635_- (YP_004445519.1) | TonB family protein / TonB-dependent receptor |
| | fig\|760192.3.peg.815 | Halhy_0750 CDS_947502_950315_+ (YP_004445530.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|760192.3.peg.976 | Halhy_0904 CDS_1113693_1116101_- (YP_004445678.1) | TonB-dependent receptor |
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.100 | Cpin_0106 CDS_128386_130833_+ (YP_003119810.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|485918.6.peg.1022 | Cpin_1034 CDS_1240012_1242792_+ (YP_003120733.1) | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.1085 | Cpin_1097 CDS_1315270_1318653_+ (YP_003120796.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.115 | Cpin_0121 CDS_148670_150844_- (YP_003119825.1) | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|485918.6.peg.1187 | Cpin_1197 CDS_1424967_1426091_+ (YP_003120896.1) | Anhydro-N-acetylmuramic acid kinase (EC 2.7.1.170) |
| | fig\|485918.6.peg.1269 | Cpin_1279 CDS_1529335_1530612_- (YP_003120978.1) | N-acetylglucosamine deacetylase (EC 3.5.1.-) / 3-hydroxyacyl-[acyl-carrier-protein] dehydratase, FabZ form (EC 4.2.1.59) |
| | fig\|485918.6.peg.129 | Cpin_0134 CDS_162103_164793_- (YP_003119838.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|485918.6.peg.131 | Cpin_0136 CDS_166355_169681_- (YP_003119840.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.1324 | Cpin_1334 CDS_1590515_1591870_+ (YP_003121033.1) | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|485918.6.peg.1343 | Cpin_1354 CDS_1618219_1621845_- (YP_003121053.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.1449 | Cpin_1458 CDS_1786038_1786865_+ (YP_003121156.1) | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|485918.6.peg.1451 | Cpin_1460 CDS_1788040_1791237_+ (YP_003121158.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.1454 | Cpin_1463 CDS_1793733_1794857_- (YP_003121161.1) | N-acetylglucosamine related transporter, NagX |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.1459 | Cpin_1468 CDS_1800044_1802092_- (YP_003121166.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.1479 | Cpin_1487 CDS_1825077_1827677_+ (YP_003121185.1) | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|485918.6.peg.1486 | Cpin_1494 CDS_1835420_1838515_+ (YP_003121192.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.1544 | Cpin_1552 CDS_1894702_1897788_- (YP_003121249.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.1581 | Cpin_1589 CDS_1932118_1935525_+ (YP_003121286.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.167 | Cpin_0171 CDS_211125_214358_- (YP_003119873.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.1706 | Cpin_1716 CDS_2064274_2067429_+ (YP_003121413.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.172 | Cpin_0176 CDS_221066_223768_- (YP_003119878.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.1725 | Cpin_1734 CDS_2088081_2091332_+ (YP_003121431.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.1741 | Cpin_1750 CDS_2119885_2122224_- (YP_003121447.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|485918.6.peg.1771 | Cpin_1781 CDS_2159362_2162496_+ (YP_003121478.1) | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.1787 | Cpin_1798 CDS_2181816_2183672_+ (YP_003121495.1) | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|485918.6.peg.1790 | Cpin_1800 CDS_2186845_2190021_+ (YP_003121497.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.1800 | Cpin_1810 CDS_2210312_2212696_+ (YP_003121507.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|485918.6.peg.1827 | Cpin_1838 CDS_2242611_2243966_+ (YP_003121535.1) | beta-glucosidase (EC 3.2.1.21) |
| | fig\|485918.6.peg.1835 | Cpin_1847 CDS_2252285_2253667_- (YP_003121544.1) | Phosphomannomutase (EC 5.4.2.8) / Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|485918.6.peg.1874 | Cpin_1886 CDS_2294751_2297756_- (YP_003121583.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.1876 | Cpin_1888 CDS_2299807_2303115_- (YP_003121585.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.1881 | Cpin_1893 CDS_2307980_2311045_+ (YP_003121590.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.1886 | Cpin_1897 CDS_2315851_2318877_+ (YP_003121594.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.1906 | Cpin_1915 CDS_2337717_2340047_- (YP_003121612.1) | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|485918.6.peg.1991 | Cpin_2000 CDS_2438688_2440523_- (YP_003121696.1) | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.20 | Cpin_0020 CDS_23772_25049_- (YP_003119730.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.201 | Cpin_0204 CDS_253225_255261_+ (YP_003119906.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2052 | Cpin_2062 CDS_2512408_2514783_+ (YP_003121757.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2069 | Cpin_2080 CDS_2543208_2546522_+ (YP_003121774.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2087 | Cpin_2097 CDS_2568247_2571543_+ (YP_003121791.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2100 | Cpin_2109 CDS_2583170_2586592_+ (YP_003121803.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2179 | Cpin_2184 CDS_2681099_2685187_+ (YP_003121877.1) | Chitinase (EC 3.2.1.14) |
| | fig\|485918.6.peg.2181 | Cpin_2186 CDS_2687008_2688618_+ (YP_003121879.1) | Chitinase (EC 3.2.1.14) |
| | fig\|485918.6.peg.2182 | Cpin_2187 CDS_2688726_2689868_+ (YP_003121880.1) | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|485918.6.peg.2187 | Cpin_2191 CDS_2694371_2698030_+ (YP_003121884.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.22 | Cpin_0022 CDS_26755_28398_+ (YP_003119732.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|485918.6.peg.2242 | Cpin_2250 CDS_2768414_2770930_- (YP_003121942.1) | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.2243 | Cpin_2251 CDS_2771330_2773588_- (YP_003121943.1) | beta-hexosaminidase precursor |
| | fig\|485918.6.peg.2257 | Cpin_2264 CDS_2790265_2793561_+ (YP_003121956.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2267 | Cpin_2275 CDS_2803635_2806184_+ (YP_003121967.1) | Beta-mannosidase (EC 3.2.1.25) |
| | fig\|485918.6.peg.231 | Cpin_0236 CDS_283325_286144_+ (YP_003119937.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2312 | Cpin_2320 CDS_2854274_2857090_+ (YP_003122012.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2415 | Cpin_2420 CDS_2976039_2976770_- (YP_003122112.1) | Predicted transcriptional regulator of N-Acetylglucosamine utilization, GntR family |
| | fig\|485918.6.peg.2431 | Cpin_2435 CDS_2994779_2998306_- (YP_003122127.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2439 | Cpin_2445 CDS_3004953_3008237_+ (YP_003122135.1) | SusC, outer membrane protein involved in starch binding |
| | fig\|485918.6.peg.2440 | Cpin_2446 CDS_3008272_3009888_+ (YP_003122136.1) | SusD, outer membrane protein |
| | fig\|485918.6.peg.2461 | Cpin_2469 CDS_3041523_3044660_+ (YP_003122158.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|485918.6.peg.2477 | Cpin_2482 CDS_3056062_3057093_+ (YP_003122171.1) | Endo-1,4-beta-xylanase D |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.2501 | Cpin_2507 CDS_3083086_3084372_+ (YP_003122196.1) | Membrane-bound lytic murein transglycosylase |
| | fig\|485918.6.peg.2562 | Cpin_2580 CDS_3156584_3159859_+ (YP_003122264.1) | Chitinase (EC 3.2.1.14) |
| | fig\|485918.6.peg.2587 | Cpin_2605 CDS_3191978_3194224_- (YP_003122289.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|485918.6.peg.2600 | CDS_3206873_3210685_- | TonB-dependent receptor |
| | fig\|485918.6.peg.2652 | Cpin_2671 CDS_3264412_3267942_- (YP_003122354.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.268 | Cpin_0274 CDS_323190_325565_- (YP_003119974.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2699 | Cpin_2720 CDS_3320046_3323207_- (YP_003122401.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|485918.6.peg.2701 | Cpin_2722 CDS_3324994_3327786_- (YP_003122403.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2714 | Cpin_2734 CDS_3344761_3347466_+ (YP_003122414.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|485918.6.peg.2719 | Cpin_2739 CDS_3355658_3358804_+ (YP_003122419.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2747 | Cpin_2768 CDS_3393365_3396568_+ (YP_003122448.1) | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.2800 | Cpin_2820 CDS_3473916_3477023_+ (YP_003122500.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2824 | Cpin_2845 CDS_3507646_3510723_+ (YP_003122525.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2836 | Cpin_2857 CDS_3530707_3534153_+ (YP_003122537.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2840 | Cpin_2861 CDS_3538706_3541660_+ (YP_003122541.1) | Beta-mannosidase (EC 3.2.1.25) |
| | fig\|485918.6.peg.2841 | Cpin_2862 CDS_3541687_3543930_+ (YP_003122542.1) | beta-glucosidase (EC 3.2.1.21) |
| | fig\|485918.6.peg.2845 | Cpin_2866 CDS_3546304_3547767_+ (YP_003122546.1) | endo-1,4-beta-xylanase |
| | fig\|485918.6.peg.2860 | Cpin_2881 CDS_3559782_3563399_- (YP_003122561.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.2929 | Cpin_2947 CDS_3625558_3628947_- (YP_003122626.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3021 | Cpin_3048 CDS_3759849_3763430_+ (YP_003122722.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3035 | CDS_3778896_3779711_- () | TonB-dependent receptor |
| | fig\|485918.6.peg.3068 | Cpin_3094 CDS_3811277_3813217_+ (YP_003122764.1) | 1,4-alpha-glucan (glycogen) branching enzyme, GH-13-type (EC 2.4.1.18) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.3095 | Cpin_3123 CDS_3842146_3845490_+ (YP_003122792.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3120 | Cpin_3148 CDS_3880078_3883437_- (YP_003122816.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3132 | Cpin_3164 CDS_3898609_3901881_+ (YP_003122832.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3157 | Cpin_3194 CDS_3938954_3941671_- (YP_003122862.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|485918.6.peg.3159 | Cpin_3196 CDS_3943164_3946463_- (YP_003122864.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3167 | Cpin_3206 CDS_3955928_3959242_+ (YP_003122874.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3175 | Cpin_3214 CDS_3965741_3969076_- (YP_003122882.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.318 | Cpin_0323 CDS_371461_373884_+ (YP_003120023.1) | beta-glucosidase (EC 3.2.1.21) |
| | fig\|485918.6.peg.3248 | Cpin_3288 CDS_4048448_4051735_+ (YP_003122956.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3305 | Cpin_3346 CDS_4113356_4113778_- (YP_003123014.1) | Peptidoglycan-binding LysM |
| | fig\|485918.6.peg.335 | Cpin_0339 CDS_402104_405235_- (YP_003120039.1) | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.3485 | Cpin_3538 CDS_4405174_4408113_+ (YP_003123203.1) | TonB family protein / TonB-dependent receptor |
| | fig\|485918.6.peg.3501 | Cpin_3553 CDS_4427541_4430771_- (YP_003123218.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.351 | Cpin_0356 CDS_423682_427068_- (YP_003120056.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3548 | Cpin_3601 CDS_4490895_4494167_+ (YP_003123265.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3579 | Cpin_3633 CDS_4559334_4562852_+ (YP_003123296.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3589 | Cpin_3643 CDS_4573815_4577075_+ (YP_003123306.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.365 | Cpin_0369 CDS_442341_445565_+ (YP_003120069.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3670 | Cpin_3725 CDS_4679786_4682542_- (YP_003123388.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3694 | Cpin_3749 CDS_4713600_4717007_- (YP_003123412.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3737 | Cpin_3794 CDS_4770439_4772151_- (YP_003123457.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|485918.6.peg.3751 | Cpin_3808 CDS_4789144_4792407_+ (YP_003123471.1) | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.38 | Cpin_0043 CDS_50784_54053_- (YP_003119750.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3853 | Cpin_3941 CDS_4898128_4900140_+ (YP_003123604.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3872 | Cpin_3959 CDS_4921603_4924737_+ (YP_003123622.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.3906 | Cpin_3992 CDS_4966532_4967188_+ (YP_003123655.1) | Beta-phosphoglucomutase (EC 5.4.2.6) |
| | fig\|485918.6.peg.3956 | CDS_5032882_5034006_+ () | TonB-dependent receptor |
| | fig\|485918.6.peg.4 | Cpin_0004 CDS_3973_7302_+ (YP_003119714.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.400 | Cpin_0405 CDS_485451_488486_+ (YP_003120104.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4045 | Cpin_4133 CDS_5151177_5153729_- (YP_003123793.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4057 | Cpin_4146 CDS_5170759_5173644_- (YP_003123806.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4084 | Cpin_4173 CDS_5215073_5216122_- (YP_003123831.1) | N-acetylglucosamine related transporter, NagX |
| | fig\|485918.6.peg.4087 | Cpin_4176 CDS_5218676_5221936_- (YP_003123834.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4134 | CDS_5292372_5292503_- () | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.4142 | Cpin_4240 CDS_5301737_5302675_- (YP_003123895.1) | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|485918.6.peg.4144 | Cpin_4242 CDS_5304850_5306766_+ (YP_003123897.1) | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|485918.6.peg.4272 | Cpin_4371 CDS_5436435_5439059_- (YP_003124019.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4293 | Cpin_4389 CDS_5459171_5461120_+ (YP_003124036.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4318 | Cpin_4414 CDS_5501509_5503404_+ (YP_003124061.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4320 | Cpin_4416 CDS_5505619_5508477_+ (YP_003124063.1) | SusC, outer membrane protein involved in starch binding |
| | fig\|485918.6.peg.4402 | Cpin_4497 CDS_5596822_5600343_- (YP_003124140.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4406 | Cpin_4501 CDS_5603425_5604375_- (YP_003124144.1) | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|485918.6.peg.4409 | Cpin_4504 CDS_5606957_5610145_- (YP_003124147.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4435 | Cpin_4532 CDS_5650790_5653816_- (YP_003124175.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4448 | Cpin_4545 CDS_5676081_5678399_- (YP_003124188.1) | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.4451 | Cpin_4548 CDS_5680821_5684012_+ (YP_003124191.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4459 | Cpin_4556 CDS_5693602_5695434_- (YP_003124199.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|485918.6.peg.4462 | Cpin_4559 CDS_5698617_5701607_- (YP_003124202.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.450 | Cpin_0454 CDS_539509_542286_- (YP_003120153.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4512 | Cpin_4609 CDS_5750648_5753917_+ (YP_003124251.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.452 | Cpin_0456 CDS_544497_547229_- (YP_003120155.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.454 | Cpin_0458 CDS_549250_552465_- (YP_003120157.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4540 | Cpin_4637 CDS_5787682_5790867_- (YP_003124279.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4559 | Cpin_4660 CDS_5813872_5816556_+ (YP_003124299.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4706 | Cpin_4811 CDS_5999523_6002438_- (YP_003124446.1) | Beta-mannosidase (EC 3.2.1.25) |
| | fig\|485918.6.peg.4707 | Cpin_4812 CDS_6002468_6005524_- (YP_003124447.1) | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.4709 | Cpin_4814 CDS_6007288_6010365_- (YP_003124449.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4714 | Cpin_4819 CDS_6015230_6018214_- (YP_003124454.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4719 | Cpin_4824 CDS_6026218_6029472_- (YP_003124459.1) | TonB family protein / TonB-dependent receptor |
| | fig\|485918.6.peg.4745 | Cpin_4849 CDS_6061233_6064391_- (YP_003124484.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4872 | Cpin_4975 CDS_6221890_6225057_- (YP_003124609.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.4893 | Cpin_4994 CDS_6243620_6245917_- (YP_003124628.1) | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|485918.6.peg.4969 | Cpin_5069 CDS_6313877_6314995_- (YP_003124702.1) | N-acetylglucosamine related transporter, NagX |
| | fig\|485918.6.peg.4988 | Cpin_5088 CDS_6334805_6335464_- (YP_003124721.1) | Beta-phosphoglucomutase (EC 5.4.2.6) |
| | fig\|485918.6.peg.4991 | Cpin_5091 CDS_6338502_6340094_- (YP_003124724.1) | SusD, outer membrane protein |
| | fig\|485918.6.peg.4992 | Cpin_5092 CDS_6340114_6343080_- (YP_003124725.1) | SusC, outer membrane protein involved in starch binding |
| | fig\|485918.6.peg.5011 | Cpin_5109 CDS_6370913_6372142_- (YP_003124742.1) | beta-glucanase |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.5015 | Cpin_5113 CDS_6377593_6380820_+ (YP_003124746.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.5046 | Cpin_5147 CDS_6423207_6426461_- (YP_003124780.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.5049 | Cpin_5150 CDS_6429054_6430754_+ (YP_003124783.1) | Beta-xylosidase (EC 3.2.1.37) |
| | fig\|485918.6.peg.5099 | Cpin_5202 CDS_6575270_6577006_- (YP_003124834.1) | SusD, outer membrane protein |
| | fig\|485918.6.peg.51 | Cpin_0057 CDS_68586_71567_- (YP_003119764.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.5100 | Cpin_5203 CDS_6577045_6580038_- (YP_003124835.1) | SusC, outer membrane protein involved in starch binding |
| | fig\|485918.6.peg.5159 | Cpin_5260 CDS_6658435_6660222_- (YP_003124892.1) | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|485918.6.peg.5179 | Cpin_5278 CDS_6691494_6694985_- (YP_003124910.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.5215 | Cpin_5315 CDS_6786556_6789636_- (YP_003124947.1) | TonB family protein / TonB-dependent receptor |
| | fig\|485918.6.peg.5243 | Cpin_5342 CDS_6821994_6825551_+ (YP_003124974.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.5298 | Cpin_5396 CDS_6900445_6903516_- (YP_003125027.1) | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.5301 | Cpin_5399 CDS_6906424_6907452_+ (YP_003125030.1) | N-acylglucosamine 2-epimerase (EC 5.1.3.8) |
| | fig\|485918.6.peg.5313 | Cpin_5411 CDS_6918112_6919917_- (YP_003125041.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|485918.6.peg.5392 | Cpin_5486 CDS_6995550_6998504_+ (YP_003125116.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.5415 | Cpin_5508 CDS_7027184_7030144_- (YP_003125138.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.5417 | Cpin_5511 CDS_7031812_7033422_- (YP_003125141.1) | Beta-xylosidase (EC 3.2.1.37) |
| | fig\|485918.6.peg.5428 | Cpin_5521 CDS_7043539_7046916_- (YP_003125151.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.5439 | Cpin_5532 CDS_7058148_7061426_- (YP_003125162.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.5457 | Cpin_5549 CDS_7082902_7086012_+ (YP_003125177.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.5470 | Cpin_5562 CDS_7098555_7101725_- (YP_003125190.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.55 | Cpin_0061 CDS_74551_77295_+ (YP_003119768.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|485918.6.peg.5558 | Cpin_5648 CDS_7204082_7207153_- (YP_003125273.1) | TonB family protein / TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.5651 | Cpin_5745 CDS_7308074_7310089_- (YP_003125367.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.5740 | Cpin_5835 CDS_7413320_7416421_+ (YP_003125455.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.5829 | Cpin_5926 CDS_7520252_7523026_- (YP_003125545.1) | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|485918.6.peg.5948 | Cpin_6044 CDS_7642584_7644035_+ (YP_003125657.1) | beta-xylosidase (1,4-beta-D-xylan xylosidase) |
| | fig\|485918.6.peg.5992 | Cpin_6086 CDS_7686680_7687462_+ (YP_003125696.1) | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|485918.6.peg.6032 | Cpin_6129 CDS_7729648_7731750_- (YP_003125739.1) | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|485918.6.peg.6051 | Cpin_6148 CDS_7749641_7753297_- (YP_003125757.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6059 | Cpin_6155 CDS_7761074_7764571_- (YP_003125764.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6070 | Cpin_6166 CDS_7778621_7781227_+ (YP_003125775.1) | SusC, outer membrane protein involved in starch binding |
| | fig\|485918.6.peg.6071 | Cpin_6167 CDS_7781245_7782837_+ (YP_003125776.1) | SusD, outer membrane protein |
| | fig\|485918.6.peg.6075 | Cpin_6171 CDS_7787822_7790263_+ (YP_003125780.1) | Beta-galactosidase (EC 3.2.1.23) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.6077 | Cpin_6173 CDS_7791580_7793424_- (YP_003125782.1) | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|485918.6.peg.6095 | Cpin_6191 CDS_7809977_7811647_+ (YP_003125798.1) | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|485918.6.peg.6105 | Cpin_6201 CDS_7823193_7825553_+ (YP_003125808.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.626 | Cpin_0633 CDS_768696_771791_+ (YP_003120332.1) | TonB family protein / TonB-dependent receptor |
| | fig\|485918.6.peg.6264 | Cpin_6368 CDS_8004774_8007707_- (YP_003125973.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6305 | Cpin_6408 CDS_8051011_8053329_+ (YP_003126013.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6326 | Cpin_6427 CDS_8073659_8076826_+ (YP_003126032.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6348 | Cpin_6448 CDS_8101439_8104690_- (YP_003126053.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6366 | Cpin_6466 CDS_8123618_8127034_- (YP_003126071.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6369 | Cpin_6469 CDS_8128914_8129936_- (YP_003126074.1) | N-acetylglucosamine related transporter, NagX |
| | fig\|485918.6.peg.6438 | Cpin_6538 CDS_8202261_8203073_- (YP_003126143.1) | N-acetylmuramic acid 6-phosphate etherase (EC 4.2.1.126) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.6457 | Cpin_6560 CDS_8227138_8228454_- (YP_003126165.1) | N-acetyl glucosamine transporter, NagP |
| | fig\|485918.6.peg.6463 | Cpin_6566 CDS_8234829_8238242_- (YP_003126171.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6467 | Cpin_6570 CDS_8241368_8243293_+ (YP_003126175.1) | Glucosamine-6-phosphate deaminase (EC 3.5.99.6) / domain similar to N-acetylglucosaminyl-phosphatidylinositol de-N-acetylase |
| | fig\|485918.6.peg.6559 | Cpin_6660 CDS_8353629_8356856_- (YP_003126265.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6628 | Cpin_6729 CDS_8426033_8429014_- (YP_003126331.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6629 | Cpin_6730 CDS_8430021_8430992_- (YP_003126332.1) | endo-1,4-beta-xylanase D precursor |
| | fig\|485918.6.peg.6633 | Cpin_6733 CDS_8433470_8436766_+ (YP_003126335.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6637 | Cpin_6737 CDS_8441523_8442383_- (YP_003126339.1) | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|485918.6.peg.6640 | Cpin_6740 CDS_8445026_8448259_- (YP_003126342.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6693 | Cpin_6791 CDS_8502979_8506290_- (YP_003126393.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6698 | Cpin_6796 CDS_8510433_8513513_+ (YP_003126398.1) | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.6713 | Cpin_6810 CDS_8531595_8532788_- (YP_003126412.1) | Anhydro-N-acetylmuramic acid kinase (EC 2.7.1.170) |
| | fig\|485918.6.peg.6714 | Cpin_6811 CDS_8532794_8533594_- (YP_003126413.1) | N-acetylmuramic acid 6-phosphate etherase (EC 4.2.1.126) |
| | fig\|485918.6.peg.6764 | Cpin_6860 CDS_8592128_8593255_+ (YP_003126462.1) | Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13) |
| | fig\|485918.6.peg.6782 | Cpin_6880 CDS_8617332_8620577_- (YP_003126482.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6844 | Cpin_6944 CDS_8697892_8700105_- (YP_003126546.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6940 | Cpin_7043 CDS_8811225_8814794_- (YP_003126645.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6952 | Cpin_7055 CDS_8828577_8831147_- (YP_003126657.1) | Beta-mannosidase (EC 3.2.1.25) |
| | fig\|485918.6.peg.6960 | Cpin_7063 CDS_8838479_8839711_- (YP_003126665.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.6964 | Cpin_7067 CDS_8843633_8844544_+ (YP_003126669.1) | Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13) |
| | fig\|485918.6.peg.7001 | Cpin_7106 CDS_8892811_8895267_- (YP_003126708.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.7074 | Cpin_7176 CDS_8961881_8963437_- (YP_003126777.1) | SusD, outer membrane protein |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.7075 | Cpin_7177 CDS_8963534_8966449_- (YP_003126778.1) | SusC, outer membrane protein involved in starch binding |
| | fig\|485918.6.peg.7079 | Cpin_7182 CDS_8970572_8974132_- (YP_003126783.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.7091 | Cpin_7193 CDS_8987010_8988638_- (YP_003126794.1) | Beta-xylosidase (EC 3.2.1.37) |
| | fig\|485918.6.peg.7116 | Cpin_7217 CDS_9017316_9019376_- (YP_003126818.1) | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|485918.6.peg.7118 | Cpin_7219 CDS_9020826_9022889_- (YP_003126820.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.7126 | Cpin_7227 CDS_9029217_9031847_- (YP_003126828.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.7132 | Cpin_7233 CDS_9037371_9038987_- (YP_003126834.1) | SusD, outer membrane protein |
| | fig\|485918.6.peg.7133 | Cpin_7234 CDS_9039006_9042227_- (YP_003126835.1) | SusC, outer membrane protein involved in starch binding |
| | fig\|485918.6.peg.7152 | Cpin_7253 CDS_9063368_9066358_- (YP_003126854.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.7169 | Cpin_7271 CDS_9086893_9090147_- (YP_003126871.1) | TonB-dependent receptor |
| | fig\|485918.6.peg.786 | Cpin_0797 CDS_961963_962586_+ (YP_003120496.1) | Lyzozyme M1 (1,4-beta-N-acetylmuramidase) (EC 3.2.1.17) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Chitinophaga pinensis* (NC_013132) | fig\|485918.6.peg.802 | Cpin_0813 CDS_975944_977266_+ (YP_003120512.1) | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|485918.6.peg.828 | Cpin_0839 CDS_1004516_1007500_+ (YP_003120538.1) | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|485918.6.peg.964 | Cpin_0973 CDS_1169115_1172444_+ (YP_003120672.1) | TonB-dependent receptor |
| Bin87 | fig\|6666666.223310.peg.3286 | contig01408_23897_28609_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223310.peg.759 | contig00241_103420_108264_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223310.peg.4139 | contig02949_22631_19056_- | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223310.peg.4141 | contig02949_25130_24921_- | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223310.peg.962 | contig00292_59465_62485_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223310.peg.963 | contig00292_62521_66609_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223310.peg.984 | contig00292_97740_95728_- | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223310.peg.879 | contig00269_108583_114465_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223310.peg.1988 | contig00699_74190_72409_- | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223310.peg.4380 | contig04344_8198_13150_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223310.peg.1524 | contig00547_17752_16337_- | 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) |
| | fig\|6666666.223310.peg.3361 | contig01549_34522_35913_+ | 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) |
| | fig\|6666666.223310.peg.4302 | contig03827_6171_5860_- | 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) |
| | fig\|6666666.223310.peg.4303 | contig03827_6836_6210_- | 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) |
| | fig\|6666666.223310.peg.3939 | contig02555_2285_3382_+ | Anhydro-N-acetylmuramic acid kinase (EC2.7.1.) |
| | fig\|6666666.223310.peg.3231 | contig01338_13133_16018_+ | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.223310.peg.2906 | contig01262_4899_7859_+ | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.223310.peg.3232 | contig01338_16038_17588_+ | SusD, outer membrane protein |
| | fig\|6666666.223310.peg.2907 | contig01262_7881_10064_+ | SusD, outer membrane protein |
| | fig\|6666666.223310.peg.1879 | contig00689_37211_40111_+ | TonB-dependent receptor |
| | fig\|6666666.223310.peg.3733 | contig02209_22091_25078_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223310.peg.729 | contig00241_57555_60647_+ | TonB family protein / TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin87 | fig\|6666666.223310.peg.3692 | contig02051_9610_6437_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223310.peg.2583 | contig01043_56667_54271_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223310.peg.4412 | contig04451_5135_5545_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223310.peg.4802 | contig13779_5496_2335_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223310.peg.2046 | contig00726_52480_55695_+ | TonB-dependent outer membrane receptor |
| | fig\|6666666.223310.peg.2419 | contig00965_54249_56480_+ | TonB-dependent outer membrane receptor |
| | fig\|6666666.223310.peg.2824 | contig01234_25530_28358_+ | TonB-dependent receptor |
| | fig\|6666666.223310.peg.2327 | contig00921_4123_6333_+ | TonB-dependent receptor |
| | fig\|6666666.223310.peg.1217 | contig00389_100424_98139_- | TonB-dependent receptor |
| | fig\|6666666.223310.peg.3975 | contig02657_16864_19536_+ | TonB-dependent receptor |
| | fig\|6666666.223310.peg.2235 | contig00856_51413_49173_- | TonB-dependent receptor |
| | fig\|6666666.223310.peg.355 | contig00130_50993_53797_+ | TonB-dependent receptor |
| | fig\|6666666.223310.peg.4736 | contig09508_3611_6004_+ | TonB-dependent receptor |
| | fig\|6666666.223310.peg.266 | contig00041_321008_323506_+ | TonB-dependent receptor plug domain protein |
| | fig\|6666666.223310.peg.2004 | contig00726_7166_4818_- | TonB-dependent receptor, plug precursor |
| | fig\|6666666.223310.peg.72 | contig00041_83481_80860_- | TonB-dependent receptor, putative |
| | fig\|6666666.223310.peg.685 | contig00241_9893_7506_- | TonB-dependent receptor, putative |
| | fig\|6666666.223310.peg.1439 | contig00500_29553_31994_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223310.peg.2021 | contig00726_26722_24278_- | TonB-dependent receptor, putative |
| | fig\|6666666.223310.peg.1050 | contig00378_27468_25015_- | TonB-dependent receptor, putative |
| | fig\|6666666.223310.peg.935 | contig00292_27812_30358_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223310.peg.370 | contig00130_72655_70250_- | TonB-dependent receptor, putative |
| | fig\|6666666.223310.peg.454 | contig00130_184305_186740_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223310.peg.4585 | contig05429_3682_6465_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223310.peg.760 | contig00241_110649_108352_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223310.peg.2854 | contig01255_6231_8630_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin87 | fig\|6666666.223310.peg.464 | contig00130_194809_197127_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223310.peg.2317 | contig00899_60285_62474_+ | TonB-dependent siderophore receptor |
| | fig\|6666666.223310.peg.1583 | contig00547_81922_79724_- | TPR domain protein, putative component of TonB system |
| | fig\|6666666.223310.peg.1878 | contig00689_36916_36536_- | Peptidoglycan-binding LysM |
| | fig\|6666666.223310.peg.3409 | contig01663_16660_19797_+ | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223310.peg.1162 | contig00389_35865_33007_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223310.peg.2625 | contig01060_35900_33990_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223310.peg.2626 | contig01060_38490_35914_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223310.peg.2937 | contig01262_38110_35795_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223310.peg.3585 | contig02017_25686_25024_- | Beta-phosphoglucomutase (EC 5.4.2.6) |
| | fig\|6666666.223310.peg.2609 | contig01060_23366_24013_+ | Beta-phosphoglucomutase (EC 5.4.2.6) |
| | fig\|6666666.223310.peg.3699 | contig02051_16904_18355_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.223310.peg.3959 | contig02639_12419_13468_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.223310.peg.2179 | contig00837_60988_61881_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.223310.peg.2076 | contig00759_11963_11097_- | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.223310.peg.4414 | contig04451_7025_5946_- | L-alanine-DL-glutamate epimerase |
| | fig\|6666666.223310.peg.1858 | contig00689_15755_17212_+ | Aminoacyl-histidine dipeptidase (Peptidase D) (EC 3.4.13.3) |
| | fig\|6666666.223310.peg.3939 | contig02555_2285_3382_+ | Anhydro-N-acetylmuramic acid kinase (EC 2.7.1.-) |
| | fig\|6666666.223310.peg.3460 | contig01762_37126_38481_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223310.peg.1126 | contig00378_114065_112146_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin87 | fig\|6666666.223310.peg.3404 | contig01663_12305_10461_- | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |
| | fig\|6666666.223310.peg.659 | contig00159_190646_188721_- | Glucosamine-6-phosphate deaminase (EC 3.5.99.6) |
| | fig\|6666666.223310.peg.1419 | contig00500_10336_11070_+ | Lyzozyme M1 (1,4-beta-N-acetylmuramidase) (EC 3.2.1.17) |
| | fig\|6666666.223310.peg.3700 | contig02051_18390_19409_+ | Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13) |
| | fig\|6666666.223310.peg.974 | contig00292_87014_86199_- | N-Acetyl-D-glucosamine ABC transport system, permease protein 2 |
| | fig\|6666666.223310.peg.4417 | contig04451_12424_11306_- | N-acetylglucosamine related transporter, NagX |
| | fig\|6666666.223310.peg.2942 | contig01262_42130_42957_+ | N-acetylmuramic acid 6-phosphate etherase |
| | fig\|6666666.223310.peg.3748 | contig02264_14108_14740_+ | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.223310.peg.3279 | contig01408_16548_15136_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.223310.peg.1404 | contig00478_100818_99775_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.223310.peg.4276 | contig03511_2_1090_+ | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.223310.peg.236 | contig00041_288873_287476_- | Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|6666666.223310.peg.3261 | contig01408_2458_1082_- | Phosphomannomutase (EC 5.4.2.8) / Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|6666666.223310.peg.4138 | contig02949_18856_17426_- | N-acetylglucosamine deacetylase (EC 3.5.1.-) / 3-hydroxyacyl-[acyl-carrier-protein] dehydratase, FabZ form (EC 4.2.1.59) |
| | fig\|6666666.223310.peg.1184 | contig00389_68615_65214_- | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|6666666.223310.peg.1557 | contig00547_56285_54591_- | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|6666666.223310.peg.1674 | contig00559_79168_76841_- | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|6666666.223310.peg.2553 | contig01043_14349_13150_- | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|6666666.223310.peg.2554 | contig01043_15356_14430_- | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|6666666.223310.peg.3732 | contig02209_19572_21824_+ | Beta-glucosidase (EC 3.2.1.21) |
| | fig\|6666666.223310.peg.1260 | contig00450_35922_33208_- | Beta-glucosidase (EC 3.2.1.21) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin87 | fig\|6666666.223310.peg.1503 | contig00500_103473_105890_+ | Beta-glucosidase (EC 3.2.1.21) |
| | fig\|6666666.223310.peg.848 | contig00269_66003_63688_- | Beta-glucosidase (EC 3.2.1.21) |
| | fig\|6666666.223310.peg.929 | contig00292_22797_20287_- | Beta-mannosidase (EC 3.2.1.25) |
| | fig\|6666666.223310.peg.30 | contig00041_32052_34274_+ | beta-hexosaminidase precursor |
| | fig\|6666666.223310.peg.4070 | contig02760_15279_14341_- | ROK family Glucokinase with ambiguous substrate specificity |
| Bin74 | fig\|6666666.223311.peg.1001 | contig00112_192218_194164_+ | Glucosamine-6-phosphate deaminase (EC 3.5.99.6) |
| | fig\|6666666.223311.peg.1019 | contig00112_222840_219913_- | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.223311.peg.1027 | contig00112_230896_233265_+ | TonB-dependent receptor |
| | fig\|6666666.223311.peg.116 | contig00045_149005_146801_- | SusD, outer membrane protein |
| | fig\|6666666.223311.peg.117 | contig00045_151995_149032_- | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.223311.peg.1184 | contig00168_163303_164496_+ | Anhydro-N-acetylmuramic acid kinase (EC 2.7.1.-) |
| | fig\|6666666.223311.peg.1270 | contig00220_39162_41474_+ | TonB-dependent receptor, plug precursor |
| | fig\|6666666.223311.peg.1296 | contig00220_62539_61460_- | Endoglucanase |
| | fig\|6666666.223311.peg.1302 | contig00220_68542_70776_+ | TonB-dependent receptor |
| | fig\|6666666.223311.peg.1307 | contig00220_74005_77181_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223311.peg.1334 | contig00220_118433_117453_- | 6-phosphofructokinase (EC 2.7.1.11) |
| | fig\|6666666.223311.peg.1344 | contig00220_138205_139494_+ | SusD/RagB family protein |
| | fig\|6666666.223311.peg.1358 | contig00220_155936_155001_- | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|6666666.223311.peg.1361 | contig00220_161702_158589_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223311.peg.1419 | contig00226_58421_59785_+ | Phosphomannomutase (EC 5.4.2.8) / Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|6666666.223311.peg.1529 | contig00240_5119_3719_- | N-acetylglucosamine deacetylase (EC 3.5.1.-) / 3-hydroxyacyl-[acyl-carrier-protein] dehydratase, FabZ form (EC 4.2.1.59) |
| | fig\|6666666.223311.peg.161 | contig00045_210094_208751_- | LysM-repeat proteins and domains |
| | fig\|6666666.223311.peg.1679 | contig00240_157257_159464_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin74 | fig\|6666666.223311.peg.1698 | contig00247_6303_9068_+ | TonB-dependent receptor |
| | fig\|6666666.223311.peg.1872 | contig00255_43950_43006_- | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|6666666.223311.peg.1886 | contig00255_63908_62526_- | 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) |
| | fig\|6666666.223311.peg.1899 | contig00255_77552_79543_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223311.peg.1900 | contig00255_79506_79958_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223311.peg.1923 | contig00255_105123_104068_- | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|6666666.223311.peg.1934 | contig00255_127586_130750_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223311.peg.1945 | contig00255_147345_150020_+ | Beta-glucosidase (EC 3.2.1.21) |
| | fig\|6666666.223311.peg.2003 | contig00267_55827_55159_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223311.peg.2007 | contig00267_63139_60998_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223311.peg.2017 | contig00267_74901_73141_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223311.peg.2033 | contig00267_101673_104072_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223311.peg.2127 | contig00279_45347_47719_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223311.peg.2242 | contig00304_40929_39718_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223311.peg.2274 | contig00304_72958_74787_+ | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223311.peg.2275 | contig00304_77226_74788_- | putative TonB-dependent receptor |
| | fig\|6666666.223311.peg.2328 | contig00304_140984_143419_+ | TonB-dependent receptor |
| | fig\|6666666.223311.peg.2337 | contig00382_8100_6649_- | 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) |
| | fig\|6666666.223311.peg.2437 | contig00397_6636_6178_- | Endo-1,4-beta-D-glucanase |
| | fig\|6666666.223311.peg.2528 | contig00397_112906_115158_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223311.peg.2619 | contig00402_102021_105035_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223311.peg.2622 | contig00402_109703_113794_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223311.peg.2677 | contig00425_60820_59981_- | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|6666666.223311.peg.2730 | contig00508_31223_30228_- | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|6666666.223311.peg.2737 | contig00508_41128_38465_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin74 | fig\|6666666.223311.peg.2955 | contig00569_76743_80138_+ | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|6666666.223311.peg.2970 | contig00569_96424_94202_- | Regulatory sensor-transducer, BlaR1/MecR1 family / TonB-dependent receptor |
| | fig\|6666666.223311.peg.3030 | contig00596_59291_57069_- | TonB-dependent outer membrane receptor |
| | fig\|6666666.223311.peg.3040 | contig00596_65489_67858_+ | TonB-dependent receptor plug domain protein |
| | fig\|6666666.223311.peg.3093 | contig00626_26064_24610_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223311.peg.3110 | contig00626_50555_47733_- | TonB-dependent receptor, putative |
| | fig\|6666666.223311.peg.3119 | contig00626_61594_60371_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.223311.peg.3122 | contig00626_68322_66313_- | TonB-dependent receptor |
| | fig\|6666666.223311.peg.3148 | contig00737_3931_2468_- | Aminoacyl-histidine dipeptidase (Peptidase D) (EC 3.4.13.3) |
| | fig\|6666666.223311.peg.3169 | contig00737_28415_30895_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223311.peg.3320 | contig00756_45629_46717_+ | N-acetylglucosamine related transporter, NagX |
| | fig\|6666666.223311.peg.3339 | contig00756_71994_72746_+ | N-acetylmuramic acid 6-phosphate etherase |
| | fig\|6666666.223311.peg.3346 | contig00765_1256_183_- | L-alanine-DL-glutamate epimerase |
| | fig\|6666666.223311.peg.3425 | contig00812_18382_21549_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223311.peg.3449 | contig00812_56832_54574_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223311.peg.3557 | contig00945_6557_3576_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223311.peg.3563 | contig00945_19227_17845_- | 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) |
| | fig\|6666666.223311.peg.3594 | contig00945_57159_54226_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223311.peg.3654 | contig00949_53177_51966_- | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.223311.peg.3785 | contig01226_26143_25394_- | Regulatory sensor-transducer, BlaR1/MecR1 family / TonB-dependent receptor |
| | fig\|6666666.223311.peg.379 | contig00076_69282_71282_+ | 1,4-alpha-glucan (glycogen) branching enzyme, GH-13-type (EC 2.4.1.18) |
| | fig\|6666666.223311.peg.3830 | contig01229_27352_24728_- | TonB-dependent receptor |
| | fig\|6666666.223311.peg.3903 | contig01331_7526_9787_+ | Beta-hexosaminidase (EC 3.2.1.52) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin74 | fig\|6666666.223311.peg.3928 | contig01331_45425_42024_- | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|6666666.223311.peg.3951 | contig01419_17123_19513_+ | putative TonB-dependent receptor |
| | fig\|6666666.223311.peg.3976 | contig01419_44307_46109_+ | TonB-dependent receptor |
| | fig\|6666666.223311.peg.3980 | contig01466_3117_1423_- | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|6666666.223311.peg.4228 | contig01656_29573_32188_+ | TonB-dependent receptor |
| | fig\|6666666.223311.peg.4294 | contig01669_27802_25382_- | TonB-dependent receptor |
| | fig\|6666666.223311.peg.4370 | contig02798_12529_14154_+ | putative TonB-dependent receptor |
| | fig\|6666666.223311.peg.4521 | contig04220_5572_7947_+ | Beta-glucosidase (EC 3.2.1.21) |
| | fig\|6666666.223311.peg.4584 | contig05480_7265_4464_- | TonB-dependent receptor |
| | fig\|6666666.223311.peg.4628 | contig06439_2604_1462_- | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|6666666.223311.peg.4632 | contig06439_5535_7199_+ | Beta-xylosidase (EC 3.2.1.37) |
| | fig\|6666666.223311.peg.493 | contig00076_193129_191471_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223311.peg.620 | contig00078_32668_30371_- | TonB-dependent receptor, putative |
| | fig\|6666666.223311.peg.63 | contig00045_68520_71111_+ | TonB-dependent receptor |
| | fig\|6666666.223311.peg.656 | contig00078_78428_76413_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223311.peg.682 | contig00078_103499_100278_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223311.peg.853 | contig00112_23855_26587_+ | TonB-dependent receptor, plug precursor |
| | fig\|6666666.223311.peg.905 | contig00112_84028_86976_+ | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223311.peg.929 | contig00112_114813_113962_- | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.223311.peg.967 | contig00112_150074_151912_+ | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |
| Bin55 | fig\|6666666.223476.peg.1067 | contig00784_74831_72564_- | Beta-glucosidase (EC 3.2.1.21) |
| | fig\|6666666.223476.peg.1087 | contig00808_15052_17469_+ | TonB-dependent receptor |
| | fig\|6666666.223476.peg.1159 | contig00824_33825_31711_- | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223476.peg.1288 | contig00840_33654_30850_- | TonB-dependent receptor |
| | fig\|6666666.223476.peg.1344 | contig00852_30678_29587_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223476.peg.1693 | contig01072_15409_18348_+ | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.223476.peg.1694 | contig01072_18406_20604_+ | SusD, outer membrane protein |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin55 | fig\|6666666.223476.peg.1707 | contig01072_38549_36216_- | TonB-dependent receptor |
| | fig\|6666666.223476.peg.1806 | contig01107_27622_29961_+ | TonB-dependent receptor, plug precursor |
| | fig\|6666666.223476.peg.1818 | contig01107_38137_39189_+ | LysM-repeat proteins and domains |
| | fig\|6666666.223476.peg.1847 | contig01220_12847_15282_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223476.peg.1893 | contig01223_2000_4807_+ | TonB-dependent receptor |
| | fig\|6666666.223476.peg.1901 | contig01223_11476_14253_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223476.peg.1935 | contig01248_3572_4621_+ | Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13) |
| | fig\|6666666.223476.peg.207 | contig00475_106590_105505_- | Endoglucanase |
| | fig\|6666666.223476.peg.2089 | contig01357_17494_14495_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223476.peg.2102 | contig01357_47727_45292_- | TonB-dependent receptor |
| | fig\|6666666.223476.peg.2282 | contig01533_24071_24994_+ | Lyzozyme M1 (1,4-beta-N-acetylmuramidase) (EC 3.2.1.17) |
| | fig\|6666666.223476.peg.2304 | contig01539_2327_3403_+ | L-alanine-DL-glutamate epimerase |
| | fig\|6666666.223476.peg.2415 | contig01572_10836_19487_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223476.peg.2446 | contig01614_15656_17902_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223476.peg.2519 | contig01648_2821_1412_- | N-acetylglucosamine deacetylase (EC 3.5.1.-) / 3-hydroxyacyl-[acyl-carrier-protein] dehydratase, FabZ form (EC 4.2.1.59) |
| | fig\|6666666.223476.peg.2528 | contig01648_11484_9175_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223476.peg.2558 | contig01661_2549_3307_+ | Lyzozyme M1 (1,4-beta-N-acetylmuramidase) (EC 3.2.1.17) |
| | fig\|6666666.223476.peg.2611 | contig01662_16910_20323_+ | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|6666666.223476.peg.265 | contig00498_64573_62612_- | TonB-dependent receptor |
| | fig\|6666666.223476.peg.2679 | contig01690_22740_20404_- | TonB-dependent receptor |
| | fig\|6666666.223476.peg.2743 | contig01772_15348_16298_+ | N-acetyl-D-glucosamine kinase (EC 2.7.1.59) |
| | fig\|6666666.223476.peg.2901 | contig01889_25480_23675_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223476.peg.3095 | contig02009_19473_16924_- | TonB-dependent receptor, putative |
| | fig\|6666666.223476.peg.3100 | contig02009_30763_27827_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223476.peg.3101 | contig02009_33080_30852_- | Beta-glucosidase (EC 3.2.1.21) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin55 | fig\|6666666.223476.peg.3134 | contig02016_31889_30447_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223476.peg.3158 | contig02023_19222_18047_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223476.peg.329 | contig00503_42662_41682_- | Membrane-bound lytic murein transglycosylase A precursor (EC 3.2.1.-) |
| | fig\|6666666.223476.peg.3296 | contig02102_17799_20195_+ | TonB-dependent receptor plug domain protein |
| | fig\|6666666.223476.peg.3404 | contig02276_7009_6155_- | D-alanyl-D-alanine carboxypeptidase |
| | fig\|6666666.223476.peg.3416 | contig02276_22610_23512_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.223476.peg.3663 | contig02477_4061_2382_- | putative TonB-dependent receptor |
| | fig\|6666666.223476.peg.3807 | contig02694_27306_25516_- | TonB-dependent receptor, putative |
| | fig\|6666666.223476.peg.3847 | contig02784_12154_10964_- | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|6666666.223476.peg.3848 | contig02784_13008_12181_- | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|6666666.223476.peg.3852 | contig02784_18677_15630_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223476.peg.4017 | contig03231_19853_17313_- | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223476.peg.4095 | contig03547_1419_1808_+ | LysM domain protein |
| | fig\|6666666.223476.peg.4106 | contig03547_10148_12475_+ | TonB-dependent receptor, plug precursor |
| | fig\|6666666.223476.peg.4131 | contig03600_19261_16733_- | TonB-dependent receptor, putative |
| | fig\|6666666.223476.peg.4139 | contig03651_5597_4086_- | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.223476.peg.4150 | contig03662_6753_11651_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223476.peg.4163 | contig03727_4783_2123_- | TonB-dependent receptor |
| | fig\|6666666.223476.peg.4174 | contig03727_14407_17370_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223476.peg.4254 | contig04018_7481_9250_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223476.peg.4257 | contig04018_10625_13090_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223476.peg.4444 | contig04844_16_2307_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223476.peg.4461 | contig04863_27_1877_+ | Chitinase (EC 3.2.1.14) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin55 | fig\|6666666.223476.peg.4562 | contig05207_14729_12381_- | TonB-dependent receptor, putative |
| | fig\|6666666.223476.peg.4588 | contig05634_3409_4611_+ | Chitinase |
| | fig\|6666666.223476.peg.4631 | contig05784_2262_3992_+ | endo-1,3-1,4-beta-glucanase |
| | fig\|6666666.223476.peg.4772 | contig07534_1697_393_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.223476.peg.48 | contig00411_49070_47799_- | N-acetyl glucosamine transporter, NagP |
| | fig\|6666666.223476.peg.4805 | contig07775_9160_9540_+ | TonB family protein |
| | fig\|6666666.223476.peg.507 | contig00587_53251_50519_- | TonB-dependent receptor |
| | fig\|6666666.223476.peg.5151 | contig27350_2904_1570_- | Aminoacyl-histidine dipeptidase (Peptidase D) (EC 3.4.13.3) |
| | fig\|6666666.223476.peg.5154 | contig28171_2872_2393_- | TonB-dependent outer membrane receptor |
| | fig\|6666666.223476.peg.580 | contig00624_54243_52567_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223476.peg.600 | contig00624_81199_82377_+ | Beta-glucosidase (EC 3.2.1.21) |
| | fig\|6666666.223476.peg.621 | contig00639_10998_11807_+ | N-acetylmuramic acid 6-phosphate etherase |
| | fig\|6666666.223476.peg.69 | contig00411_68343_65764_- | Beta-mannosidase (EC 3.2.1.25) |
| | fig\|6666666.223476.peg.698 | contig00672_9628_8210_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223476.peg.72 | contig00411_70340_72529_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| Bin09 | fig\|6666666.223312.peg.1042 | contig00037_352442_351063_- | LysM-repeat proteins and domains |
| | fig\|6666666.223312.peg.150 | contig00003_162082_163365_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.223312.peg.230 | contig00003_252560_253447_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.223312.peg.179 | contig00003_193802_196627_+ | putative TonB-dependent receptor |
| | fig\|6666666.223312.peg.469 | contig00003_525060_527486_+ | putative TonB-dependent receptor |
| | fig\|6666666.223312.peg.2317 | contig00266_118609_116384_- | TonB-dependent outer membrane receptor |
| | fig\|6666666.223312.peg.959 | contig00037_239898_237430_- | TonB-dependent receptor |
| | fig\|6666666.223312.peg.1549 | contig00160_201982_205044_+ | TonB-dependent receptor |
| | fig\|6666666.223312.peg.2138 | contig00258_54230_56557_+ | TonB-dependent receptor |
| | fig\|6666666.223312.peg.2811 | contig00601_72455_75136_+ | TonB-dependent receptor |
| | fig\|6666666.223312.peg.3005 | contig00895_20054_17286_- | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin09 | fig\|6666666.223312.peg.1699 | contig00165_159234_161588_+ | TonB-dependent receptor, plug precursor |
| | fig\|6666666.223312.peg.1709 | contig00165_176046_178526_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223312.peg.2533 | contig00453_84531_82135_- | TonB-dependent receptor, putative |
| | fig\|6666666.223312.peg.2930 | contig00783_41481_43799_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223312.peg.1407 | contig00160_58334_55749_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223312.peg.1563 | contig00165_7815_9944_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223312.peg.2457 | contig00453_1057_3477_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223312.peg.2527 | contig00453_73998_76163_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223312.peg.731 | contig00003_819369_822506_+ | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|6666666.223312.peg.580 | contig00003_654862_652571_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223312.peg.3408 | contig05478_14188_12680_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223312.peg.2012 | contig00250_77641_75167_- | Beta-mannosidase (EC 3.2.1.25) |
| | fig\|6666666.223312.peg.1041 | contig00037_349974_351053_+ | Endoglucanase |
| | fig\|6666666.223312.peg.2079 | contig00250_153630_152728_- | Endoglucanase (EC 3.2.1.4) |
| | fig\|6666666.223312.peg.3058 | contig01181_30419_29601_- | N-acetylmuramic acid 6-phosphate etherase |
| | fig\|6666666.223312.peg.1544 | contig00160_196550_195738_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.223312.peg.1634 | contig00165_90655_92046_+ | N-acetylglucosamine deacetylase (EC 3.5.1.-) / 3-hydroxyacyl-[acyl-carrier-protein] dehydratase, FabZ form (EC 4.2.1.59) |
| | fig\|6666666.223312.peg.246 | contig00003_267166_268242_+ | Anhydro-N-acetylmuramic acid kinase (EC 2.7.1.-) |
| | fig\|6666666.223312.peg.939 | contig00037_216606_215335_- | 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) |
| | fig\|6666666.223312.peg.2604 | contig00485_45932_47317_+ | Phosphomannomutase (EC 5.4.2.8) / Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|6666666.223312.peg.410 | contig00003_456759_454921_- | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin09 | fig\|6666666.223312.peg.1224 | contig00074_143085_144785_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223312.peg.1229 | contig00074_148304_149977_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223312.peg.1991 | contig00250_56246_57988_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223312.peg.2116 | contig00258_26577_25126_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| Bin56 | fig\|6666666.223473.peg.1011 | contig00067_213378_215849_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223473.peg.1043 | contig00067_249936_252713_+ | TonB-dependent receptor |
| | fig\|6666666.223473.peg.1054 | contig00067_269642_271675_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223473.peg.1112 | contig00107_11237_13075_+ | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |
| | fig\|6666666.223473.peg.1130 | contig00107_33534_32143_- | 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) |
| | fig\|6666666.223473.peg.130 | contig00021_149056_151725_+ | Beta-glucosidase (EC 3.2.1.21) |
| | fig\|6666666.223473.peg.1479 | contig00194_155812_153245_- | TonB-dependent receptor |
| | fig\|6666666.223473.peg.1661 | contig00214_146332_144128_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223473.peg.1665 | contig00214_152869_150725_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223473.peg.1728 | contig00320_40498_43248_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223473.peg.1788 | contig00320_120094_116948_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223473.peg.1861 | contig00324_47081_45747_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223473.peg.203 | contig00021_232155_233258_+ | N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25) |
| | fig\|6666666.223473.peg.2044 | contig00338_96302_99427_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223473.peg.2047 | contig00338_103083_103982_+ | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|6666666.223473.peg.2048 | contig00338_104040_105056_+ | Beta-glucanase precursor (EC 3.2.1.73) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin56 | fig\|6666666.223473.peg.2049 | contig00338_105083_105979_+ | Beta-1,3(4)-glucanase precursor (EC 3.2.1.6) |
| | fig\|6666666.223473.peg.2072 | contig00338_128956_126920_- | 1,4-alpha-glucan (glycogen) branching enzyme, GH-13-type (EC 2.4.1.18) |
| | fig\|6666666.223473.peg.2091 | contig00356_10379_11533_+ | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223473.peg.2278 | contig00363_71091_73514_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223473.peg.2371 | contig00434_45896_44574_- | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.223473.peg.2458 | contig00435_28318_25934_- | TonB-dependent receptor, putative |
| | fig\|6666666.223473.peg.2476 | contig00435_58094_57915_- | N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25) |
| | fig\|6666666.223473.peg.2478 | contig00435_61723_58718_- | TonB-dependent receptor |
| | fig\|6666666.223473.peg.2574 | contig00471_41363_40284_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.223473.peg.2648 | contig00471_102437_101664_- | Lyzozyme M1 (1,4-beta-N-acetylmuramidase) (EC 3.2.1.17) |
| | fig\|6666666.223473.peg.2672 | contig00515_26497_29472_+ | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.223473.peg.2673 | contig00515_29490_31103_+ | SusD, outer membrane protein |
| | fig\|6666666.223473.peg.2813 | contig00527_86239_84860_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223473.peg.2839 | contig00607_15014_12387_- | TonB-dependent receptor, putative |
| | fig\|6666666.223473.peg.3069 | contig00754_3517_1943_- | SusD, outer membrane protein |
| | fig\|6666666.223473.peg.3121 | contig00754_69440_68094_- | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223473.peg.3191 | contig00857_65337_65759_+ | N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25) |
| | fig\|6666666.223473.peg.3293 | contig00874_29698_29276_- | Peptidoglycan-binding LysM |
| | fig\|6666666.223473.peg.3298 | contig00874_33233_35605_+ | TonB-dependent receptor, plug precursor |
| | fig\|6666666.223473.peg.3361 | contig00898_27466_25523_- | Glucosamine-6-phosphate deaminase (EC 3.5.99.6) |
| | fig\|6666666.223473.peg.3362 | contig00898_30121_27629_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223473.peg.3395 | contig00898_61763_64219_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223473.peg.3396 | contig00898_64491_66965_+ | TonB-dependent receptor, putative |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin56 | fig\|6666666.223473.peg.3422 | contig00942_33739_34626_+ | Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13) |
| | fig\|6666666.223473.peg.3471 | contig00976_23235_20953_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223473.peg.3549 | contig01071_40355_38976_- | Phosphomannomutase (EC 5.4.2.8) / Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|6666666.223473.peg.3587 | contig01105_20114_17673_- | TonB-dependent receptor, putative |
| | fig\|6666666.223473.peg.3594 | contig01105_28180_26513_- | putative TonB-dependent receptor |
| | fig\|6666666.223473.peg.3600 | contig01105_35858_34458_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223473.peg.3647 | contig01455_25735_23369_- | TonB-dependent receptor plug domain protein |
| | fig\|6666666.223473.peg.365 | contig00021_404417_407722_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223473.peg.3669 | contig01483_3713_1314_- | TonB-dependent receptor, putative |
| | fig\|6666666.223473.peg.3711 | contig01579_6230_7999_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223473.peg.3729 | contig01579_26630_25584_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.223473.peg.3815 | contig01645_39732_38209_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223473.peg.387 | contig00021_439291_438554_- | Glucosamine-6-phosphate deaminase (EC 3.5.99.6) |
| | fig\|6666666.223473.peg.3906 | contig01802_31039_30749_- | 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) |
| | fig\|6666666.223473.peg.400 | contig00021_453344_455566_+ | Beta-glucosidase (EC 3.2.1.21) |
| | fig\|6666666.223473.peg.4040 | contig02476_29046_28624_- | N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25) |
| | fig\|6666666.223473.peg.4120 | contig03253_10709_11503_+ | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|6666666.223473.peg.4274 | contig20267_2376_1558_- | N-acetylmuramic acid 6-phosphate etherase |
| | fig\|6666666.223473.peg.458 | contig00047_1635_4847_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223473.peg.464 | contig00047_11687_13294_+ | Beta-xylosidase (EC 3.2.1.37) |
| | fig\|6666666.223473.peg.512 | contig00047_59615_60682_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin56 | fig\|6666666.223473.peg.573 | contig00047_131318_132721_+ | N-acetylglucosamine deacetylase (EC 3.5.1.-) / 3-hydroxyacyl-[acyl-carrier-protein] dehydratase, FabZ form (EC 4.2.1.59) |
| | fig\|6666666.223473.peg.6 | contig00021_7366_5192_- | TonB-dependent receptor, putative |
| | fig\|6666666.223473.peg.652 | contig00047_218066_220273_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223473.peg.679 | contig00047_252902_254830_+ | TonB-dependent receptor |
| | fig\|6666666.223473.peg.736 | contig00047_315296_316399_+ | N-acetylglucosamine related transporter, NagX |
| | fig\|6666666.223473.peg.748 | contig00047_327365_326547_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.223473.peg.749 | contig00047_328354_327401_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.223473.peg.804 | contig00067_6165_6596_+ | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.223473.peg.823 | contig00067_33611_30663_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223473.peg.830 | contig00067_43336_41951_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223473.peg.840 | contig00067_51701_52756_+ | TonB-dependent receptor |
| | fig\|6666666.223473.peg.999 | contig00067_200702_197577_- | TonB family protein / TonB-dependent receptor |
| Bin08 | fig\|6666666.223313.peg.1027 | contig00006_61173_59977_- | N-acylglucosamine 2-epimerase (EC 5.1.3.8) |
| | fig\|6666666.223313.peg.1085 | contig00006_131934_133337_+ | N-acetylglucosamine deacetylase (EC 3.5.1.-) / 3-hydroxyacyl-[acyl-carrier-protein] dehydratase, FabZ form (EC 4.2.1.59) |
| | fig\|6666666.223313.peg.109 | contig00002_137649_142319_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223313.peg.11 | contig00002_12126_12851_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.12 | contig00002_12966_15533_+ | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.223313.peg.1253 | contig00006_312644_313813_+ | N-acetylglucosamine related transporter, NagX |
| | fig\|6666666.223313.peg.1285 | contig00006_347927_347112_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.223313.peg.1289 | contig00006_350700_353162_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223313.peg.13 | contig00002_15551_17224_+ | SusD, outer membrane protein |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin08 | fig\|6666666.223313.peg.1371 | contig00006_437379_434923_- | TonB-dependent receptor, putative |
| | fig\|6666666.223313.peg.1385 | contig00006_451590_454550_+ | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.223313.peg.1435 | contig00006_510839_507867_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.1446 | contig00006_526154_523212_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223313.peg.1451 | contig00006_533184_531811_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223313.peg.1459 | contig00006_540027_541091_+ | TonB-dependent receptor |
| | fig\|6666666.223313.peg.1531 | contig00006_605056_603908_- | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223313.peg.1557 | contig00006_625132_626511_+ | N-acylglucosamine 2-epimerase |
| | fig\|6666666.223313.peg.1637 | contig00006_710899_708584_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.1696 | contig00027_20521_19751_- | Lyzozyme M1 (1,4-beta-N-acetylmuramidase) (EC 3.2.1.17) |
| | fig\|6666666.223313.peg.175 | contig00002_208768_207464_- | Endo-1,4-beta-xylanase B precursor |
| | fig\|6666666.223313.peg.1758 | contig00027_88747_87602_- | N-acetylglucosamine related transporter, NagX |
| | fig\|6666666.223313.peg.1770 | contig00027_97120_98235_+ | L-alanine-DL-glutamate epimerase |
| | fig\|6666666.223313.peg.1829 | contig00027_153471_156788_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.1861 | contig00027_197802_200765_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.1862 | contig00027_200785_202347_+ | SusD, outer membrane protein |
| | fig\|6666666.223313.peg.1879 | contig00027_223693_226674_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.19 | contig00002_27258_26185_- | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|6666666.223313.peg.1904 | contig00027_254193_255938_+ | Beta-xylosidase (EC 3.2.1.37) |
| | fig\|6666666.223313.peg.1911 | contig00027_261940_262875_+ | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|6666666.223313.peg.1952 | contig00027_306287_304380_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223313.peg.2006 | contig00027_359061_361418_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223313.peg.2061 | contig00027_420063_422879_+ | TonB-dependent receptor |
| | fig\|6666666.223313.peg.212 | contig00002_239587_242916_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.213 | contig00002_242956_244479_+ | RagB/SusD domain protein |
| | fig\|6666666.223313.peg.2140 | contig00053_77631_75313_- | TonB-dependent receptor |
| | fig\|6666666.223313.peg.2143 | contig00053_78736_79386_+ | Beta-phosphoglucomutase (EC 5.4.2.6) |
| | fig\|6666666.223313.peg.2165 | contig00053_109745_107712_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin08 | fig\|6666666.223313.peg.2184 | contig00053_134423_131340_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.2227 | contig00053_186668_184791_- | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223313.peg.2228 | contig00053_188397_186679_- | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223313.peg.2236 | contig00053_193355_196063_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223313.peg.2257 | contig00053_215167_216288_+ | Endoglucanase |
| | fig\|6666666.223313.peg.228 | contig00002_260778_259399_- | 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) |
| | fig\|6666666.223313.peg.2362 | contig00053_302504_307426_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223313.peg.2382 | contig00053_320928_321659_+ | Glucosamine-6-phosphate deaminase (EC 3.5.99.6) |
| | fig\|6666666.223313.peg.2400 | contig00053_340121_343048_+ | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.223313.peg.2401 | contig00053_343065_344633_+ | SusD, outer membrane protein |
| | fig\|6666666.223313.peg.2483 | contig00060_66934_68004_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.223313.peg.2555 | contig00060_143030_140484_- | TonB-dependent receptor |
| | fig\|6666666.223313.peg.26 | contig00002_37670_34467_- | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.223313.peg.2641 | contig00060_241852_240722_- | Anhydro-N-acetylmuramic acid kinase (EC 2.7.1.-) |
| | fig\|6666666.223313.peg.2770 | contig00066_28109_25533_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223313.peg.2773 | contig00066_31993_30809_- | N-acylglucosamine 2-epimerase (EC 5.1.3.8) |
| | fig\|6666666.223313.peg.2777 | contig00066_36642_35680_- | N-acetylneuraminate lyase (EC 4.1.3.3) |
| | fig\|6666666.223313.peg.2779 | contig00066_41786_38382_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.2787 | contig00066_49238_50530_+ | N-acetyl glucosamine transporter, NagP |
| | fig\|6666666.223313.peg.287 | contig00002_334097_335497_+ | Phosphomannomutase (EC 5.4.2.8) / Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|6666666.223313.peg.2905 | contig00066_167606_165264_- | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin08 | fig\|6666666.223313.peg.2945 | contig00066_211296_208162_- | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.223313.peg.2965 | contig00066_229292_231721_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223313.peg.2983 | contig00066_248130_250559_+ | putative TonB-dependent receptor |
| | fig\|6666666.223313.peg.3040 | contig00066_326235_322756_- | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.223313.peg.3118 | contig00122_91334_90417_- | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.223313.peg.3133 | contig00122_105296_106675_+ | 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) |
| | fig\|6666666.223313.peg.3147 | contig00122_122599_120761_- | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |
| | fig\|6666666.223313.peg.3332 | contig00132_47530_49641_+ | Beta-xylosidase (EC 3.2.1.37) |
| | fig\|6666666.223313.peg.3378 | contig00132_91960_94650_+ | TonB-dependent receptor |
| | fig\|6666666.223313.peg.3395 | contig00132_113272_110660_- | TonB-dependent receptor, putative |
| | fig\|6666666.223313.peg.3503 | contig00132_222756_221560_- | N-acetylglucosamine related transporter, NagX |
| | fig\|6666666.223313.peg.3611 | contig00135_89548_92862_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.3612 | contig00135_92873_94252_+ | RagB/SusD domain protein |
| | fig\|6666666.223313.peg.371 | contig00002_400436_402178_+ | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.223313.peg.3773 | contig00151_36528_37658_+ | N-acetylglucosamine related transporter, NagX |
| | fig\|6666666.223313.peg.3776 | contig00151_41381_43453_+ | TonB-dependent receptor |
| | fig\|6666666.223313.peg.3842 | contig00151_112361_111615_- | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|6666666.223313.peg.3843 | contig00151_113640_112618_- | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|6666666.223313.peg.3844 | contig00151_114545_113646_- | Beta-glucanase precursor (EC 3.2.1.73) |
| | fig\|6666666.223313.peg.3847 | contig00151_121305_118192_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.3851 | contig00151_127483_128742_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223313.peg.3906 | contig00151_164123_163257_- | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|6666666.223313.peg.3917 | contig00151_172346_173398_+ | D-alanyl-D-alanine carboxypeptidase( EC:3.4.16.4 ) |
| | fig\|6666666.223313.peg.400 | contig00002_427024_428211_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223313.peg.4012 | contig00176_38497_37604_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin08 | fig\|6666666.223313.peg.4051 | contig00176_78081_75556_- | TonB-dependent receptor |
| | fig\|6666666.223313.peg.4117 | contig00176_135527_134415_- | beta-hexosaminidase precursor |
| | fig\|6666666.223313.peg.4186 | contig00176_202017_200812_- | Membrane-bound lytic murein transglycosylase |
| | fig\|6666666.223313.peg.42 | contig00002_54546_52621_- | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|6666666.223313.peg.4226 | contig00201_37784_40276_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223313.peg.4239 | contig00201_52237_55461_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.44 | contig00002_55737_54661_- | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|6666666.223313.peg.4451 | contig00257_81977_80895_- | conserved hypothetical protein, with LysM-repeats |
| | fig\|6666666.223313.peg.4509 | contig00257_149436_147082_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.4552 | contig00286_42777_40450_- | Beta-glucosidase (EC 3.2.1.21) |
| | fig\|6666666.223313.peg.4556 | contig00286_48909_46288_- | TonB-dependent receptor, putative |
| | fig\|6666666.223313.peg.4561 | contig00286_51903_52946_+ | Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13) |
| | fig\|6666666.223313.peg.4563 | contig00286_54600_55511_+ | Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13) |
| | fig\|6666666.223313.peg.4614 | contig00286_129457_128564_- | Endoglucanase (EC 3.2.1.4) |
| | fig\|6666666.223313.peg.4615 | contig00286_129792_129577_- | Endoglucanase (EC 3.2.1.4) |
| | fig\|6666666.223313.peg.4620 | contig00286_135640_134195_- | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|6666666.223313.peg.4623 | contig00286_142049_138897_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.47 | contig00002_58200_56764_- | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|6666666.223313.peg.476 | contig00002_520474_518546_- | Glucosamine-6-phosphate deaminase (EC 3.5.99.6) |
| | fig\|6666666.223313.peg.477 | contig00002_523851_521503_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223313.peg.4984 | contig00380_68852_70024_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223313.peg.5018 | contig00380_96211_98319_+ | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|6666666.223313.peg.5088 | contig00406_46681_43514_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.5090 | contig00406_51645_48637_- | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.223313.peg.5118 | contig00406_78593_76941_- | putative TonB-dependent receptor |
| | fig\|6666666.223313.peg.5125 | contig00406_87453_86071_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223313.peg.5226 | contig00492_63262_61160_- | Beta-galactosidase (EC 3.2.1.23) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin08 | fig\|6666666.223313.peg.5243 | contig00492_90271_87182_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.5248 | contig00492_99516_96424_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.5259 | contig00553_3705_877_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.223313.peg.5356 | contig00595_14828_15949_+ | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|6666666.223313.peg.5359 | contig00595_19730_20896_+ | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|6666666.223313.peg.5367 | contig00595_28546_29406_+ | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|6666666.223313.peg.537 | contig00002_571631_572515_+ | endo-1,4-beta-xylanase B |
| | fig\|6666666.223313.peg.5377 | contig00595_38918_40627_+ | Beta-xylosidase (EC 3.2.1.37) |
| | fig\|6666666.223313.peg.5419 | contig00595_93138_88621_- | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.223313.peg.5487 | contig00764_10180_10965_+ | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223313.peg.5488 | contig00764_10970_13123_+ | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223313.peg.5489 | contig00764_13281_15947_+ | Beta-glucosidase (EC 3.2.1.21) |
| | fig\|6666666.223313.peg.5492 | contig00764_19766_17583_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223313.peg.5496 | contig00764_27437_25596_- | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|6666666.223313.peg.55 | contig00002_70996_68381_- | Beta-glucosidase (EC 3.2.1.21) |
| | fig\|6666666.223313.peg.5503 | contig00764_44143_40922_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.223313.peg.5551 | contig00798_13278_15737_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223313.peg.5591 | contig00798_55855_54734_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.223313.peg.5721 | contig00846_45315_44134_- | N-acylglucosamine 2-epimerase (EC 5.1.3.8) |
| | fig\|6666666.223313.peg.5729 | contig00846_54072_53032_- | Endoglucanase E precursor (EC 3.2.1.4) (EgE) (Endo-1,4-beta-glucanase E) (Cellulase E) |
| | fig\|6666666.223313.peg.5887 | contig00930_60173_57849_- | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|6666666.223313.peg.590 | contig00002_625398_626186_+ | Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8) |
| | fig\|6666666.223313.peg.5943 | contig01020_48635_49813_+ | LysM-repeat proteins and domains |
| | fig\|6666666.223313.peg.6206 | contig08920_491_66_- | Peptidoglycan-binding LysM |
| | fig\|6666666.223313.peg.642 | contig00002_677264_679675_+ | TonB-dependent receptor, putative |
| | fig\|6666666.223313.peg.775 | contig00002_816298_818103_+ | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.223313.peg.776 | contig00002_818130_819296_+ | SusC, outer membrane protein involved in starch binding |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin08 | fig\|6666666.223313.peg.777 | contig00002_819309_821000_+ | SusD, outer membrane protein |
| | fig\|6666666.223313.peg.779 | contig00002_822269_824275_+ | 1,4-alpha-glucan (glycogen) branching enzyme, GH-13-type (EC 2.4.1.18) |
| | fig\|6666666.223313.peg.793 | contig00002_835215_836030_+ | N-acetylmuramic acid 6-phosphate etherase |
| | fig\|6666666.223313.peg.813 | contig00002_862681_861731_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.223313.peg.857 | contig00002_907707_909275_+ | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.223313.peg.931 | contig00002_980911_979931_- | Chitinase (EC 3.2.1.14) |
| *Dechloromonas aromatica* RCB (NC_007298) | fig\|159087.6.peg.652 | Daro_0643 CDS_706428_707039_- (YP_283870.1) | FIG016425: Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains) |
| | fig\|159087.6.peg.3683 | Daro_3704 CDS_3981962_3982936_- (YP_286903.1) | Membrane-bound lytic murein transglycosylase A |
| | fig\|159087.6.peg.2506 | Daro_2520 CDS_2718014_2719057_+ (YP_285723.1) | Membrane-bound lytic murein transglycosylase B |
| | fig\|159087.6.peg.1315 | Daro_1320 CDS_1436224_1437762_- (YP_284540.1) | Membrane-bound lytic murein transglycosylase D |
| | fig\|159087.6.peg.2435 | Daro_2441 CDS_2642738_2643055_+ (YP_285646.1) | Membrane-bound lytic murein transglycosylase D |
| | fig\|159087.6.peg.1298 | Daro_1301 CDS_1411728_1413263_- (YP_284523.1) | Membrane-bound lytic murein transglycosylase F (EC 4.2.2.n1) |
| | fig\|159087.6.peg.4120 | Daro_4132 CDS_4430201_4432129_+ (YP_287328.1) | Soluble lytic murein transglycosylase precursor (EC 3.2.1.-) |
| | fig\|159087.6.peg.1042 | Daro_1044 CDS_1144121_1146301_+ (YP_284270.1) | Probable tonB-dependent receptor yncD precursor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Dechloromonas aromatica* RCB (NC_007298) | fig\|159087.6.peg.2932 | Daro_2953 CDS_3189600_3191612_+ (YP_286153.1) | Probable tonB-dependent receptor yncD precursor |
| | fig\|159087.6.peg.21 | Daro_0020 CDS_25900_26931_- (YP_283249.1) | Uncharacterized protein with LysM domain, COG1652 |
| | fig\|159087.6.peg.2022 | Daro_2034 CDS_2184790_2185809_+ (YP_285249.1) | beta-N-acetylglucosaminidase (EC 3.2.1.52) |
| | fig\|159087.6.peg.1429 | Daro_1444 CDS_1570887_1572524_+ (YP_284663.1) | Phosphoglucomutase (EC 5.4.2.2) |
| | fig\|159087.6.peg.3285 | Daro_3299 CDS_3547851_3549227_+ (YP_286499.1) | Phosphoglucomutase (EC 5.4.2.2) @ Phosphomannomutase (EC 5.4.2.8) |
| | fig\|159087.6.peg.594 | Daro_0586 CDS_657897_659762_+ (YP_283813.1) | 1,4-alpha-glucan (glycogen) branching enzyme, GH-13-type (EC 2.4.1.18) |
| | fig\|159087.6.peg.945 | Daro_0946 CDS_1026760_1028037_+ (YP_284172.1) | Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|159087.6.peg.3918 | Daro_3931 CDS_4226551_4228374_+ (YP_287129.1) | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |
| | fig\|159087.6.peg.3028 | Daro_3049 CDS_3293633_3294982_- (YP_286249.1) | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|159087.6.peg.299 | Daro_0291 CDS_332243_333382_- (YP_283520.1) | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| Bin79 | fig\|6666666.177556.peg.1211 | contig00145_177247_175868_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.177556.peg.1293 | contig00157_36058_37419_+ | Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|6666666.177556.peg.1308 | contig00157_51391_52131_+ | LysM domain protein |
| | fig\|6666666.177556.peg.1311 | contig00157_53976_55442_+ | LysM domain protein |
| | fig\|6666666.177556.peg.1388 | contig00157_139491_137566_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin79 | fig\|6666666.177556.peg.1409 | contig00157_159389_158217_- | 6-phosphofructokinase (EC 2.7.1.11) |
| | fig\|6666666.177556.peg.157 | contig00070_155763_157340_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.177556.peg.1996 | contig00275_16532_14553_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.177556.peg.2003 | contig00275_23936_21846_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.177556.peg.203 | contig00070_202830_203645_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.177556.peg.2226 | contig00334_121635_122315_+ | Beta-phosphoglucomutase (EC 5.4.2.6) |
| | fig\|6666666.177556.peg.2239 | contig00334_132908_132294_- | Membrane-bound lytic murein transglycosylase C precursor (EC 3.2.1.-) |
| | fig\|6666666.177556.peg.2389 | contig00353_7229_6402_- | TonB-dependent receptor, putative |
| | fig\|6666666.177556.peg.2583 | contig00365_82258_84357_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.177556.peg.2613 | contig00365_114427_116004_+ | LysM domain protein |
| | fig\|6666666.177556.peg.3035 | contig00575_50137_48263_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.177556.peg.3353 | contig00673_75051_73549_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.177556.peg.344 | contig00089_44064_46283_+ | Soluble lytic murein transglycosylase precursor (EC 3.2.1.-) |
| | fig\|6666666.177556.peg.3622 | contig00828_31689_33803_+ | TonB-dependent receptor, plug |
| | fig\|6666666.177556.peg.3724 | contig00907_62697_61696_- | Membrane-bound lytic murein transglycosylase B precursor (EC 3.2.1.-) |
| | fig\|6666666.177556.peg.3747 | contig01085_13271_14977_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.177556.peg.3751 | contig01085_17800_18750_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin79 | fig\|6666666.177556.peg.3756 | contig01085_23554_22376_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.177556.peg.3814 | contig01134_20543_21832_+ | 6-phosphofructokinase (EC 2.7.1.11) |
| | fig\|6666666.177556.peg.3945 | contig01392_11522_12592_+ | Endoglucanase (EC 3.2.1.4) |
| | fig\|6666666.177556.peg.4003 | contig01444_22191_22988_+ | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |
| | fig\|6666666.177556.peg.4377 | contig08296_5810_3981_- | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |
| | fig\|6666666.177556.peg.906 | contig00142_72437_74512_+ | TonB-dependent receptor |
| Bin03 | fig\|6666666.177616.peg.1010 | contig01754_35480_34947_- | TonB-dependent receptor |
| | fig\|6666666.177616.peg.1181 | contig01849_34708_35709_+ | Endoglucanase precursor (EC 3.2.1.4) |
| | fig\|6666666.177616.peg.1534 | contig02222_15941_14736_- | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|6666666.177616.peg.1559 | contig02254_10274_11650_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.177616.peg.1674 | contig02407_15134_18271_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.177616.peg.1784 | contig02480_10778_10251 | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|6666666.177616.peg.1791 | contig02480_20467_19163_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.177616.peg.1904 | contig02721_2647_1409_- | Membrane-bound lytic murein transglycosylase A precursor (EC 3.2.1.-) |
| | fig\|6666666.177616.peg.1999 | contig02832_21200_19056_- | Beta-glucosidase (EC 3.2.1.21) |
| | fig\|6666666.177616.peg.204 | contig01024_31464_33398_+ | TonB-dependent receptor |
| | fig\|6666666.177616.peg.2040 | contig02963_21220_19853_- | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.177616.peg.2160 | contig03224_9937_9041_- | Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13) |
| | fig\|6666666.177616.peg.2426 | contig03674_5398_6252_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.177616.peg.2529 | contig03980_2293_2_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.177616.peg.2719 | contig04339_15131_17374_+ | TonB-dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin03 | fig\|6666666.177616.peg.2899 | contig04874_2378_1560_- | N-Acetyl-D-glucosamine ABC transport system, permease protein 2 |
| | fig\|6666666.177616.peg.2943 | contig04948_8107_7040_- | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.177616.peg.3032 | contig05447_37_3633_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.177616.peg.3139 | contig05880_6219_7358_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.177616.peg.3267 | contig06866_10_591_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.177616.peg.3305 | contig07140_6825_5065_- | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|6666666.177616.peg.3334 | contig07226_3338_4552_+ | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|6666666.177616.peg.3335 | contig07226_4620_6278_+ | Phosphoglucomutase (EC 5.4.2.2) |
| | fig\|6666666.177616.peg.3431 | contig07740_3128_1092 | TonB-dependent receptor |
| | fig\|6666666.177616.peg.3612 | contig08532_5174_6985_+ | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |
| | fig\|6666666.177616.peg.367 | contig01177_56051_54561_- | TonB-dependent receptor |
| | fig\|6666666.177616.peg.4197 | contig20293_256_1062_+ | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.177616.peg.597 | contig01295_43691_45055_+ | Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|6666666.177616.peg.620 | contig01313_13882_12941_- | N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25) |
| | fig\|6666666.177616.peg.704 | contig01501_16823_18115_+ | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.177616.peg.756 | contig01505_35825_34473_- | Beta-glucosidase (EC 3.2.1.21) |
| | fig\|6666666.177616.peg.764 | contig01505_43767_42652_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.177616.peg.824 | contig01615_24351_25661_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.177616.peg.832 | contig01615_35613_41477_+ | Beta-galactosidase (EC 3.2.1.23) |
| | fig\|6666666.177616.peg.906 | contig01694_5653_7437_+ | Chitinase (EC 3.2.1.14) |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Geobacter metallireducens* GS-15 (NC_007517) | fig\|269799.8.peg.1025 | Gmet_0990 CDS_1098034_1099452_+ (YP_383957.1) | Phosphoglucomutase (EC 5.4.2.2) @ Phosphomannomutase (EC 5.4.2.8) |
| | fig\|269799.8.peg.117 | Gmet_0104 CDS_129995_131824_+ (YP_383078.1) | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |
| | fig\|269799.8.peg.1194 | Gmet_1158 CDS_1284271_1285164_+ (YP_384121.1) | Putative TonB-dependent receptor |
| | fig\|269799.8.peg.1224 | Gmet_1188 CDS_1319834_1320817_+ (YP_384150.1) | N-acetylmuramoyl-L-alanine amidase |
| | fig\|269799.8.peg.1283 | Gmet_1245 CDS_1403629_1405539_+ (YP_384206.1) | Putative TonB-dependent receptor |
| | fig\|269799.8.peg.1415 | Gmet_1377 CDS_1539404_1541563_- (YP_384336.1) | Soluble lytic murein transglycosylase precursor (EC 3.2.1.-) |
| | fig\|269799.8.peg.1461 | Gmet_1425 CDS_1593423_1594667_+ (YP_384384.1) | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|269799.8.peg.148 | Gmet_0135 CDS_163535_164944_- (YP_383109.1) | Phosphoglucomutase (EC 5.4.2.2) @ Phosphomannomutase (EC 5.4.2.8) |
| | fig\|269799.8.peg.1527 | Gmet_1487 CDS_1680833_1682662_+ (YP_384446.1) | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |
| | fig\|269799.8.peg.1679 | Gmet_1640 CDS_1839853_1840944_+ (YP_384596.1) | 6-phosphofructokinase (EC 2.7.1.11) |
| | fig\|269799.8.peg.1708 | Gmet_1669 CDS_1874607_1876637_- (YP_384625.1) | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| *Geobacter metallireducens* GS-15 (NC_007517) | fig\|269799.8.peg.1929 | Gmet_1886 CDS_2100595_2101950_- (YP_384840.1) | Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|269799.8.peg.1998 | Gmet_1953 CDS_2179681_2180292_- (YP_384907.1) | Membrane-bound lytic murein transglycosylase C (EC 3.2.1.n1) |
| | fig\|269799.8.peg.2335 | Gmet_2294 CDS_2600934_2602004_- (YP_385244.1) | Endoglucanase |
| | fig\|269799.8.peg.2594 | Gmet_2556 CDS_2901389_2902888_- (YP_385500.1) | Membrane-bound lytic murein transglycosylase D |
| | fig\|269799.8.peg.3301 | Gmet_3262 CDS_3671554_3672327_- (YP_386200.1) | Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains) |
| | fig\|269799.8.peg.974 | Gmet_0938 CDS_1039789_1040748_- (YP_383905.1) | 6-phosphofructokinase (EC 2.7.1.11) |
| Bin78 | fig\|6666666.177606.peg.1028 | contig00359_114953_117109_+ | TonB-dependent receptor |
| | fig\|6666666.177606.peg.107 | contig00173_114082_115305_+ | Membrane-bound lytic murein transglycosylase A precursor (EC 3.2.1.-) |
| | fig\|6666666.177606.peg.1535 | contig00780_64767_62602_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.177606.peg.1621 | contig00815_72000_73358_+ | Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|6666666.177606.peg.1625 | contig00822_2936_4990_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.177606.peg.1839 | contig00929_17892_19208_+ | Beta-hexosaminidase (EC 3.2.1.52) |
| | fig\|6666666.177606.peg.2047 | contig01049_33830_32007_- | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |
| | fig\|6666666.177606.peg.2061 | contig01049_47688_48581_+ | TonB protein |
| | fig\|6666666.177606.peg.2073 | contig01078_3513_1210_- | TonB dependent receptor |

**Table C.9 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| Bin78 | fig\|6666666.177606.peg.215 | contig00197_16177_18117_+ | Soluble lytic murein transglycosylase precursor (EC 3.2.1.-) |
| | fig\|6666666.177606.peg.2184 | contig01130_48057_49064_+ | Beta N-acetyl-glucosaminidase (EC 3.2.1.52) |
| | fig\|6666666.177606.peg.2345 | contig01296_40943_42409_+ | Aminoacyl-histidine dipeptidase (Peptidase D) (EC 3.4.13.3) |
| | fig\|6666666.177606.peg.2529 | contig01554_8465_7206_- | 6-phosphofructokinase (EC 2.7.1.11) |
| | fig\|6666666.177606.peg.3098 | contig02822_8025_9095_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.177606.peg.310 | contig00197_116911_115886_- | Uncharacterized protein with LysM domain, COG1652 |
| | fig\|6666666.177606.peg.3221 | contig03312_17722_15746_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.177606.peg.329 | contig00197_136940_137647_+ | N-acetylmuramoyl-L-alanine amidase |
| | fig\|6666666.177606.peg.3594 | contig10661_26_1750_+ | TonB-dependent receptor |
| | fig\|6666666.177606.peg.3613 | contig17609_1754_3187_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.177606.peg.523 | contig00231_157293_158183_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.177606.peg.627 | contig00285_99904_98594_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.177606.peg.627 | contig00285_99904_98594_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.177606.peg.793 | contig00312_134961_133585_- | Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|6666666.177606.peg.867 | contig00345_73381_72761_- | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.177606.peg.877 | contig00345_81055_80438_- | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.177606.peg.980 | contig00359_43408_41540_- | 1,4-alpha-glucan (glycogen) branching enzyme, GH-13-type (EC 2.4.1.18) |

## C.1 Supplemental descriptions for additional metagenomic datasets in Chapter 4

In the chapter 4, ecological roles of the microbial community, selectively enriched in the DHS reactor for biological degradation of SMP, was revealed by providing the community structure and the functionality in both community and population levels using coupled metagenomic and metatranscriptomic approaches. To verify that the microbial community shift and the functional preservation between Phase III and Phase V were in a temporal continuity, additional samples from the upper part of the reactor at days 528 and 602 in Phase II, and 723 in Phase IV were collected (Figure 4.1). The additional samples were analyzed by as same methods as possible in the section 4.3. The detailed differences were described below.

## C.2 DNA extraction, library construction, and sequencing

The procedures for biomass collection and DNA extraction were followed as written in section 4.3.2. The concentrations of DNA in the samples were measured by a Nanodrop 1000 spectrophotometer, which were 100.2 ng/ml, 236.4 ng/ml, and 247.7 ng/ml for U528, U602, and U723, respectively. The integrity of the extracted DNA was verified by running 100ng of each sample with a DNA molecular weight marker (1kb DNA ladder, Promega) on a 1% denaturing formaldehyde agarose gel for electrophoresis prior to sequencing (Figure C.1). The extracted DNA samples were submitted to the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign (IL, USA) for sequencing and DNA and library construction. The DNA libraries were constructed for each sample using a Hyper Library construction kit (Kapa Biosystmes), and the pooled libraries were quantitated by qPCR and sequenced on one lane for 151 cycles from each end of the fragments on a HiSeq4000 sequencer (Illumina, San Diego, CA, USA) using a HiSeq 4000 sequencing kit version 1. The genomic libraries were generated and demultiplexed with the bcl2fastq v2.17.1.14 Conversion Software (Illumina, San Diego, CA, USA).

## C.3 Quality control, rRNA subtraction, and 16S rRNA gene reconstruction

The raw genomic reads were trimmed using a Q13 Phred quality score cutoff and screened with minimum length 50 bp cutoff using SolexaQA v3.1.7[1] for a quality control (QC) (Table C.10). The post QC genomic datasets were used to reconstruct full length of

16S rRNA using EMIRGE[2] with 0.99 OTU identity and default settings for the rest of conditions to reveal the microbial community compositions. The reconstructed genes for the three genomic datasets were combined and subjected to an operational taxonomic units (OTUs) assignment in QIIME,[3] and the phylogenetic affiliation of the OTUs was classified based on the Greengenes ARB database (Greengenes_16S_2011_1.arb) using ARB parsimony method and visualized in a phylogenetic tree as described in the section 4.3.4.1. The relative abundance of the representative sequences in each genomic dataset was expressed in percentage of the raw sequencing reads mapped to the representative sequences using Blastn with a cutoff of 95% identity and the parameters of X = 150, q = −1 and F = F at default settings.

**C.4 Metagenomic assembly and assembled genome bins**

The three post OC genomic dataset were subjected to be assembled together using MEGAHIT[4] with a range of k-mer sizes, 21-141 (Table C.10). The assembled contigs longer than 300 bp were submitted to the MG-RAST pipeline[5] and subjected to protein encoding genes (PEG) prediction (MG-RAST ID, 4740023.3 in the project, mgp 9993).[6] Taxonomic annotation was performed against the SEED database using a Best Hit Classification approach with a maximum e-value cutoff of 1E-5, a similarity cutoff of 60%, and a minimum alignment length of 15 measured in amino acids for protein and base pairs for RNA databases. Functional annotation was conducted by comparison to the subsystems using a hierarchical classification algorithm with a maximum e-value cutoff of 1E-5, a similarity cutoff of 60%, and a minimum alignment length of 15 amino acids. The PEGs longer than 300 bp were applied to the further expression analysis. The relative abundance of PEGs was estimated by following steps described in the section 4.3.4.4.

The assembled contigs greater than 1000 bases were subjected to cluster into genome bins, using MaxBin (v 2.2.1).[7] Assembled genome bins, the estimated completeness and contamination of which using CheckM (v 1.0.5)[8] were less than 20% and more than 10%, respectively, were discarded. AMPHORA2[9] was used to estimate the taxonomic affiliation of the assembled genome bins, and the resulted marker lineage was reported when 75% of the classifications reached a consensus taxonomic level.[10] A genome-wide phylogenetic analysis of the assembled genome bins was conducted using PhyloPhlAn.[11] The predicted

protein encoding genes for the assembled genome bins were identified and aligned on a subset of 400 conserved protein sequences. The assembled genome bins and reference genomes were integrated into the tree of life with 3,171 microbial genomes.

**C.5 Carbohydrate-active enzyme annotation**

The clustered contigs for the most abundant thirty four assembled genome bins (Table C.14) were subjected to gene prediction using FragGeneScan v1.30.[12] A carbohydrate-active enzyme (CAZy)-family specific hidden Markov model (HMMs) were downloaded from the dbCAN database (http://csbl.bmb.uga.edu/dbCAN/)[13] and used in screening amino acid sequences of the predicted ORFs for similarity to 385 families (13 auxiliary activity (AA), 81 carbohydrate-binding module (CBM), 16 carbohydrate esterase (CE), 145 glycoside hydrolase (GH), 103 glycosyl transferase (GT), and 27 polysaccharide lyase (PL) families) in the CAZy database.[14] The protein sequences were compared and sorted as described in the section 4.3.4.6 using hmmscan. The CAZy families which the related genes of the major bins belong to were plotted using the heatmap.2 function of the gplots package (v 3.0.1) in R. The relative abundances of the CAZy families were normalized as described in the section, 4.2.4.4.

**C.6 Microbial phylogenetic community structure in Phase II and Phase IV**

The additional samples collected were named U528 and U602 in Phase II and U723 in Phase IV to determine the temporal continuity between the samples collected in Phase III and V. To compare the phylogenetic community structures among those samples, the three microbial community samples were sequenced using Illumina, and the sequencing results provided paired-end 150 bp metagenomic reads with a range of fragment size 150 bp to 800 bp ($2.4 \times 10^8$ reads for U528, $2.4 \times 10^8$ reads for U602, $2.2 \times 10^8$ reads for U723) (Table C.10). The post QC genomic reads were blasted to the EMIRGE-based reconstructed 16S rRNA gene sequences.[2] The dominant bacterial EMIRGE-constructed representative sequences that indicated relative abundance >1% of the total number of 16s rRNA gene sequences in any sample were included in constructing a phylogenetic tree with closely related references (Figure C.9). In U528, *Acidithiobacillales*-related member in *Gammaproteobacteria* (DHS_Emg 28, 6.0%), *Saprospiraceae*-related member in

*Sphingobacteriales* (DHS_Emg 0, 5.5%), and *Cytophaga*-related members (DHS_Emg 33, 2.4%) were most abundant. Compared to U528, in U602, *Saprospiraceae*-related members (DHS_Emg 0, 5.8%, DHS_Emg 29, 3.0%, and DHS_Emg 19, 2.2%) became more abundant followed by *Cytophaga*-related members (DHS_Emg 33, 2.3%), whereas DHS_Emg 28 dramatically decreased to less than 0.2%. As previously indicated in a comparison between U648 and &798, in U723 a clear shift among the abundant *Saprospiraceae*-related members was observed; the abundance of all three *Saprospiraceae*-related members, DHS_Emg 0, DHS_Emg 29, and DHS_Emg 19 decreased to about 0.1% in U723. Instead, another *Saprospiraceae*-relative (DHS_Emg 22, 7.4%) became most abundant. In addition, *Dechloromonas*-related members (DHS_Emg 8, 2.9%) and *Geobacter*-related members (DHS_Emg 18, 5.8%) increased in U798. These population shifts which observed in U723 indicated transitional microbial community in the upper part of the DHS reactor between U648 and U798.

**Table C.10** Assembly statistic of metagenomic datasets.

| Assembly | Assembler | kmer size | Total trimmed reads[a] | | Assembled reads (95% similarity) | | Total contig size | Number of contig | Max contig size | N50 |
|---|---|---|---|---|---|---|---|---|---|---|
| TGTHR | MEGAHIT | 21 - 141 | U528 | 238,473,718 | 212,229,902 | 89% | 3,519,636,225 | 1,972,366 | 2,892,810 | 3,813 |
| | | | U602 | 238,432,590 | 217,195,520 | 91% | | | | |
| | | | U723 | 219,542,298 | 193,956,863 | 88% | | | | |

a. The raw genomic reads were trimmed using a Q13 Phred quality score cutoff and screened with minimum length 50 bp cutoff using SolexaQA v3.1.7.

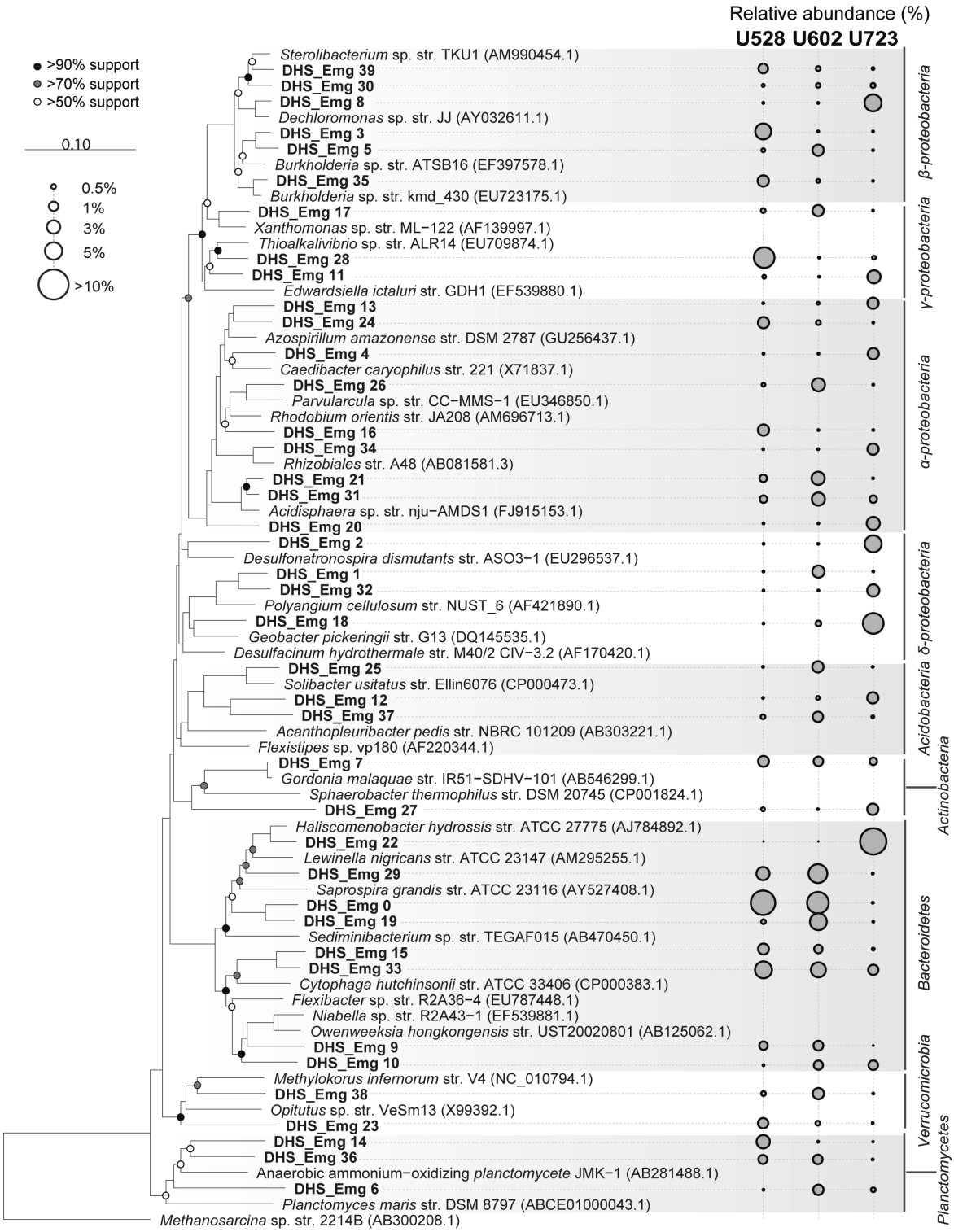



**Figure C.9** Microbial phylogenetic composition in Phase II (U528 and U602) and Phase IV (U723). In the 16S rRNA gene-based phylogenetic tree (bootstrap 1000: >90% black node, >70% gray node, and >50% white node), DHS_Emg refers to reconstructed ribosomal sequences using EMIRGE. The relative abundance is normalized to total number of bacterial 16s rRNA gene sequences in each dataset.

### C.7 Microbial global functionality and expressions in Phase II and Phase IV

The de novo assembly produced using MEGAHIT[4] included 89% of the 238 million reads in U528, 91% of the 238 million reads in U602, and 88% of the 220 million reads in U723 (Table C10). The assembly contained 1,972,366 contigs with a total sequence size of 3.5 Gb, a maximum contig size of 2.9 Mb and N50 of 3,813 bp with cutoff 300 bp. Using MG-RAST functional annotation, 3,404,512 ORFs were predicted, 1,950,961 ORFs of which were annotated with putative protein functions and 1,579,174 ORFs were assigned to a functional classification (Table C11). Among the annotated ORFs, 86.0% of features were classified as SEED Subsystems-based PEGs (Table C12). 85.2%, 82.4%, and 78.9% of the PEGs by Subsystems encoded in U528, U602, and U723 post QC datasets, respectively (Table C13). The relative abundance of the genomic encodes at the SEED Subsystem level 1 were exhibited (Figure C.10). Cluster-based subsystems (12.0-13.9%) and Carbohydrates (10.5-11.4%) were the two systems most abundantly encoded, followed by Amino acids and derivatives (9.4-10.7%) and Protein metabolism (7.8-8.3%). These subsystems indicated the constant metabolic categories and abundances as analyzed in the three datasets in Phase III and Phase V. U528 and U602 were closely clustered together rather than with U723, which indicated that change of the global functionality was likely subjected to temporal variation.

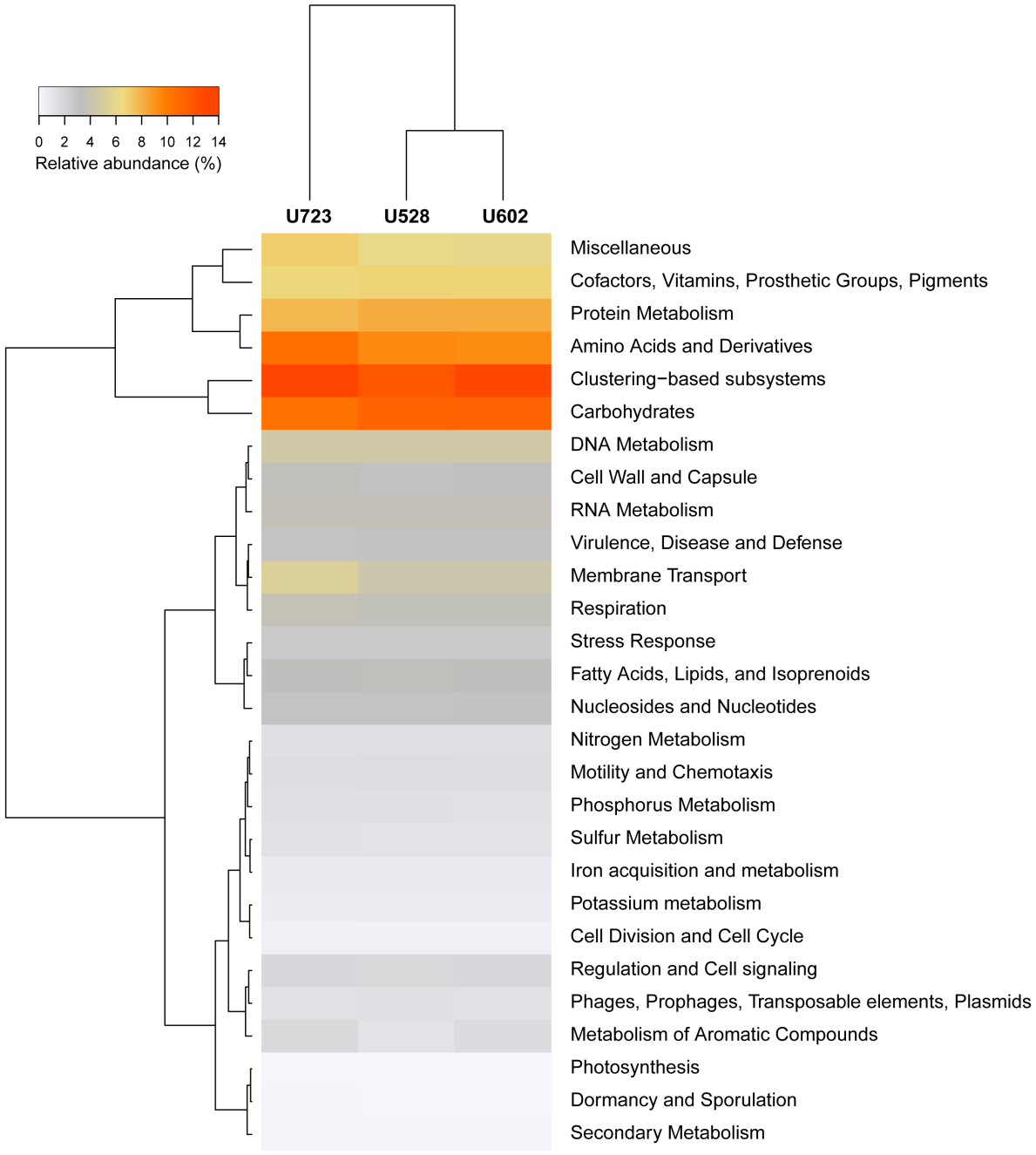



**Figure C.10** Global analysis of metabolic potential and functional activities in the DHS communities. Clustering of the three metagenomic and triplicated metatranscriptomic datasets based on normalized relative abundance of SEED subsystem level 1. Hierarchical clustering of the metagenomic and the metatranscriptomic datasets were separately conducted with Euclidean distance using R package (Stats v3.2.0).

**Table C.11** MG-RAST annotation of Assembly (contigs > 300 bp).

| Item | Statistics |
|---|---|
| Contigs | 1,972,366 |
| Average length (bp) | 1,784 ± 10,535 |
| Total length (bp) | 3,519,636,225 |
| Predicted ORFs | 3,404,512 |
| Annotated | 1,950,961 |
| rRNAs | 1,493 |
| Functional category | 1,579,174 |
| Unrecognized | 371,787 |

**Table C.12** Summary of protein encoding genes annotated by SEED Subsystems.

| Total number of protein encoding genes | | 1,677,780 |
|---|---|---|
| Summary of protein encoding genes (length) | Minimum | 47 |
| | 1st Quantile | 505 |
| | Median | 804 |
| | Mean | 1,174 |
| | 3rd Quantile | 1,404 |
| | Maximum | 71,960 |

**Table C.13** Protein encoding genes aligned with coding-DNA.

| Genomic sample | Number of aligned reads (95% similarity blasted to protein encoding genes among the trimmed reads) | | Present protein encoding genes | |
|---|---|---|---|---|
| U528 | 141,047,525 | 59% | 1,428,607 | 85.15% |
| U602 | 160,287,239 | 67% | 1,382,469 | 82.40% |
| U723 | 125,549,466 | 57% | 1,323,504 | 78.88% |

**C.8 Potential encoding of CAZy families in the assembled genome bins**

As a result of a metagenomic binning to reconstruct assembled genome bins, the assembled contigs were clustered into 244 bins with less than 10% contamination and more than 20% completeness.[10] 34 assembled genome bins that contributed top 50% of relative abundance of PEG in any dataset were subjected to construction of a genome-wide phylogenetic analysis with other 3,171 reference genomes using PhyloPhlAn[11] (Figure C.11 and Table C14). Additionally, their taxonomic affiliations were assigned using AMPHORA2[9] software with 31 conserved bacterial phylogenetic protein marker genes (Table C14). 12 bins were assigned in *Bacteroidetes*, six of which (BinN002, BinN013, BinN022, BinN024, BinN026, and BinN424) constituted a deep branch with *Haliscomenobacter hydrossis* DSM 1100 (IMG taxon ID: 2504756004). BinN013 and BinN040 constructed a branch with *Chitinophaga pinensis* UQM 2034 (IMG taxon ID: 644736340). As observed in the community structure (Figure C.9), a shift was observed in the relative abundance of the genomic encodes; BinN002 was the most abundant in Phase II while BinN008 and BinN024 became abundant in Phase IV. In *Proteobacteria*, 14 bins were classified affiliated: *alphaproteobacteria* (5), *betaproteobacteria* (6), *deltaproteobacteria* (2), and *gammaproteobacteria* (2). The relative abundance was shifted from BinN001 in *gammproteobacteria*, constituting a deep branch with *Alkalilimnicola ehrlichii* MLHE-1(IMG taxon ID: 637000324), in Phase II to BinN042 in *deltaproteobacteria* and BinN115 and BinN213 in *betaproteobacteria*. This shift may indicate a continuity to the abundance of the *Geobacter* and *Dechloromonas*-related assembled genome bins later in U798. The rest of them were affiliated with *Acidobacteria* (2), *Gemmatinomonadetes* (1), *Planctomycetes* (2), and *Verrucomicrobia* (2).

To further investigate how the functionally dominant microbial populations were involved in polysaccharide and glycan degradation, the 34 major assembled genomic bins were subjected to the carbohydrate-active enzyme analysis using the profile hidden Markov model specifying CAZy database. The normalized genomic abundances of each CAZy family, which were significantly abundant at the 98% confidence level, were shown (Figure C.12). As previously observed in the datasets in Phase III and Phase V, the assembled genomic bins affiliating *Bacteroidetes* indicated the most abundant genomic encodes in the predicted enzymes. The abundance among the *Bacteroidetes*-related genomic bins were

changed from the *Haliscomenobacter*-related bin, BinN002, in U528 and U602, to another *Haliscomenobacter*-related bin, BinN024, and *Chitinophaga*-related bin, BinN008, in U723. The most genomically predicted enzyme families by them were CBM families 32, 37, 40, 44, and 50, together with Cohesin and Dockerin, of which the glucan specific CBM family, CBM44, was most highly encoded. The most predicted glycoside hydrolytic GH families were endoglucanase (GH74), GalNAc hydrolase (GH109), oligo-alpha-glucosidase (GH13) and peptidoglycan lyase (GH23). The predicted glycoside hydrolytic GH families were mostly endoglucanase (GH74), GalNAc hydrolase (GH109), and peptidoglycan lyase (GH23). Carboxyl esterase enzyme families (CE1 and 10) were also highly encoded by the *Bacteroidetes*-related assembled genome bins. In spite of the temporal variance, the *Bacteroidetes*-related bins were equipped with the CAZy families involved in both binding modules to glucan and glycan substrates and following glycoside hydrolases and esterases. The gene inventory of the *Bacteroidetes*-related bins, further, showed that these bins were fully equipped with exo-enzymes and intracellular genes (chitinase, glucuronidase, hex, nagZ, nagK, murQ, nagA, and nagB), which were necessary to bind and degrade N-substituted oligosaccharides (Table C15). Laster increasing abundance of these CAZy families in *Geobacter* and *Dechloromonas*-related assembled genome bins were observed, but insignificant in U723.

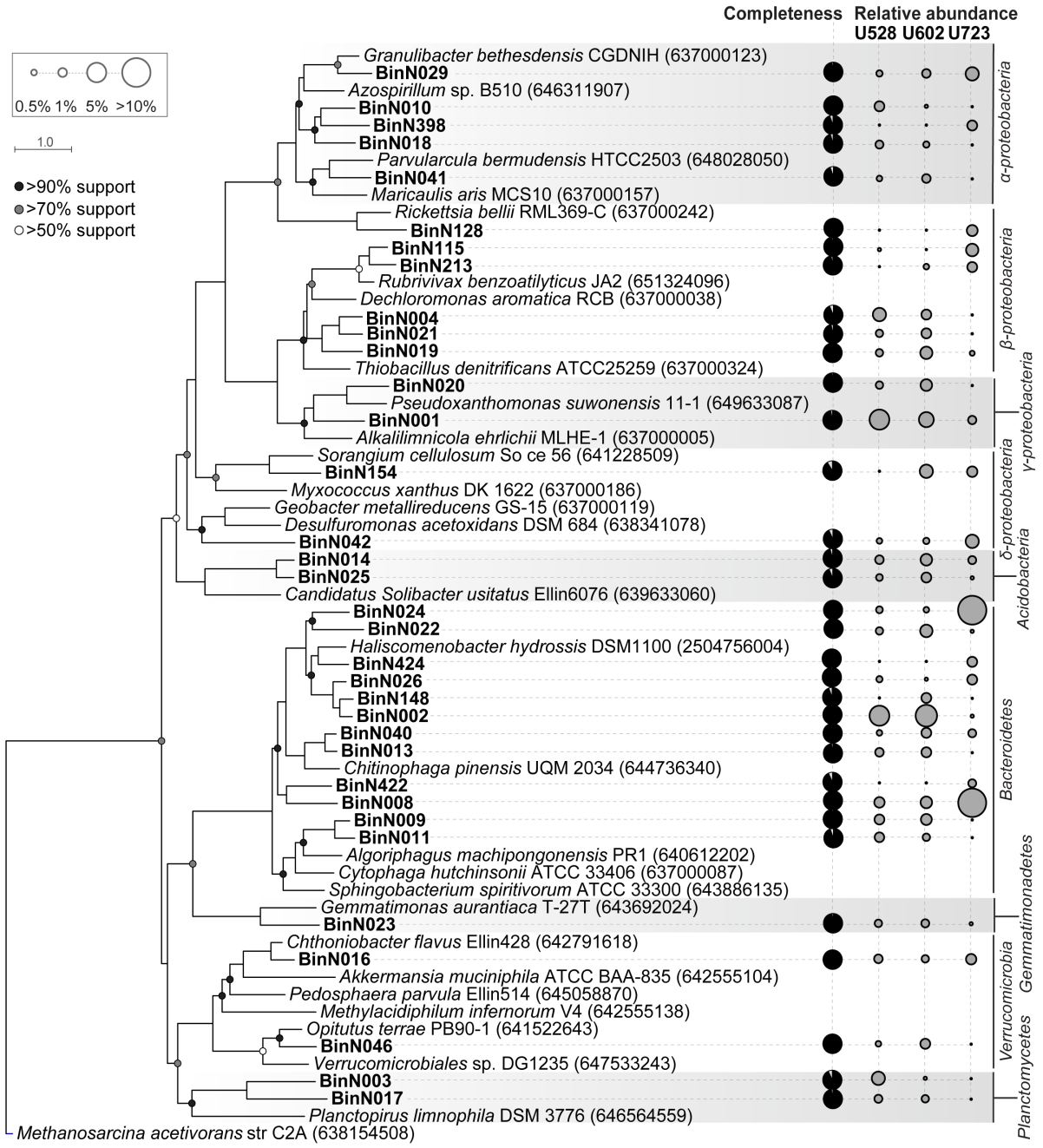



**Figure C.11** The genome-wide phylogenetic analysis and the abundance profile of the major assembled genome bins contributing cumulative top 50% of relative abundance for each dataset. The phylogenetic tree was generated by PhyloPhlAn and iTOL from predicted protein sequences of the major bins and 3,171 other reference genomes (bootstrap 1000: >90% black node, >70% gray node, and >50% white node; IMG taxon ID of the reference genomes in parenthesis.

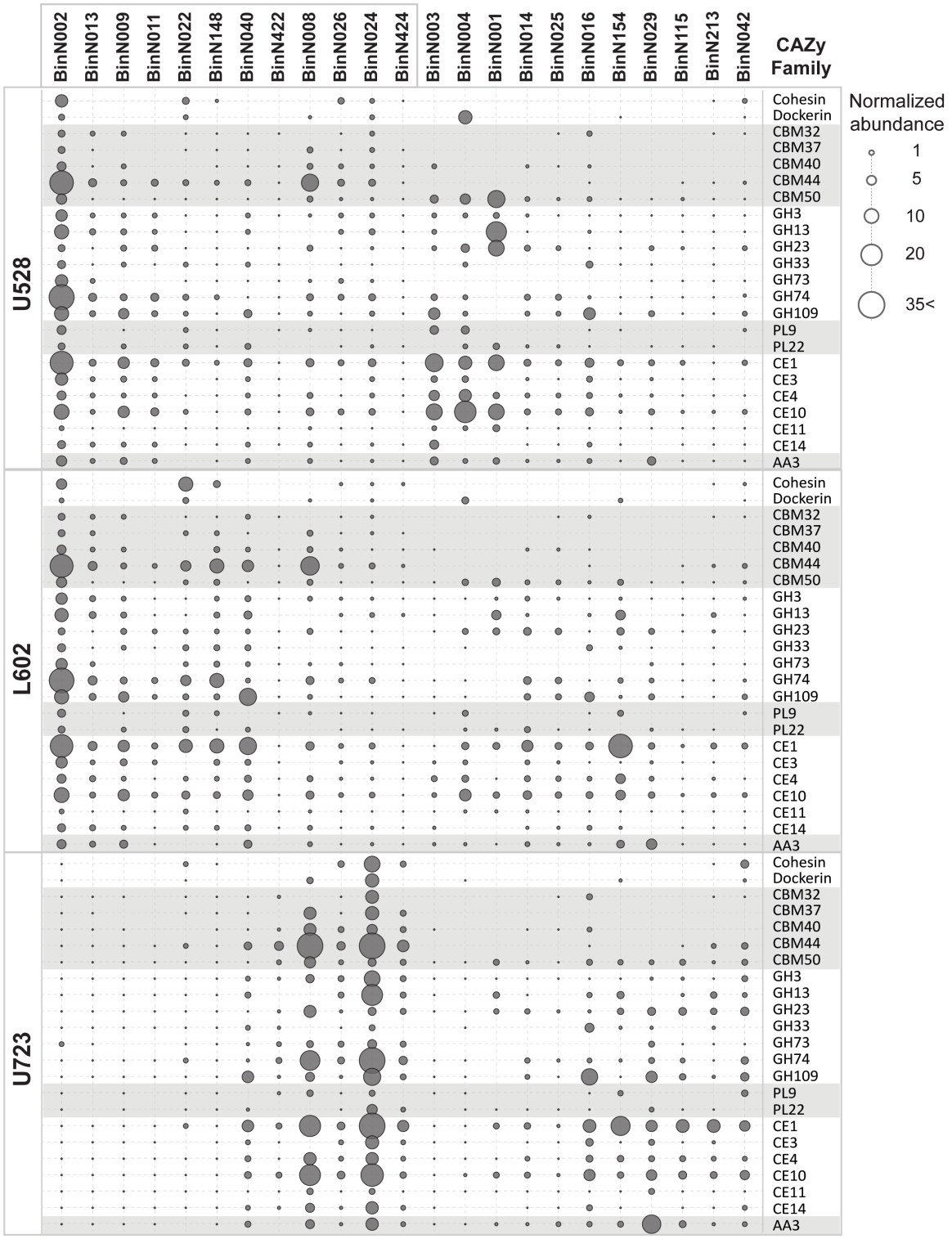



**Figure C.12** Potential encoding and expression of CAZy by the dominant assembled genome bins. The genomic normalized abundance of each CAZy family, which was significantly abundant at the 98% confidence level, were plotted with closed circles.

**Table C.14** Assembled genome bins that contribute top 50 % of abundance in each genomic dataset.

| Bin_Id | U528[a] | U602[a] | U723[a] | Marker lineage[c] | Marker gene copies 0 | 1 | 2 | 3 | Completeness [b] | Contamination[b] | Size (Mb) | GC (%) | Contig count |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BinN001 | 6.4 | 1.9 | 0.9 | *o_Chromatiales* | 8 | 266 | 1 | 0 | 97.1 | 0.1 | 2.6 | 63.4 | 1164 |
| BinN002 | 5.2 | 5.5 | 0.2 | *g_Haliscomenobacter* | 3 | 298 | 1 | 0 | 98.5 | 0.3 | 6.7 | 56.8 | 364 |
| BinN003 | 3.0 | 0.2 | 0.0 | *f_Planctomycetaceae* | 8 | 143 | 1 | 0 | 93.2 | 1.1 | 5.7 | 64.5 | 2921 |
| BinN004 | 3.3 | 1.3 | 0.0 | *o_Burkholderiales* | 21 | 403 | 1 | 0 | 93.4 | 0.1 | 3.9 | 58.6 | 223 |
| BinN008 | 1.1 | 1.4 | 6.7 | *g_Fluviicola* | 2 | 273 | 3 | 0 | 98.9 | 1.6 | 4.3 | 60.6 | 434 |
| BinN009 | 1.2 | 1.4 | 0.1 | *g_Cytophaga* | 6 | 441 | 7 | 0 | 98.5 | 0.7 | 4.5 | 41.7 | 47 |
| BinN010 | 1.1 | 0.2 | 0.0 | *c_Alphaproteobacteria* | 0 | 349 | 0 | 0 | 100.0 | 0.0 | 3.0 | 56.9 | 564 |
| BinN011 | 1.0 | 0.5 | 0.0 | *g_Cytophaga* | 25 | 404 | 23 | 2 | 94.5 | 6.5 | 3.9 | 49.1 | 47 |
| BinN013 | 0.8 | 1.1 | 0.0 | *g_Chitinophaga* | 6 | 295 | 1 | 0 | 97.5 | 0.5 | 4.7 | 43.8 | 364 |
| BinN014 | 0.6 | 1.4 | 0.5 | *o_Nitrospirales* | 5 | 179 | 4 | 0 | 95.7 | 2.0 | 3.8 | 53.2 | 2258 |
| BinN016 | 1.1 | 0.9 | 2.1 | *o_Verrucomicrobiales* | 1 | 208 | 19 | 1 | 99.3 | 6.3 | 6.3 | 62.6 | 88 |
| BinN017 | 0.5 | 0.5 | 0.0 | *f_Gemmatimonadaceae* | 2 | 139 | 2 | 0 | 97.7 | 2.3 | 4.9 | 70.7 | 2921 |
| BinN018 | 1.5 | 1.0 | 0.1 | *o_Rhodospirillales* | 14 | 301 | 20 | 1 | 95.7 | 5.8 | 7.1 | 67.8 | 63 |
| BinN019 | 0.7 | 1.8 | 0.3 | *o_Burkholderiales* | 2 | 417 | 5 | 1 | 99.1 | 1.6 | 4.1 | 69.0 | 223 |
| BinN020 | 0.5 | 1.4 | 0.0 | *f_Xanthomonadaceae* | 11 | 637 | 11 | 0 | 98.3 | 1.4 | 4.8 | 62.2 | 55 |
| BinN021 | 0.6 | 1.2 | 0.1 | *c_Betaproteobacteria* | 10 | 401 | 14 | 0 | 96.4 | 3.8 | 3.8 | 61.5 | 223 |
| BinN022 | 0.4 | 1.4 | 0.1 | *g_Haliscomenobacter* | 2 | 299 | 1 | 0 | 99.0 | 0.3 | 4.2 | 36.2 | 364 |
| BinN023 | 0.7 | 0.8 | 0.2 | *f_Gemmatimonadaceae* | 2 | 140 | 5 | 0 | 97.8 | 4.8 | 5.8 | 69.6 | 2993 |
| BinN024 | 0.6 | 0.4 | 7.6 | *g_Haliscomenobacter* | 1 | 295 | 6 | 0 | 99.5 | 1.4 | 5.8 | 36.3 | 364 |
| BinN025 | 0.4 | 0.8 | 0.2 | *c_Solibacteres* | 8 | 174 | 6 | 0 | 94.0 | 5.1 | 4.0 | 51.5 | 2258 |
| BinN026 | 0.5 | 0.2 | 0.9 | *g_Haliscomenobacter* | 3 | 298 | 1 | 0 | 98.5 | 0.5 | 5.7 | 55.1 | 364 |
| BinN029 | 0.8 | 1.3 | 3.7 | *o_Rhodospirillales* | 27 | 286 | 22 | 1 | 91.7 | 6.8 | 5.0 | 66.2 | 63 |
| BinN040 | 0.5 | 2.0 | 1.0 | *g_Chitinophaga* | 1 | 297 | 2 | 1 | 99.5 | 1.5 | 6.7 | 42.0 | 364 |
| BinN041 | 0.3 | 0.6 | 0.1 | *c_Alphaproteobacteria* | 60 | 326 | 2 | 0 | 94.6 | 0.8 | 3.1 | 62.6 | 468 |
| BinN042 | 0.4 | 0.5 | 2.2 | *o_Desulfuromonadales* | 11 | 233 | 3 | 0 | 93.3 | 1.9 | 6.0 | 64.4 | 83 |
| BinN046 | 0.3 | 0.9 | 0.0 | *o_Verrucomicrobiales* | 1 | 228 | 1 | 0 | 99.3 | 0.7 | 3.7 | 66.2 | 88 |
| BinN115 | 0.2 | 0.2 | 2.4 | *o_Burkholderiales* | 8 | 395 | 22 | 1 | 98.0 | 5.6 | 6.3 | 68.5 | 193 |
| BinN128 | 0.0 | 0.0 | 0.6 | *o_Rickettsiales* | 3 | 306 | 11 | 3 | 98.6 | 8.3 | 1.7 | 33.8 | 83 |
| BinN148 | 0.1 | 1.4 | 0.0 | *g_Haliscomenobacter* | 15 | 274 | 13 | 0 | 94.2 | 4.3 | 6.4 | 48.0 | 364 |
| BinN154 | 0.1 | 3.6 | 2.2 | *o_Myxococcales* | 15 | 227 | 4 | 1 | 91.2 | 2.8 | 9.4 | 69.9 | 83 |
| BinN213 | 0.1 | 0.5 | 1.3 | *o_Burkholderiales* | 53 | 369 | 5 | 0 | 97.2 | 2.3 | 5.3 | 68.4 | 193 |
| BinN398 | 0.0 | 0.0 | 1.1 | *o_Rhizobiales* | 26 | 299 | 24 | 0 | 92.7 | 6.4 | 3.0 | 64.0 | 564 |
| BinN422 | 0.0 | 0.0 | 0.5 | *g_Fluviicola* | 16 | 298 | 2 | 0 | 93.3 | 1.0 | 3.6 | 33.9 | 350 |
| BinN424 | 0.0 | 0.0 | 1.2 | *g_Haliscomenobacter* | 2 | 285 | 10 | 1 | 99.0 | 4.7 | 6.1 | 52.5 | 364 |

**Table C.14 (cont.)**

a. Normalized abundance of genomic datasets aligned to the protein-coding genes. The bins contributing top 50% of abundance for each dataset were listed in the table.
b. Relative abundance ratio (%), defined by the actual coverage levels divided by summed coverage levels of all assembled genome bins that are retrieved from the results using MaxBin (v.2.2.1). The assembled genome bins that contribute cumulative top 50% of relative abundance were listed.
c. Completeness and contamination of the assembled genome bins were assessed using CheckM. Bins that were less than 90% complete or with greater than 10% contamination were discarded.
d. Marker lineage was analyzed using AMPHORA2 and reported if 75% of the classifications were in agreement at a particular taxonomic level.
e. The listed bins were submitted under the MG-RAST project (ID: mgp9993).

**Table C.15** Gene inventory analysis related to N-substituted biomass structural detritus utilization in the *Bacteroidetes*-related genome bins.

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| BinN002 | fig\|6666666.273398.peg.103 | contig1035842_8895_7054_- | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |
| | fig\|6666666.273398.peg.1076 | contig1349688_118669_121344_+ | TonB-dependent receptor |
| | fig\|6666666.273398.peg.108 | contig1035842_13474_11981_- | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.273398.peg.12 | contig1032987_15020_17965_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.273398.peg.1203 | contig1383735_90043_91290_+ | N-acetylglucosamine related transporter, NagX |
| | fig\|6666666.273398.peg.1204 | contig1383735_92895_91402_- | SusD, outer membrane protein |
| | fig\|6666666.273398.peg.1205 | contig1383735_96015_93022_- | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.273398.peg.1212 | contig1383735_106999_104462_- | TonB-dependent receptor |
| | fig\|6666666.273398.peg.1219 | contig1383735_113438_111045_- | TonB-dependent receptor |
| | fig\|6666666.273398.peg.127 | contig1035842_36180_33691_- | TonB-dependent receptor, putative |
| | fig\|6666666.273398.peg.1571 | contig176149_132715_131336_- | Phosphomannomutase (EC 5.4.2.8) / Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|6666666.273398.peg.1585 | contig176149_151447_150149_- | N-acylglucosamine 2-epimerase (EC 5.1.3.8) |
| | fig\|6666666.273398.peg.1590 | contig176149_157374_160664_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.273398.peg.1616 | contig176149_191066_192037_+ | N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) |
| | fig\|6666666.273398.peg.1643 | contig176149_228293_237829_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.273398.peg.1848 | contig1871866_43359_41020_- | TonB-dependent receptor, plug precursor |
| | fig\|6666666.273398.peg.1918 | contig1871866_140886_139576_- | N-acetylglucosaminyltransferase (EC 2.4.1.-) |
| | fig\|6666666.273398.peg.192 | contig1035842_106495_109506_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.273398.peg.193 | contig1035842_109630_110964_+ | SusD/RagB family protein |
| | fig\|6666666.273398.peg.1939 | contig1871866_163665_160750_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.273398.peg.1958 | contig1871866_189217_187796_- | N-acetylglucosamine deacetylase (EC 3.5.1.-) / 3-hydroxyacyl-[acyl-carrier-protein] dehydratase, FabZ form (EC 4.2.1.59) |
| | fig\|6666666.273398.peg.1996 | contig1888521_25362_22282_- | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.273398.peg.2021 | contig1903360_27132_25714_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.273398.peg.208 | contig1041127_15712_13019_- | TonB-dependent receptor, putative |
| | fig\|6666666.273398.peg.2150 | contig1937582_61602_60550_- | Membrane-bound lytic murein transglycosylase A precursor (EC 3.2.1.-) |
| | fig\|6666666.273398.peg.2475 | contig211462_101956_100382_- | SusD, outer membrane protein |
| | fig\|6666666.273398.peg.2476 | contig211462_105130_102098_- | SusC, outer membrane protein involved in starch binding |

**Table C.15 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| BinN002 | fig\|6666666.273398.peg.2515 | contig211462_150397_147224_- | chitinase II |
| | fig\|6666666.273398.peg.258 | contig1046462_47961_51509_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.273398.peg.2581 | contig211462_271645_272184_+ | Phospholipid-lipopolysaccharide ABC transporter |
| | fig\|6666666.273398.peg.2812 | contig233572_105312_107666_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.273398.peg.2898 | contig233572_218617_217778_- | N-acetylmuramic acid 6-phosphate etherase |
| | fig\|6666666.273398.peg.2947 | contig245956_19602_18817_- | D-alanyl-D-alanine carboxypeptidase |
| | fig\|6666666.273398.peg.2962 | contig249485_18315_20753_+ | TonB-dependent receptor, putative |
| | fig\|6666666.273398.peg.2972 | contig249485_35622_38432_+ | TonB-dependent receptor |
| | fig\|6666666.273398.peg.3066 | contig268405_139855_141075_+ | N-acetyl-L,L-diaminopimelate deacetylase (EC 3.5.1.47) |
| | fig\|6666666.273398.peg.3067 | contig268405_147419_141132_- | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.273398.peg.3081 | contig268405_160961_163420_+ | TonB-dependent receptor, putative |
| | fig\|6666666.273398.peg.3144 | contig283311_25868_23640_- | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.273398.peg.34 | contig1034615_6503_9550_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.273398.peg.3415 | contig394373_60481_59360_- | N-acetylglucosamine related transporter, NagX |
| | fig\|6666666.273398.peg.3534 | contig484478_14503_12296_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.273398.peg.3641 | contig530325_34960_32105_- | putative TonB-dependent receptor |
| | fig\|6666666.273398.peg.3862 | contig591739_37050_38990_+ | Glucosamine-6-phosphate deaminase (EC 3.5.99.6) |
| | fig\|6666666.273398.peg.3922 | contig618580_49300_48455_- | D-alanyl-D-alanine dipeptidase (EC 3.4.13.-) |
| | fig\|6666666.273398.peg.4059 | contig646410_72600_76646_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.273398.peg.4060 | contig646410_76760_81199_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.273398.peg.4065 | contig646410_89929_94818_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.273398.peg.4199 | contig65279_58406_59905_+ | Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|6666666.273398.peg.421 | contig108488_105709_108813_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.273398.peg.425 | contig108488_113214_114122_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.273398.peg.4267 | contig668715_86168_83682_- | TonB-dependent receptor, putative |
| | fig\|6666666.273398.peg.4296 | contig686991_33618_31975_- | putative TonB-dependent receptor |
| | fig\|6666666.273398.peg.4331 | contig687193_1652_7555_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.273398.peg.4333 | contig687193_10539_17465_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.273398.peg.4337 | contig687193_20997_23525_+ | TonB-dependent receptor |
| | fig\|6666666.273398.peg.4347 | contig687193_37105_33806_- | TonB-dependent receptor |
| | fig\|6666666.273398.peg.4518 | contig735952_27979_28962_+ | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |

**Table C.15 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| BinN002 | fig\|6666666.273398.peg.4539 | contig735952_53966_55129_+ | N-acetylhexosamine 1-kinase |
| | fig\|6666666.273398.peg.4661 | contig735952_200253_207476_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.273398.peg.474 | contig108488_164942_166138_+ | Anhydro-N-acetylmuramic acid kinase (EC 2.7.1.-) |
| | fig\|6666666.273398.peg.4851 | contig814186_55849_54512_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.273398.peg.4920 | contig823276_36704_33837_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.273398.peg.5199 | contig877855_169662_167467_- | TonB-dependent siderophore receptor |
| | fig\|6666666.273398.peg.5205 | contig877855_178219_181014_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.273398.peg.5211 | contig877855_187453_186239_- | N-acetylglucosaminyltransferase (EC 2.4.1.-) |
| | fig\|6666666.273398.peg.5360 | contig999578_9709_11049_+ | D-amino acid dehydrogenase small subunit (EC 1.4.99.1) |
| | fig\|6666666.273398.peg.544 | contig109545_9499_7394_- | Probable tonB-dependent receptor yncD precursor |
| | fig\|6666666.273398.peg.606 | contig109545_79743_81800_+ | TonB-dependent receptor |
| | fig\|6666666.273398.peg.679 | contig1138890_29349_28147_- | Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13) |
| | fig\|6666666.273398.peg.72 | contig103538_11954_9390_- | TonB-dependent receptor |
| | fig\|6666666.273398.peg.804 | contig1186173_97696_100362_+ | TonB-dependent receptor |
| | fig\|6666666.273398.peg.823 | contig1186173_124841_122439_- | TonB-dependent receptor, putative |
| | fig\|6666666.273398.peg.836 | contig1202278_6429_5155_- | N-acetyl glucosamine transporter, NagP |
| | fig\|6666666.273398.peg.923 | contig1330001_36789_39995_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.273398.peg.971 | contig1349688_17674_15443_- | TonB-dependent receptor |
| BinN008 | fig\|6666666.273400.peg.1273 | contig33798_93301_94035_+ | TonB family protein |
| | fig\|6666666.273400.peg.1323 | contig33798_146272_147963_+ | putative TonB-dependent receptor |
| | fig\|6666666.273400.peg.1341 | contig33798_166718_166957_+ | Putative peptidoglycan binding domain 1 |
| | fig\|6666666.273400.peg.1458 | contig446840_8844_9671_+ | Glutamate racemase (EC 5.1.1.3) |
| | fig\|6666666.273400.peg.169 | contig125349_17128_15728_- | Phosphoglucosamine mutase (EC 5.4.2.10) |
| | fig\|6666666.273400.peg.174 | contig125349_20029_21105_+ | Anhydro-N-acetylmuramic acid kinase (EC 2.7.1.-) |
| | fig\|6666666.273400.peg.1766 | contig501300_101050_102189_+ | TonB family protein |
| | fig\|6666666.273400.peg.1797 | contig501300_130902_129058_- | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |
| | fig\|6666666.273400.peg.198 | contig125349_43600_41177_- | TonB-dependent receptor, plug precursor |
| | fig\|6666666.273400.peg.1995 | contig566071_23008_20789_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |

**Table C.15 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| BinN008 | fig\|6666666.273400.peg.224 | contig125349_75255_77549_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.273400.peg.2294 | contig57381_92081_90567_- | N-acetylgalactosamine 6-sulfatase |
| | fig\|6666666.273400.peg.2380 | contig585643_61379_62476_+ | N-acetylglucosaminyltransferase (EC 2.4.1.-) |
| | fig\|6666666.273400.peg.2468 | contig615680_43075_40496_- | TonB-dependent receptor |
| | fig\|6666666.273400.peg.2491 | contig615680_67960_66974_- | Glucose-6-phosphate isomerase, archaeal II (EC 5.3.1.9) / Mannose-6-phosphate isomerase, archaeal (EC 5.3.1.8) |
| | fig\|6666666.273400.peg.2497 | contig615680_75048_73582_- | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.273400.peg.2622 | contig625043_14111_17206_+ | TonB-dependent receptor |
| | fig\|6666666.273400.peg.2656 | contig69057_30654_29839_- | N-acetylmuramic acid 6-phosphate etherase |
| | fig\|6666666.273400.peg.2873 | contig720972_48096_45718_- | TonB-dependent receptor, putative |
| | fig\|6666666.273400.peg.3004 | contig777453_47453_46137_- | N-acetyl glucosamine transporter, NagP |
| | fig\|6666666.273400.peg.3089 | contig840221_36786_39263_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.273400.peg.3108 | contig840221_60269_61738_+ | N-acetylglucosamine deacetylase (EC 3.5.1.-) / 3-hydroxyacyl-[acyl-carrier-protein] dehydratase, FabZ form (EC 4.2.1.59) |
| | fig\|6666666.273400.peg.3111 | contig840221_63162_67325_+ | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.273400.peg.3214 | contig94918_37882_37070_- | N-acetylmannosaminyltransferase (EC 2.4.1.187) |
| | fig\|6666666.273400.peg.3286 | contig94918_110583_111623_+ | L-alanine-DL-glutamate epimerase |
| | fig\|6666666.273400.peg.3384 | contig94918_232694_235135_+ | TonB-dependent receptor |
| | fig\|6666666.273400.peg.350 | contig153540_35147_33759_- | Chitinase (EC 3.2.1.14) |
| | fig\|6666666.273400.peg.3571 | contig972386_127409_126204_- | N-acetyl-L,L-diaminopimelate deacetylase (EC 3.5.1.47) |
| | fig\|6666666.273400.peg.382 | contig157047_36908_38104_+ | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.273400.peg.583 | contig1790038_26596_25640_- | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.273400.peg.723 | contig1985150_7435_8682_+ | Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) |
| | fig\|6666666.273400.peg.770 | contig222786_26390_23628_- | TonB-dependent receptor |
| | fig\|6666666.273400.peg.926 | contig24165_31057_28730_- | TonB-dependent receptor |
| | fig\|6666666.273400.peg.93 | contig1234741_22762_24267_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.273400.peg.97 | contig1234741_29342_29091_- | TonB family protein |
| BinN024 | fig\|6666666.273401.peg.108 | contig1093387_8181_9665_+ | Chitin binding protein |
| | fig\|6666666.273401.peg.1148 | contig1467467_146203_147396_+ | Periplasmic septal ring factor with murein hydrolase activity EnvC/YibP |

**Table C.15 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| BinN024 | fig\|6666666.273401.peg.1186 | contig1467467_185372_186565_+ | Anhydro-N-acetylmuramic acid kinase (EC 2.7.1.-) |
| | fig\|6666666.273401.peg.1269 | contig1510968_33543_36206_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.273401.peg.1331 | contig1541227_19711_21471_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.273401.peg.1509 | contig1554492_85347_88487_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.273401.peg.1512 | contig1554492_93155_97246_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.273401.peg.1536 | contig1595497_18036_20294_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.273401.peg.1560 | contig1595497_56486_53319_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.273401.peg.1601 | contig1595497_105205_103751_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.273401.peg.1618 | contig1595497_129696_126874_- | TonB-dependent receptor, putative |
| | fig\|6666666.273401.peg.1628 | contig1595497_140735_139512_- | N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) |
| | fig\|6666666.273401.peg.1631 | contig1595497_147463_145454_- | TonB-dependent receptor |
| | fig\|6666666.273401.peg.167 | contig1127358_12599_9957_- | TonB-dependent receptor |
| | fig\|6666666.273401.peg.1753 | contig1683636_83063_85543_+ | TonB-dependent receptor, putative |
| | fig\|6666666.273401.peg.1830 | contig1739719_27403_24779_- | TonB-dependent receptor |
| | fig\|6666666.273401.peg.2035 | contig220997_15913_19026_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.273401.peg.2052 | contig220997_39410_38121_- | SusD/RagB family protein |
| | fig\|6666666.273401.peg.2089 | contig220997_103610_100434_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.273401.peg.2094 | contig220997_109073_106839_- | TonB-dependent receptor |
| | fig\|6666666.273401.peg.2126 | contig220997_138453_136141_- | TonB-dependent receptor, plug precursor |
| | fig\|6666666.273401.peg.217 | contig1136211_12168_13268_+ | D-galactose 1-dehydrogenase (EC 1.1.1.48) |
| | fig\|6666666.273401.peg.2259 | contig30784_58766_59518_+ | D-alanyl-D-alanine dipeptidase (EC 3.4.13.-) |
| | fig\|6666666.273401.peg.237 | contig116609_13942_12140_- | TonB-dependent receptor |
| | fig\|6666666.273401.peg.2390 | contig30784_216416_218431_+ | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.273401.peg.2430 | contig30784_262191_264488_+ | TonB-dependent receptor, putative |
| | fig\|6666666.273401.peg.2467 | contig309883_13005_15938_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.273401.peg.2505 | contig309883_63607_66588_+ | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.273401.peg.2578 | contig332824_68611_71202_+ | TonB-dependent receptor |
| | fig\|6666666.273401.peg.2631 | contig332824_149097_146893_- | SusD, outer membrane protein |
| | fig\|6666666.273401.peg.2632 | contig332824_152087_149124_- | SusC, outer membrane protein involved in starch binding |

**Table C.15 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| BinN024 | fig\|6666666.273401.peg.3067 | contig365579_121910_123310_+ | N-acetylglucosamine deacetylase (EC 3.5.1.-) / 3-hydroxyacyl-[acyl-carrier-protein] dehydratase, FabZ form (EC 4.2.1.59) |
| | fig\|6666666.273401.peg.3174 | contig407875_28137_25726_- | TonB-dependent receptor plug domain protein |
| | fig\|6666666.273401.peg.3184 | contig407875_34293_36515_+ | TonB-dependent outer membrane receptor |
| | fig\|6666666.273401.peg.3324 | contig449887_19914_22334_+ | TonB-dependent receptor |
| | fig\|6666666.273401.peg.3379 | contig49397_25101_22732_- | TonB-dependent receptor |
| | fig\|6666666.273401.peg.3387 | contig49397_33157_36084_+ | SusC, outer membrane protein involved in starch binding |
| | fig\|6666666.273401.peg.3405 | contig49397_63779_61833_- | Glucosamine-6-phosphate deaminase (EC 3.5.99.6) |
| | fig\|6666666.273401.peg.3439 | contig49397_105905_104067_- | Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) |
| | fig\|6666666.273401.peg.3477 | contig49397_141166_142017_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.273401.peg.3553 | contig49397_232124_229392_- | TonB-dependent receptor, plug precursor |
| | fig\|6666666.273401.peg.357 | contig1206644_33706_30542_- | TonB family protein / TonB-dependent receptor |
| | fig\|6666666.273401.peg.3744 | contig496638_193323_191665_- | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.273401.peg.391 | contig1206644_81786_81334_- | TonB-dependent receptor, putative |
| | fig\|6666666.273401.peg.392 | contig1206644_83746_81749_- | TonB-dependent receptor, putative |
| | fig\|6666666.273401.peg.3948 | contig61346_17316_18437_+ | N-acetylglucosaminyltransferase (EC 2.4.1.-) |
| | fig\|6666666.273401.peg.3979 | contig61346_57976_60777_+ | TonB-dependent receptor |
| | fig\|6666666.273401.peg.4004 | contig636711_12587_13798_+ | Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-) |
| | fig\|6666666.273401.peg.403 | contig1206644_93891_95090_+ | N-acetyl-L,L-diaminopimelate deacetylase (EC 3.5.1.47) |
| | fig\|6666666.273401.peg.4218 | contig653883_157985_155220_- | TonB-dependent receptor |
| | fig\|6666666.273401.peg.4312 | contig700315_9451_7199_- | TonB-dependent receptor, putative |
| | fig\|6666666.273401.peg.4630 | contig892113_75670_76740_+ | Protein often near L-alanine-DL-glutamate epimerase (cell wall recycling) |
| | fig\|6666666.273401.peg.4631 | contig892113_76737_77810_+ | L-alanine-DL-glutamate epimerase |
| | fig\|6666666.273401.peg.4649 | contig912080_10758_8551_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.273401.peg.497 | contig125023_45829_46917_+ | N-acetylglucosamine related transporter, NagX |
| | fig\|6666666.273401.peg.516 | contig125023_72194_72946_+ | N-acetylmuramic acid 6-phosphate etherase |
| | fig\|6666666.273401.peg.528 | contig1321850_5938_3503_- | TonB-dependent receptor |
| | fig\|6666666.273401.peg.614 | contig1321850_105993_107204_+ | D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) |
| | fig\|6666666.273401.peg.717 | contig1321850_199521_200498_+ | N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) |

**Table C.15 (cont.)**

| Genome | Feature ID | Protein locus tag (accession) | Functional role |
|---|---|---|---|
| BinN024 | fig\|6666666.273401.peg.742 | contig1321850_234408_235646_+ | D-amino acid dehydrogenase small subunit (EC 1.4.99.1) |
| | fig\|6666666.273401.peg.832 | contig1404041_106577_104205_- | TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins |
| | fig\|6666666.273401.peg.985 | contig1404041_267853_266489_- | Phosphomannomutase (EC 5.4.2.8) / Phosphoglucosamine mutase (EC 5.4.2.10) |

## C.9 References


1. Cox, M.; Peterson, D.; Biggs, P., SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **2010,** *11* (1), 485.

2. Miller, C. S.; Baker, B. J.; Thomas, B. C.; Singer, S. W.; Banfield, J. F., EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biology* **2011,** *12* (5).

3. Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Pena, A. G.; Goodrich, J. K.; Gordon, J. I.; Huttley, G. A.; Kelley, S. T.; Knights, D.; Koenig, J. E.; Ley, R. E.; Lozupone, C. A.; McDonald, D.; Muegge, B. D.; Pirrung, M.; Reeder, J.; Sevinsky, J. R.; Tumbaugh, P. J.; Walters, W. A.; Widmann, J.; Yatsunenko, T.; Zaneveld, J.; Knight, R., QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **2010,** *7* (5), 335-336.

4. Li, D.; Liu, C.-M.; Luo, R.; Sadakane, K.; Lam, T.-W., MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **2015,** *31* (10), 1674-1676.

5. Meyer, F.; Paarmann, D.; D'Souza, M.; Olson, R.; Glass, E. M.; Kubal, M.; Paczian, T.; Rodriguez, A.; Stevens, R.; Wilke, A.; Wilkening, J.; Edwards, R. A., The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *Bmc Bioinformatics* **2008,** *9*.

6. Wilke, A.; Bischof, J.; Harrison, T.; Brettin, T.; D'Souza, M.; Gerlach, W.; Matthews, H.; Paczian, T.; Wilkening, J.; Glass, E. M., A RESTful API for accessing microbial community data for MG-RAST. *Plos Comput Biol* **2015,** *11* (1), e1004008.

7. Wu, Y.-W.; Tang, Y.-H.; Tringe, S. G.; Simmons, B. A.; Singer, S. W., MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2014,** *2*, 26-26.

8. Parks, D. H.; Imelfort, M.; Skennerton, C. T.; Hugenholtz, P.; Tyson, G. W., CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **2015**.

9. Wu, M.; Scott, A. J., Phylogenomic Analysis of Bacterial and Archaeal Sequences with AMPHORA2. *Bioinformatics* **2012**.

10. Hultman, J.; Waldrop, M. P.; Mackelprang, R.; David, M. M.; McFarland, J.; Blazewicz, S. J.; Harden, J.; Turetsky, M. R.; McGuire, A. D.; Shah, M. B., Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* **2015**.

11. Segata, N.; Börnigen, D.; Morgan, X. C.; Huttenhower, C., PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature communications* **2013,** *4*.

12. Rho, M. N.; Tang, H. X.; Ye, Y. Z., FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Research* **2010,** *38* (20).

13. Yin, Y. B.; Mao, X. Z.; Yang, J. C.; Chen, X.; Mao, F. L.; Xu, Y., dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **2012,** *40* (W1), W445-W451.

14. Lombard, V.; Ramulu, H. G.; Drula, E.; Coutinho, P. M.; Henrissat, B., The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **2014,** *42* (D1), D490-D495.