Extracellular processes in wastewater treatment

By:

Reihaneh Bashiri

A thesis submitted to the Newcastle University for the degree of Doctor of Philosophy

School of Engineering

Newcastle University

United Kingdom

October 2021

Abstract

One of the limitations of low-temperature anaerobic treatment of domestic wastewater is poor lipid degradation. Even when psychrophiles are used as an inoculum, the lipids degrade relatively less than carbohydrates and proteins. The first step towards the rational engineering of lipolysis in any system is to identify the lipolytic bacteria.

In this study the combination of metagenomics and metaproteomics is used to screen for potential and actual lipolytic bacteria and their extracellular lipases in anaerobic membrane bioreactors treating domestic wastewater at 4℃ and 15℃. The reactors were inoculated by psychrophilic biomass collected from the sediment and soils of Lake Geneva, Switzerland (annual temperature range -11 – 21 °C) and Svalbard, Norway (annual temperature range -16 – 6 °C), respectively. The feed of the reactors was primary influent collected from an activated sludge plant. The bacterial psychrophilic community and their lipases at 4℃ and 15℃ were compared.

Of the 40 recovered putative lipolytic metagenome-assembled genomes (MAGs), only three (*Chlorobium*, *Desulfobacter*, and *Mycolicibacterium*) were common and abundant (relative abundance ≥ 1%) in all reactors. Notably, some MAGs that represented aerobic autotrophs (*Nitrosomonas*) contained lipases. Therefore, the lipases found may not always be associated with exogenous lipid degradation and may have other roles such as polyhydroxyalkanoates accumulation/degradation and interference with the outer membranes of other bacteria.

Different protein classification tools were used for the putative lipase sequences identified by metagenomics to verify if they have potential lipolytic activity. None of the current tools, including *InterProScan*, could precisely assign lipolytic activity to these sequences. Enrichment of public databases by lipase sequences that have been experimentally tested can alleviate this problem.

Metaproteomics did not provide sufficient proteome coverage for relatively lower abundant proteins such as lipases. The expression of *fadL* genes (long-chain fatty acid transporters) was confirmed for four genera (*Dechloromonas*, *Azoarcus*, *Aeromonas* and *Sulfurimonas*), but none of them was recovered as putative lipolytic MAGs. Metaproteomics also confirmed the presence of 15 relatively abundant (≥1%) genera in all reactors, of which at least 6 can potentially accumulate lipid/polyhydroxyalkanoates. For most putative lipolytic MAGs, there was no statistically significant correlation between the read abundance and

reactor conditions such as temperature, phase (biofilm and bulk liquid), and feed type (treated by ultraviolet light or not). Reactor temperature had no statistical correlation with the length of the lipases either. Results obtained by metagenomics and metaproteomics did not confirm each other and further work is required to identify the true lipid degraders in these systems.

Keywords: Anaerobic treatment, domestic wastewater, psychrophilic extracellular lipases, metagenomics, metaproteomics

Acknowledgment

It is the 12th of October, late in the evening. These are the last bit of modifications I am applying to my thesis. On the same date, four years ago, I arrived in the UK. Looking back on that day, I was both excited and scared. I was not sure what I could achieve and where I would stand in my four-year time. After all, I had left my family and many dreams for an uncertain future. Many times, I wanted to give up, but tonight I am happy that I did not.

Doing a Ph.D. is an everyday challenge, and it is even more challenging when you are far from your family, when you are in the minority, and you do not know how to cope with the new culture. Needless to mention that two years of it passed in the Covid-19 pandemic.

But here I am, and my mind, just like my thesis, has many pages. Some pages are blank, some are full, some have even colourful illustrations, and some are summaries of the past.

In these years, I had support from many friends, colleagues, and lab technicians. I know that I cannot thank them enough, but since the chain of kindness is continual, I am hopeful that what they provided will return to them one day.

I would like to express my deepest appreciation to my supervisors Dr. Dana Ofiteru and Prof. Tom Curtis. Their encouragement and critiques were the driving force that always helped me to move forward. They both did their best to teach me a lot about the research.

I would like to extend my sincere thanks to my examiners and appreciate the time they will allocate to review my thesis.

Last but not least, my deepest and warmest gratitude goes to my parents and my sister to whom this dissertation is dedicated. The completion of this thesis would not have been possible without their love and support.

List of conferences and publications related to this study

Manuscripts published

Bashiri, R., Allen, B., Shamurad, B., Pabst, M., Curtis, T.P. and Ofiţeru, I.D. (2022) 'Looking for lipases and lipolytic organisms in low-temperature anaerobic reactors treating domestic wastewater', Water Research, 212, p. 118115.

Manuscripts submitted

Bashiri R, Allen B, Curtis T P, Ofiteru I D. The limitations of the current protein classification tools in identifying lipolytic features in putative bacterial lipase sequences. Submitted to the Journal of Biotechnology.

Conferences

Oral Presentations

- 2021: EBNet Early Career Researcher (ECR) Conference
- 2020: 18th International Symposium on Microbial Ecology (postponed to 2022)
- 2020: EBNet Early Career Researcher (ECR) Conference, Manchester University
- 2020: Chemical Engineering Postgraduate Conference, Newcastle University
- 2020: IWA conference on Water Management in Cold Climate, Harbin, China
- 2019: EBNet Early Career Researcher (ECR) Conference, Sheffield Hallam University

Poster (pitch)

- 2020: IWA World Water Congress & Exhibition (postponed to 2022)
- 2020: Global Water Security Symposium (H70), Newcastle University
- 2019: Anaerobic Digestion Conference AD16, Delft University, Netherlands

Nomenclature

AA: Amino acid
ABC: ATP-binding Cassette
Acyl-ACP: acyl-acyl carrier protein
Acyl-CoA: Acyl coenzyme A
Acyl-PO_4: Acyl-phosphates
ANOVA: Analysis of variance
AnMBR: Anaerobic membrane bioreactors
bp: Base pair
btuB: Vitamin B12 transporters
C: Carbon
CAD: Collision-activated dissociation
CER: Cation exchange resins
CCR: Carbon catabolite repression
CID: Collision induced dissociation
COD: Chemical oxygen demand
DDA: Data dependent acquisition
DIA: Data independent acquisition
EBI: European Bioinformatics Institute
EC number: Enzyme commission number
ECD: Electron capture dissociation
EPS: Extracellular polymeric substances
ENA: European Nucleotide Archive
ESI: Electrospray ionization
ETD: Electron transfer dissociation
E-value: Expect-values
FadD: Acyl coenzyme A synthetase
FadL: Long-chain fatty acid transporter
Fak: Fatty acid kinase
FDR: False discovery rate
GHG: Global greenhouse gas
HCD: Higher-energy collisional dissociation
HRT: Hydraulic retention time
HMM: Hidden Markov models
KEGG: Kyoto encyclopedia of genes and genomes
MAGs: Metagenome-assembled genomes
MALDI: Matrix assisted laser desorption ionization
MFP: Membrane fusion proteins
MS/MS: Tandem mass spectrometry
m/z: Mass-over-charge
N: Nitrogen
NCBI: National centre for biotechnology information
norB/ C: Nitric oxide subunit B/ C
OMP: Outer membrane proteins
Omp32: Outer membrane porin proteins
P: Phosphorous
PHA: Polyhydroxyalkanoates
PhaC: PHA synthesizing genes
porA: major outer membrane proteins P. IA
RED: Relative evolutionary divergence
SDS-PAGE: Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SRT: Solid retention time
susC: TonB-dependent starch-binding receptors
TOF: Times-of-flight
UASB: Upflow anaerobic sludge blanket
UniProtKB: Universal protein knowledgebase
UV: Ultraviolet
UVPD: Ultraviolet photodissociation
VSS: Volatile suspended solid
WWTP: Wastewater treatment plants

Table of contents

List of figures

List of tables

List of appendices

Chapter 1: Introduction

1.1. Scope and goal of the study

The anaerobic treatment of domestic wastewater has the potential (through biogas production) to generate energy. Besides, it produces less sludge than the aerobic treatment systems (typically activated sludge). However, full-scale anaerobic treatment of domestic wastewater is only undertaken in tropical regions where the average temperature of sewage is above 20°C (Bressani-Ribeiro *et al.*, 2019). For 60% of the world population that live in countries with variable seasonal temperatures (below 20°C), using anaerobic treatment for treating domestic wastewater is problematic.

The major limitation of the anaerobic process is related to its first step, the hydrolysis. In this step, fermentative bacteria produce certain extracellular enzymes to degrade large biopolymers like carbohydrates, proteins, and lipids. Yet, at low temperatures, the rate of biological reactions drops, and hydrolysis becomes rate-limiting (Lettinga *et al.*, 2001; Van Lier *et al.*, 2008). Although a psychrophilic microbial community can perform at temperatures below 20°C, they do not hydrolyse all biopolymers at the same rate.

Lipids are more sensitive to lower temperatures and remain relatively undegraded compared to carbohydrates and proteins (Petropoulos *et al.*, 2018). However, we do not know whether poor lipid degradation is due to lack of lipase (the enzyme that degrade lipids) production, lack of lipase activity or further uptake and degradation of hydrolysed long-chain fatty acids.

The present study, therefore, aims to investigate lipolytic potential of the psychrophilic bacteria, during the low-temperature anaerobic treatment of domestic wastewater.

The specific objectives of the research are as follows:

- To identify and compare psychrophilic bacterial community from the anaerobic membrane bioreactors at 4°C and 15°C (Chapter 3).
- To identify and compare potential cold-adapted lipolytic genes, other hydrolytic enzymes genes and their producers at 4°C and 15°C (Chapter 3).
- To identify expressed extracellular lipases and other hydrolytic enzymes or marker proteins (Chapter 4).
- To evaluate and compare protein classification tools for identifying and classifying lipases (Chapter 5).

Chapter 2 : Literature review

2.1. Anaerobic treatment of domestic wastewater

2.1.1. Why should we use anaerobic treatment?

The main aim of wastewater treatment is to provide public sanitation and protect the environment. Yet this process is not carbon-neutral; it contributes 3% of global greenhouse gas (GHG) emissions (Maktabifard *et al.*, 2019), and due to the growth of population its impact on global warming is likely to increase.

Most large to medium wastewater treatment plants (WWTP) are aerobic, exemplified by conventional activated sludge plants, where the biological conversion of organic compounds occurs in the presence of the oxygen.

A major direct and indirect carbon footprint in aerobic plants is related to the aeration tanks and electricity consumption (Maktabifard *et al.*, 2018; Demir and Yapıcıoğlu, 2019). Per year we globally produce about 312×10^9 m^3 domestic wastewater (http://www.fao.org/aquastat/statistics/query/index.html?lang=en, retrieved on 17/01/2021) and we need to consume 0.3-0.6 kwh energy for treating every cubic meter (Soares *et al.*, 2017). Therefore, for the worldwide treatment of wastewater using an aerobic process, we should consume at least 94×10^9 kwh energy per year. Supplying this amount of energy, which is coming mostly from fossil fuels, will increase the GHG emissions significantly. Only in Europe, the 2030 Climate Target Plan (https://ec.europa.eu/clima/policies/eu-climate-action/2030_ctp_en, retrieved on 05/01/2021) urges the member countries to cut the GHG emissions by at least 55% to become climate neutral by 2050. One good approach to reach that target is transitioning the WWTPs from the aerobic to the less energy-intensive processes, like the anaerobic treatment.

The chemical energy of domestic wastewater itself is estimated at about 7.6 kJ/L (2 kwh/m^3) (Heidrich *et al.*, 2011). We can recover some of this energy through anaerobic processes. Anaerobic routes require less input energy, generate biogas (hence heat and electricity), and produce less sludge (Figure 2.1). Taken together, by using anaerobic treatment we might harvest annually seven times (624×10^9 kwh/year) more energy than

what we would need to consume for the global aerobic wastewater treatment (94×10^9 kwh/year).

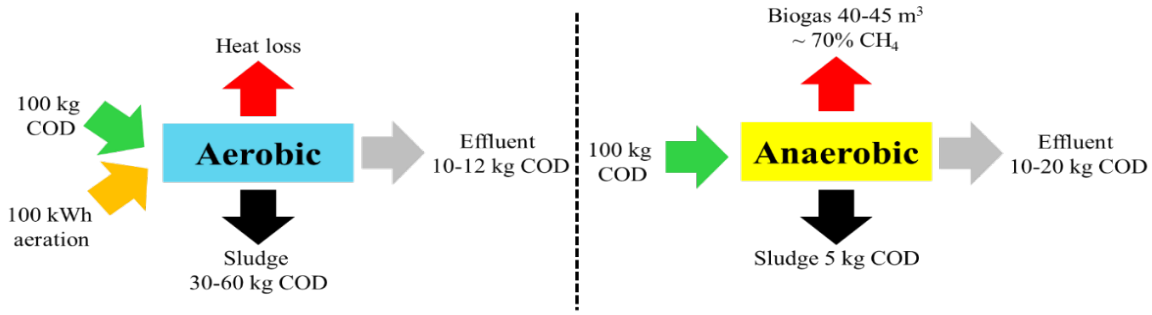


Figure 2-1. Fate of chemical oxygen demand (COD) and energy in aerobic and anaerobic treatment processes, adapted from (Van Lier et al., 2008).

2.1.2. Status and challenges

Despite many advantages, the anaerobic treatment of domestic wastewater is not yet widespread at full-scale except for South America (Bressani-Ribeiro *et al.*, 2019) where the temperature of sewage is 20°C-30°C and the reactors can work at ambient and mesophilic (23°C-34°C) temperatures (Aquino *et al.*, 2019).

Domestic wastewater is characterized as dilute with a chemical oxygen demand (COD) typically below 1000 mg/l, and concentrated in terms of suspended solids (Aquino *et al.*, 2019). At low temperature these two features add to the difficulty of the hydrolysis process, which is the first and the key step in every biological conversion process.

During the hydrolysis, some fermentative bacteria produce extracellular enzymes to hydrolyse polymers into simpler molecules. Carbohydrates, proteins, and lipids are three major classes of polymers in domestic wastewater. Extracellular bacterial enzymes would convert them into their monomeric forms of simple sugars, amino acids, and long-chain fatty acids, respectively. However, when the temperature drops, the rate of biological reactions decreases too (Lettinga *et al.*, 2001) and the hydrolysis step becomes rate-limiting for the whole process (Van Lier *et al.*, 2008; Aquino *et al.*, 2019).

Slow hydrolysis causes the accumulation of polymers in the reactor. Hence at certain solid retention time (SRT), which is required for achieving a high COD removal and biogas production, higher hydraulic retention time (HRT) is needed (Zeeman and Lettinga, 1999; Elmitwalli, 2000). For instance, a 5°C decrease in the temperature of an anaerobic digester, would add 20 days to the initial HRT required to achieve a similar biogas production rate

(Jaimes-Estévez *et al.*, 2021). Long HRTs are not desirable since they increase the operational costs by requiring larger bioreactors and space.

Many researchers have tried to improve low-temperature anaerobic treatment of domestic wastewater, following the publication of Lettinga *et al.* (2001) on the *"challenge of psychrophilic anaerobic wastewater treatment"*.

The central focus of most studies has been: i) reactor set-up and configuration, ii) stepwise adaptation of *mesophilic* inoculum to cold temperatures, iii) recovery of dissolved methane. iv) fouling control in membrane bioreactors, and v) co-digestion of domestic wastewater with other wastes. Nonetheless, in most of the research either ambient temperatures (20°C-25°C) or synthetic wastewater was used. Very few studies used *psychrophilic* microbial community, real domestic wastewater or studied the extracellular processes. Recent publications in this field are briefed in Table 2-1.

Table 2-1. Recent studies on low-temperature anaerobic treatment of wastewater.

Reactor type	Scale	Vol[3]	Feed	Temp.	Inoculum	COD^4_{rem}	CH_4 [5]	HRT	Duration (days)	OLR[6]	Ref
Submerged AnMBR	Pilot-	5000	Municipal WW	25 °C	Mesophilic anaerobic digested sludge from municipal WW	90	0.25-0.27	6 h	217	0.18-1.84	(Kong *et al.*, 2021)
UASB/EGSB[1]/ EGSB-AF2	Lab	3.5	Synthetic dairy WW	Stepwise drop from 37 °C to 15 °C	Mesophilic anaerobic digester treating ethanol production WW (20 g VSS/l).	65-83	45-69%	6.6-8 h	443	7.5–9	(McAteer *et al.*, 2020)
AnDMBR[7]	Lab	3.6	Domestic WW	20 °C-25 °C	Mesophilic anaerobic digested sludge from brewery WWTP	70-77	0.08-0.12	1-8 h	93	0.82-6.8	(Yang *et al.*, 2020)
F-UASB[8], F-AnMBR9, G-AnMBR10	Pilot	70, 70, 42.5	Domestic WW	9.7 °C	F-AnMBR/F-UASB: Municipal digested sludge (primary and secondary sludges), G-AnMBR: Granular sludge	UASB: 57, AnMBR:99	0.13, 0.2, 0.18	8 h	45	-	(Ribera-Pi *et al.*, 2020)
Gas-Sparged AnMBR	Pilot	1300	Municipal WW	12.7 °C-31.5 °C	Mesophilic anaerobic digested sludge from municipal WW	88	0.14	20 h	472	1.3	(Lim *et al.*, 2019)
Submerged AnMBR	Lab	4.5	Malting WW	23 °C	Mesophilic anaerobic seed sludge from brewery WWTP	87-92	0.21	1.8-3.3 h	287		(Maleki *et al.*, 2019)

1. Expanded granular sludge bed, 2. Expanded granular sludge bed anaerobic filter, 3. Working volume (l) 4. Percentage of COD removal, 5. Methane yield (l/g COD removed)

6. Organic loading rate (kg COD m^{-3} d^{-1}), 7. Anaerobic dynamic membrane bioreactor, 8. Flocculent biomass UASB, 9. Flocculent biomass AnMBR, 10. Granular biomass AnMBR

The two frequently used formats for anaerobic reactors are upflow anaerobic sludge blanket (UASB) reactors and anaerobic membrane bioreactors (AnMBRs). In UASB reactors (Figure 2-2), gas and suspended solids get separated at the top of the reactor. Suspended solids further granulate and settle to form a sludge blanket/bed, providing a continuous contact between the active biomass and the fresh wastewater.

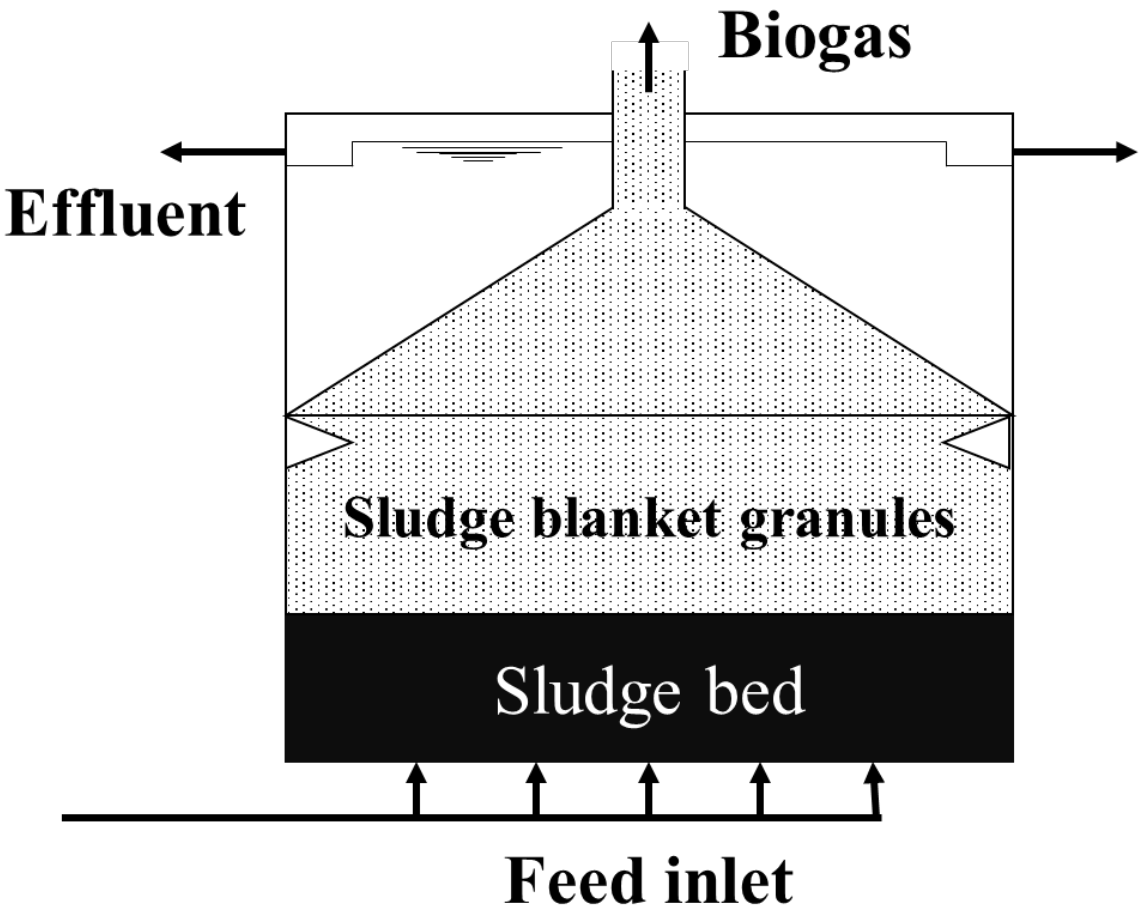


Figure 2-2. Schematic diagram of upflow anaerobic sludge blanket reactor

However, at low loading rates, or when the feed distribution system is poorly designed, wastewater and the active biomass would be barely in contact, which results in the liquid channelling (clogging) phenomenon.

For dilute and cold wastewater, the risk of channelling is even higher. Both low biogas production and formation of thinner sludge blanket due to slow hydrolysis, prevent the sufficient mixing between the phases at the top and the bottom of the reactor (Lettinga *et al.*, 1984).

Operational data from a 6 m^3 UASB reactor treating domestic wastewater at temperatures between 9.5 and 19°C (liquid retention time 8h) has shown that channelling in the sludge bed

at temperatures below 12 ℃ would cause low biogas production and poor suspended solid removal (Lettinga *et al.*, 1984).

By contrast, anaerobic membrane bioreactors (AnMBRs) separate solid/liquid phases more efficiently through membrane and retain the biomass for longer. For cold domestic wastewaters that suffer from slow hydrolysis, the longer SRTs that AnMBRs provide improve the COD removal (Smith *et al.*, 2012), provided that no membrane fouling occurs (Penfield, 2017). COD removal of more than 95% and 86% has been reported for AnMBRs at 6℃ and 3℃ using *mesophilic* and *psychrophilic* inoculum (Smith *et al.*, 2013; Smith *et al.*, 2015). However, real domestic wastewater was not used.

In addition, membrane fouling, and methane oversaturation are two drawbacks of AnMBRs in treating cold domestic wastewaters (Ozgun *et al.*, 2013; Li and Yu, 2016). As the temperature drops, methane solubility in the permeate water would increases too, which can account for losses of about 45%-88% of the total produced methane (Cookney *et al.*, 2016). Life cycle assessment analysis has shown that failure in recovering the dissolved methane from the low-temperature AnMBRs removes their benefit in terms of GHG mitigation relative to the aerobic processes (Smith *et al.*, 2015).

The other challenge of low-temperature anaerobic treatment of domestic wastewater is poor lipid degradation. Most studies have used *thermophilic* and *mesophilic* inoculum in their reactors. *Thermophilic* and *mesophilic* microbial communities are not adapted to cold and hence their reaction rate, which follow the *Arrhenius* equation (K= A exp (-E/RT), K= Reaction rate, A=Arrhenius factor, E=Activation energy, R= Universal gas constant, T=Temperature), decreases significantly as the temperature drops.

By contrast, *psychrophiles* and psychrotolerant can maintain a high reaction rate even at temperatures near zero (Figure 2.3). A psychrophilic microbial community, therefore, might perform better for hydrolysing the lipids at low temperatures. A scum layer of lipids at interface of liquid usually forms in reactors during the treatment of slaughterhouse wastewaters at 20℃ (Sayed and Lettinga, 1984). Mechanical solutions like installing a skimmer in the reactor can remove the scum layer from the lipid-rich wastewater (Lettinga *et al.*, 1984) and solve the problem. However, an increase in lipid degradation can enhance the biogas production.

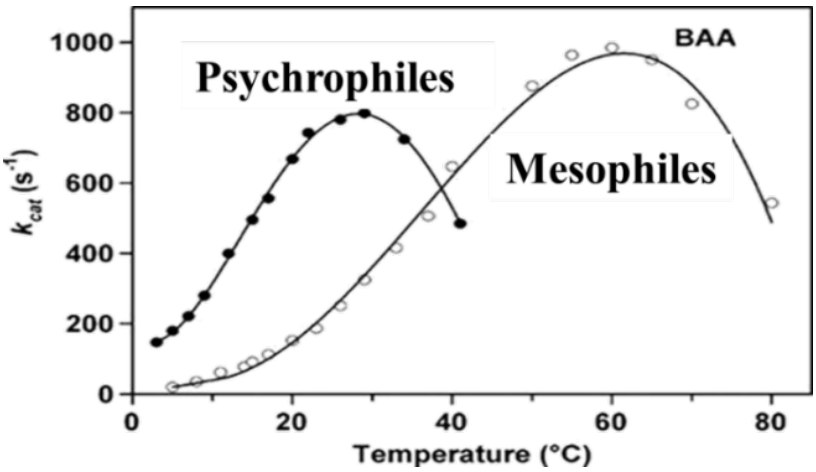


Figure 2-3. effect of temperature on the activity of α-amylase enzyme produced by psychrophilic and mesophilic bacteria (Georlette et al., 2004).

Estimated methane yield from 1 gr glycerol trioleate (an abundant natural lipid) is 1.08 L (at standard temperature and pressure) while for 1 gr glucose it is only 0.37 L (Kim and Shin, 2010). Also, aerobic assimilation of 1 g olive oil yields 1.2 g dry biomass whereas this yield from 1 g glucose is only 0.5 gr (Becker, 2010).

Yet even *psychrophilic* microbial community adapted to 4°C, 8°C, and 15°C (in an AnMBRs fed with domestic wastewater) failed to degrade lipids (Petropoulos *et al.*, 2018). Petropoulos *et al.* (2018) concluded that poor lipid degradation is rather due to lack of lipase (an extracellular enzyme that degrade lipids) activity than lack of lipase production. However, they did not investigate the extracellular processes of *psychrophiles,* their potential for producing and excreting the extracellular lipases or other hydrolytic enzymes.

2.2. Lipid degradation

Lipids comprise a wide range of molecules. In wastewater, they mostly represent natural fats and oils such as fatty acids, glycerides (esters of fatty acids with glycerol) and phosphoglycerides (esters of fatty acids and phosphoric acid with glycerol combined with other radicals) (Hrudey, 1981). The major part of lipids in raw wastewater is triacylglycerides and only a small fraction is in the form of free long-chain fatty acids (Dueholm *et al.*, 2001). The presence of lipids in aerobic systems cause problems like sludge flotation, bulking and foaming due to the growth of filamentous microorganisms which limit the oxygen transfer (Chipasa and Mdrzycka, 2008).

In the activated sludge process the efficiency of lipid degradation highly depends on the ratio of lipid to microorganism. Maintaining the content of lipid at 0.1 grams per day per gram of

mixed liquor suspended solid is suggested to prevent overloading of lipids in the reactor (Hrudey, 1981).

Aerobic-thermophilic processes (65 ºC) have reported to be advantageous for treating lipid-rich wastewaters. At high temperatures, lipids have different physical properties. For example, the operating temperature is above the melting point of the lipids which makes them more accessible to lipases. Besides, the diffusion coefficient and solubility of long-chain fatty acids increases and microbes can take them up more easily (Becker *et al.*, 1999).

However, Becker and Märkl (2000) have shown by simulations that fluctuations in lipid concentration (e.g. increasing the lipid content in the feed to 4 g/l) even at such favourable aerobic-thermophilic condition can cause complete biomass washout. They have proposed that limited *β-oxidation* of the released long-chain fatty acids rather than lipase production is the barrier to lipid degradation at high lipid concentrations. *β-oxidation* is a catabolic process during which fatty acids loose two carbon in each step and get degraded (Jimenez-Diaz *et al.*, 2017).

Models developed for lipid degradation in the activated sludge process imply that in addition to the lipolysis and fatty acid assimilation, lipid production by lipid-accumulating microorganisms can limit the lipid degradation (Chipasa and Mdrzycka, 2008). *Candidatus Microthrix parvicella* is a well-documented taxon in both the activated sludge plants and anaerobic conditions (Nielsen *et al.*, 2002). It is responsible for lipid degradation (it has 8 lipase genes) and assimilation of long-chain fatty acids to biosynthesize lipids (related gene is *wax ester synthetase/Acyl-CoA: diacylglycerol acyltransferase*).

Under anaerobic conditions, fermentative bacteria hydrolyse lipids to long-chain fatty acids though they might not oxidize them themselves. In essence, two other groups of anaerobes, obligate hydrogen-producing syntrophs and sulphate-reducing bacteria utilize long-chain fatty acids and oxidize them through the *β-oxidation* pathway. In addition to these two, sulphur-reducing bacteria and denitrifiers also degrade long-chain fatty acids if light is absent (Mackie *et al.*, 1991). Genera like *Mycobacterium*, *Rhodococcus*, and *Nocardia* have been identified in anaerobic reactors as lipid-consumers and accumulators. These taxa can be very abundant and are usually afloat on the surface as a foam (Muller *et al.*, 2014).

However, we still do not know who the major lipid-degraders are, particularly in low-temperature anaerobic treatment of wastewater. Studying the extracellular processes that microbes perform can elucidate this problem.

2.3. What is an extracellular process?

Extracellular processes are microbial activities which results in excretion of biomolecules to the extracellular medium. Such extracellular releases have various advantages for the survival of a microbial community. One of the main purposes is to transfer nutrients between cells, as recently reviewed by Fritts *et al.* (2021).

Releasing the extracellular polymeric substances (EPS), extracellular enzymes, extracellular vesicles, quorum sensing signals, siderophores (iron transporters), toxins, metabolites, nanowires (electron transfer) (Ilshadsabah and Suchithra, 2019), and nanotubes (DNA, proteins and nutrient exchange) (Pospíšil *et al.*, 2020) as shown in Figure 2.4 are all examples of extracellular processes that can aid microbes to take up the nutrients they require (Fritts *et al.*, 2021).

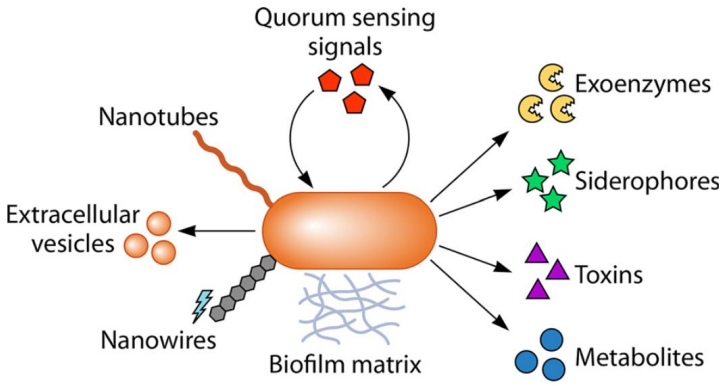


Figure 2-4. Extracellular biomolecules produced by microbial cells to promote cross-feeding (Fritts et al., 2021).

2.3.1. Extracellular polymeric substances (EPS)

Microorganisms often live in communities known as biofilms (cities of microbes). Within these metaphorical cities, microbes build 'houses' scientifically known as extracellular polymeric substances (EPS) (Flemming and Wingender, 2001; Flemming *et al.*, 2007). The production and excretion of the EPS is an extracellular process that living cells perform. Some components of the EPS can also come from the surrounding environment or from the cell lysis processes.

The EPS is a matrix that consists of biopolymers (polysaccharides, proteins, and lipids), nucleic acids and charged ions. Each component of the EPS plays a role in the survival of the microbial community. For instance, polysaccharides, such as alginate, serve as a 'shield' and protect the microbes against biocides either by limiting the diffusion of such chemicals or by forming a chemical bond with them. Lipids are surface-active and help bacteria tolerate the strong surface

tension of the surrounding water. Proteins on the other hand, influence the formation of microbial flocs by forming electrostatic and hydrophobic bonds with other components or more importantly have hydrolytic roles. Extracellular enzymes hydrolyse organic compounds and help microbes access the carbon and other necessary resources present in their surroundings (Wingender *et al.*, 1999).

2.3.2. Extracellular enzymes

Microbes assimilate nutrients for their growth and maintenance. Nonetheless, most available nutrients do not fall within the size threshold that microbial cell walls allow for the passage of molecules. For Gram-negative bacteria the allowed molecular weight for passing through the outer membrane is 600 Dalton (Arnosti, 2011). Yet in aquatic environment, about 95% of the organic compounds are in polymeric forms (Chróst, 1991) like carbohydrates, proteins and lipids. These molecules are far larger than 600 Dalton. For instance, Glycine is the smallest amino acid with a molecular weight of 75 Dalton. The smallest known protein has only 20 amino acids (Neidigh *et al.*, 2002). If it is possible to have a mini protein only built of 20 Glycine, the overall molecular weight of such protein would be 1500 Dalton which still cannot cross the membrane pores of Gram-negative bacteria.

In essence, microbes produce extracellular enzymes and hydrolyse the large polymers to monomers (e.g., proteins to amino acids) that can cross the microbial wall. However, not all microbes are producers. Cheater microorganisms steal the hydrolysis products without allocating their energy for producing such enzymes (Allison *et al.*, 2014a) and increase their fitness.

2.3.3. Wired and wireless extracellular enzymes

Extracellular enzymes are in two forms: cell-bound (attached to the cell) and cell-free (excreted in the surrounding environment?). In the past it was assumed that most of the degradation process is performed by cell-bound enzymes (Baltar *et al.*, 2010). However, the cell-free enzymes activity is as important though they are not physically connected to the cell. Some studies have estimated half-lives of up to 20 days for cell-free enzymes in seawater and shown that low temperatures and lack of ultraviolet radiation promotes their lifetime (Baltar, 2018).

The two forms of enzymes respond differently to environmental change. While the lifetime and activity of cell-bound enzymes depend on the growth, activity, and diversity of the cell,

cell-free enzymes act independent of their producers and will react with substrates as long as they are not trapped by particles (Baltar, 2018).

2.3.4. When do the cells produce extracellular enzymes?

Production comes at a cost. The building blocks of enzymes are amino acids, and cells use carbon (C) and nitrogen (N) in a ratio of 3 to 1, to synthesize them. Bacteria usually lose about 1-5 % of their metabolic productivity for enzyme production (Allison *et al.*, 2014b). Therefore, microbes would only invest their energy for synthesis of the extracellular enzymes and their transportation if they can outweigh these costs. Factors like temperature, nutrient availability, spatial structure, and competition can influence the production/activity of extracellular enzymes.

Computational models have shown that at limited concentration of nutrients, e.g., C, N and phosphorous (P), the relevant nutrient-releasing enzymes dominate the system. For instance, at low C concentration, C-hydrolysing enzymes have higher production rate than the N/P-hydrolysing enzymes. However, when N is low, total enzyme production declines and more than 50% of the C-mineralisation drops. Furthermore, the addition of N relative to C and P has a higher impact on the growth rate (Allison, 2005).

2.3.5. What are the threats for extracellular enzymes?

Ecologically, extracellular enzymes are public goods. This means that while they are costly for individual producers, cheater organisms (non-producers) gain a competitive advantage without paying the cost of production. However, the spatial structure can affect the cheating and cooperation mechanisms (Allison *et al.*, 2014b). Based on computational models, when cheaters are present, well-mixed environments would increase the opportunities of cheaters to gain hydrolysed monomers and outnumber the producers. By contrast, when the diffusion rate is limited, producers gain the advantage and increase their population size (Allison, 2005; Allison *et al.*, 2014b). Apart from the cheaters, factors like pH, temperature and inhibitors can alter the activity or production rate of the enzymes.

2.3.6. Lipolytic enzymes: Lipases vs esterases

Lipases (EC 3.1.1.3) and esterases/carboxylesterases (EC 3.1.1.1.) are members of a broader class of hydrolytic enzymes called carboxylic ester hydrolases. They both act on ester bonds; however, lipases cleave the lipids that have i) long-chain fatty acids (C_{12} and higher), and ii)

are water-insoluble (Hausmann and Jaeger, 2010). Furthermore, most lipases show a unique phenomenon called interfacial activation which esterases do not. When an emulsion forms at the interface of water and lipid, the lid which protects and cover the active site of the lipases would go through conformational change and allow the catalytic reaction to start (Verger, 1997). However, some lipases like those that are produced by *Pseudomonas aeruginosa* and *Bacillus subtilis*, do not necessarily need the interfacial activation (Jaeger *et al.*, 1994).

Both lipases and esterases share an α/β-hydrolase fold in their structure, but at the sequence level they are diverse (Verma *et al.*, 2021). Arpigny and Jaeger (1999) initially classified bacterial lipolytic enzymes into eight families based on their structures. This classification is now broader and was last updated in 2018 by Kovacic *et al.* (2018). More details about the structure of lipases and their classifications are provided in Chapter 5.

2.3.7. Cold-adapted lipases and industry

Most microbial lipases that are in use in industry are mesophilic; however, cold-adapted lipases are more desirable for industries that manufacture detergents, paper, food, and pharmaceutical products (Mhetras *et al.*, 2021; Verma *et al.*, 2021). For example, using *psychrophilic* lipases in laundry detergents can considerably reduce the energy demand of washing machines, increase the durability of the clothes, and minimize waste production (Mhetras *et al.*, 2021).

Despite the growing global market for microbial lipases, estimated at $590 million by 2023, very few lipase producing organisms are identified. Some of the bacterial genera that produce cold-active lipases are *Photobacterium lipolyticum (Ryu et al., 2006)*, *Aeromonas sp*. (Lee *et al.*, 2003), *Pseudoalteromonas sp. (Zeng et al., 2004)*, and *Psychrobacter sp.* (Joseph *et al.*, 2007).

2.3.8. Lipase inducers and inhibitors

Microbial lipase production usually requires inducers such as oils, triglycerides, long-chain fatty acids, TWEENs, hydrolysable esters, n-alkanes, bile salts and glycerol. Nonetheless, excessive concentrations of inducers might have a negative impact and inhibit the expression of lipases. In natural systems, where a diverse microbial community is interacting and the ecology impacts its responses, lipase production might be very complex. Even, different species may have different inducers. For example, for *G. thermoleovorans IHI-91*, glycerol has no inducing effect and can even repress lipases (Becker, 2010).

Glucose is an easier carbon source for microbes to catabolize and its presence at concentrations higher than 2 g/L has an inhibitory effect on lipase production (Becker, 2010). This phenomenon is called catabolite repression, a global regulatory system that prevents bacteria from using the secondary carbon sources when their preferred carbon source is present (Görke and Stülke, 2008). Furthermore, the addition of inducers cannot overcome the repression that glucose causes for lipase production (Pauli *et al.*, 1974).

The accumulation of long-chain fatty acids also has an inhibitory effect on lipase production (van den Berg, 2005; Becker, 2010). The regulon protein, *fadR*, controls the expression of all genes involved in the uptake and transport (*fadL*, long-chain fatty acid transporters), activation (*fadD*, acyl-CoA synthetases), and degradation (*fadA*, *fadB*, *fadE*, *fadF*, *fadG* and *fadH*) of long-chain fatty acids in Gram-negative bacteria (Kunau *et al.*, 1995). When the concentration of acyl-CoA exceeds a certain threshold in the intracellular medium, *fadR* represses the expression of both *fadL* and *FadD* (van den Berg, 2005). Therefore, long-chain fatty acids accumulate in the extracellular medium and inhibit lipase production. Even poor lipid degradation by thermophilic bacteria like *G. thermoleovorans IHI-91*, has been associated to repression of *fadL*, lack of long-chain fatty acid transport to the cell and limited capacity of *β-oxidation* for degrading them (Becker, 2010).

2.3.9. Extracellular vesicles

Extracellular vesicles are a package of proteins (intracellular or outer membrane based), lipids, and nucleic acids that all cells can produce and excrete. For bacterial extracellular vesicles, there are two common terms in the literature: i) outer membrane vesicles for Gram-negative bacteria (average diameter of 20–200 nm) and ii) membrane vesicles for Gram-positive bacteria (average diameter of 20–100 nm) (Kim *et al.*, 2015).

Initially it was assumed that cells eliminate certain proteins, lipids, and RNA selectively through releasing extracellular vesicles (van Niel *et al.*, 2018; Woith *et al.*, 2019). However, it is now known that cells use extracellular vesicles for different purposes including cell-cell communications, nutrient transport, invading the competitor cells, horizontal gene transfer, infection and releasing extracellular enzymes.

Some of the characterized proteins and other molecules in the extracellular vesicles are ABC transporters, porins (*OmpA*, *OmpC*, *OmpF*), TonB-dependent receptors (for the uptake of large molecules like *iron-siderophores* or *vitamin B12* that cannot diffuse through porins) (Frias *et*

al., 2010), long-chain fatty acid transporters (*fadL* in Gram-negatives), periplasmic proteins (alkaline phosphatase) (Kim *et al.*, 2015), virulence factors (adhesins, lipopolysaccharides, *β-lactamase*), viral particles (Liu *et al.*, 2018), ribosomal proteins, and RNA (Tsatsaronis *et al.*, 2018).

Extracellular vesicles of bacterial cells can transfer their cargo to the target cells through membrane fusion. This way extracellular RNAs would enter those cells and modulate, silence or enhance expression of certain genes (Tsatsaronis *et al.*, 2018). Also, transporter proteins can bind their substrate and carry them to the target cells. For example, *Prochlorococcus* (a marine cyanobacterium) releases the extracellular vesicles that contain phosphate-binding proteins. It is suggested that these proteins can scavenge extracellular phosphates and carry it to the target cells. The same mechanisms have been proposed for iron and zinc-binding proteins in the extracellular vesicles of *Neisseria meningitidis* (Biller, 2020).

What is interesting is that extracellular enzymes like lipases have been identified in the extracellular vesicles as well and have been reported as a virulence factor in Gram-positives (Lee *et al.*, 2009b; Kim *et al.*, 2015). (Baltar, 2018) has proposed that extracellular vesicles are a way for the cells to release cell-free enzymes to last longer in the extracellular medium and act independently of the cells. Long-chain fatty acid transporters have been frequently identified in the extracellular vesicles of Gram-negative bacteria (Lee *et al.*, 2008; Lee *et al.*, 2016a; Hong *et al.*, 2019). However, it is still not known what the role of these transporters in the extracellular vesicles is. Since long-chain fatty acid transporters act as bacteriophage (viruses that attack bacteria) T_2 receptors too (Black, 1988), it is possible that some bacteria might use vesicles to reduce their susceptibility against bacteriophage or expose their competitors to this virus attack. Wild-type *Escherichia coli* lower the expression of long-chain fatty acid transporters (Jeon *et al.*, 2018) which strengthen this idea that cells might regulate these proteins at a level that is safe and can protect them against bacteriophage attachment.

2.4. How do we study extracellular processes?

To study a microbial community, it is helpful to know which bacteria have which function. Metagenomics and metaproteomics are two molecular biology tools that help us to access such data. These two workflows consist of wet and dry lab approaches. The latter is known as bioinformatics which is a computational branch of biology (Claverie and Notredame, 2006).

Different tools have been developed to shape our perspective toward the microbial world, however, none of them are still able to depict an unbiased picture.

The wet-lab part of the metagenomics workflow consists of DNA extraction and sequencing to find the order of the nucleotides that comprise a DNA molecule. Sequencing machines generate thousands or millions of reads that are in essence pieces of a jigsaw puzzle. By using bioinformatics tools we put these pieces together, reconstruct the genomes, predict the gene cluster, and the potential corresponding proteins of the microbial population inhabiting an environment (Hugenholtz and Tyson, 2008).

The main steps of data processing and analysis in metagenomics comprise of i) quality control which involves trimming and removing the short and low-quality reads, ii) assembly through which longer reads like contigs are produced, iii) binning that recovers the genomes as metagenome-assembled genomes (MAGs), iv) taxonomic classification that identify who is there and v) gene prediction and annotation that determine what the community is doing (Figure 2.5).

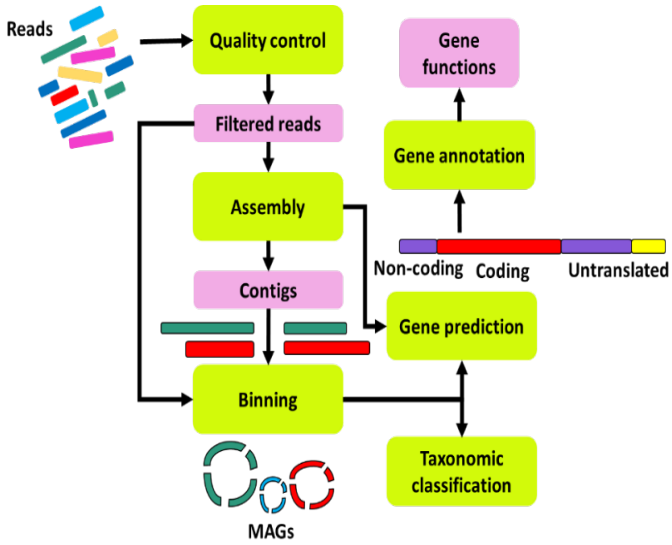


Figure 2-5. Schematic diagram of post-sequencing analysis of reads in metagenomics study.

In metaproteomics by contrast, we extract the proteins, and measure their mass spectra to measure the gene expression among the community. Nonetheless, metaproteomics presents more than the gene expression (Kleiner, 2019). Some questions that only metaproteomics can currently answer about the microbial communities are: i) what is the community structure based on the protein biomass? (Kleiner *et al.*, 2017), ii) what is the expressed metabolism and physiology of the community? (Kleiner *et al.*, 2012), iii) how do the members of a community interact? (Hamann *et al.*, 2016), iv) who uses a certain substrate (Bryson *et al.*, 2016; Jehmlich *et al.*, 2016), v) what carbon sources and assimilation pathway microbes use? (Kleiner *et al.*, 2018).

Compared to metagenomics, metaproteomics bioinformatics tools are still in their infancy. Metaproteomics still does not allow us to trace isotopically labelled substrates in small amounts or measure a growth rate of each member of a community. We cannot only identify the

extracellular enzymes when we only want to focus on hydrolytic enzymes. Additionally, with metaproteomics, measuring the abundance of viruses, or determining the age and role of cell-free proteins (e.g. proteins in the extracellular vesicles) are not yet possible. (Kleiner, 2019). For metagenomics, on-line pipelines like KBase (Arkin *et al.*, 2018) and MGnify (Mitchell *et al.*, 2017) are recently developed by the US Department of Energy and European Bioinformatics Institute (EBI), respectively. These services are free and provide a user-friendly environment for analysis and subsequently sharing the metadata publicly. However, these pipelines require longer processing time (users should stand in the queue for submitting some high memory-demanding jobs) and fail to process large data due to limited memory size. For metaproteomics by contrast, such pipelines do not exist.

Chapter 3 : Do psychrophiles have the potential to produce lipases?

3.1. Introduction

Two of the fundamental questions about a microbial community that metagenomics can, potentially, answer is: who is there and what are they doing? In any given microbial community some bacteria must, necessarily, produce extracellular enzymes to gain the carbon (C), nitrogen (N) and phosphorous (P) from the polymers in their environment. Metagenomics can help us to determine the identity of those bacteria by allowing us to determine which genomes in the microbial community have genes that code for hydrolytic enzymes. I am particularly interested in Lipases (EC 3.1.1.3). Lipases are members of carboxyl ester hydrolases (Ali *et al.*, 2012) and can degrade lipid molecules. Those microbes that can break-down lipids rather than carbohydrates, gain more energy for growth. One gram of glucose under aerobic condition can yield half a gram of dry biomass whereas the yield of olive oil is about 1.2 grams per gram (Becker, 2010).

Lipases can break down the ester bonds of triacylglycerides and diacylglycerides and release the long-chain fatty acids from the glycerol backbone (Figure 3-1). Both molecules can then enter the cell but would have a different fate. Long-chain fatty acids (C_{12} and longer) unlike short-chain (C_6 and smaller) and medium-chain (C_7-C_{11}) fatty acids which diffuse through the membranes or porin channels, need a protein-mediated apparatus. The only well characterised transporter protein for long-chain fatty acids are *FadL* genes (in the outer membrane) which are identified in *Escherichia coli* (*E. coli*) though homologues of these proteins have been seen in other gram-negative bacteria as well (Clark and Cronan, 2005). Unfortunately, it is not clear in the literature if the same transporter protein exists in the gram-positive bacteria too or due to the different outer membrane structure, they employ a different transportation mechanism.

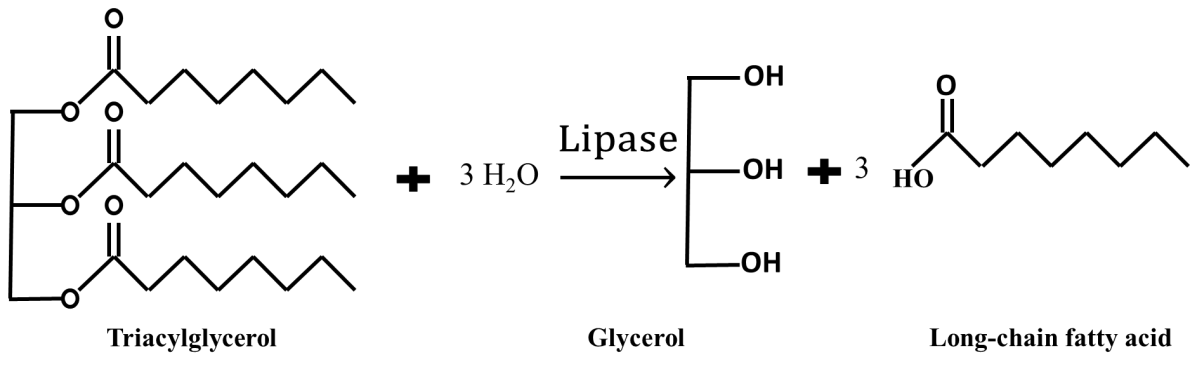


Figure 3-1. Lipolysis of triacylglyceride molecules by lipases: Lipases hydrolyse triacylglycerol into glycerol and long-chain fatty acids.

Imported exogenous fatty acids have different fates in different cells. In gram-negatives, nearly 2% of exogenous fatty acids are converted to acyl-ACP (for phosphatidic acid synthesis) and 98% to acyl coenzyme A (acyl-CoA) (Jimenez-Diaz *et al.*, 2017). The latter is done by *FadD* or acyl coenzyme A synthetase. The acyl-CoA, as well as being used for phospholipid synthesis can also be broken down by *β-Oxidation* pathway to yield energy. Gram-positive bacteria on the other hand have an alternative gene, fatty acid kinase (*Fak*), that converts exogenous fatty acids to the acyl-phosphates (acyl-PO_4). Therefore, modified fatty acids would only take part in phospholipid synthesis or forming acyl-acyl carrier protein (acyl-ACP) that can synthesize phosphatidic acid (Yao and Rock, 2017). One study has shown by metagenomics and metabolic labelling that *Staphylococcus aureus*, a gram-positive bacterium from the *Firmicutes*, do not have the genes for fatty acid degradation (*β-Oxidation* pathway) and the only fate of the exogenous fatty acids in these cells are to be incorporated as a cellular component or go through the elongation process (Parsons et al., 2011). It is not clear from the literature if lacking the genes for the *β-Oxidation* pathway is a feature of all *Firmicutes* or gram-positives. Nonetheless, *Mycolicibacterium*, are gram-positives and well-known for growing on lipidic substrates. For instance, *Mycobacterium tuberculosis*, is a pathogen and use host lipids to gain energy. This bacterium has multiple genes for *β-Oxidation* pathway. 'Redundant' enzymes probably help this bacterium to adapt to different environments and switch its metabolism (Toledo and Benach, 2015).

Bacteria can also synthesize fatty acids endogenously. These synthetic fatty acids would only be converted to either acyl-ACP or β-hydroxyacyl-ACP for phospholipid and lipopolysaccharide synthesis, respectively (Yao and Rock, 2017). This means endogenous fatty acids do not degrade through the *β-Oxidation* pathway. Yet, endogenous fatty acid synthesis is energy-intensive and can at least compete with the regulatory system that incorporates exogenous fatty acids to convert them to acyl-ACP for phosphatidic acid synthesis. In gram-negatives for instance, about 2% of exogenous fatty acids are converted to acyl-ACP (Jimenez-Diaz *et al.*, 2017). Therefore, we do not know when the cells favour fatty acid synthesis or transporting the exogenous fatty acids which for the latter cells need to express extracellular lipases and degrade lipids to fatty acids before they can transport. However, we know that the endogenous fatty acids are always converted to either acyl-ACP or β-hydroxyacyl-ACP for phospholipid and lipopolysaccharide synthesis, respectively (Yao and Rock, 2017). This means that no acyl-CoA would be formed from the endogenous fatty acids to go through fatty acid degradation process by *β-Oxidation*.

During low temperature anaerobic treatment of domestic sewage, lipids, unlike carbohydrates and proteins, remain relatively undegraded (Petropoulos *et al.*, 2018). We do not know why this happens. Plausible hypotheses include: i) lack of lipolytic genes compared to other carbon-acquiring genes ii) higher costs of lipase production for cells compared to other enzymes; iii) depression of lipase genes iv) inactivation of extracellular lipases, lower bioavailability of lipids at cold temperatures.

The first step towards understanding why lipids do not degrade at low temperatures is to understand which taxa are present and what genes they have in their genomes. Metagenomics cannot, however, tell us whether those genes were expressed inside the cells or excreted to the extracellular media. Protein expression and excretion can be confirmed by metaproteomics (discussed in Chapter 4).

In this chapter results from the metagenomes of the cold-adapted microbes taken from the lab-scale AnMBRs at 4℃ and 15℃ are presented. The purpose of the analysis was to find the lipase coding genes and see how different temperature and treatment conditions could affect them.

3.2. Materials and method

3.2.1. Reactor set-up

The reactor set-up, inoculation, feeding, and wastewater characterization is described in detail by Petropoulos et al. (2017). Four AnMBRs with 1 L working volume (and their duplicates) were operated at 4 ℃ and 15 ℃ under the Sterile (treated with the ultraviolet light to exclude mesophilic microbes of the feed) and Non-sterile conditions. The reactors were inoculated by psychrophilic biomass collected from the sediment and soils of Lake Geneva “N 46°23′04, E 6°25′ 07” (-11–17 °C) and Svalbard, “N78°, E11, 15,16°” (-16–6 °C), respectively. The feed of the reactors was primary influent collected from an activated sludge plant (Tudhoe Mill, County Durham, UK). More details about the reactor set-up and performance are included in Appendix A.

3.2.2. DNA extraction and sequencing

DNA was extracted from the anaerobic bioreactors sample (both bulk liquid and biofilm) using the CTAB method (Griffiths *et al.*, 2000) and sent for sequencing (HiSeq 2500 platform) to the Earlham Institute, Norwich. Amplification free, Illumina compatible libraries were constructed

using the Kapa Hyper Prep kit. Aliquots of each sample were run on two lanes/two flowcells to generate paired end (PE 250) reads of about 300 Mb.

3.2.3. Read processing and bioinformatics

FastQC v0.11.5 was employed to check the quality of reads, and *Cutadapt v1.18* and *Trimmomatic v0.36* were used to trim the adapters (From Read 1: *AGATCGGAAGAGCACACGTCTGAACTCCAGTCA* and From Read 2 *AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT*) and poor regions. Filtered reads were co-assembled with *MEGAHIT v.1.2.9* using a high-performance computer at *Newcastle University*. Obtained contigs were then binned with *MetaBat2 v1.7* to recover the metagenome-assembled genomes (MAGs). To evaluate the quality of the bins, *CheckM v1.0.18* was used and MAGs with more than 90% completeness and less than 10% contamination were selected as good bins (bins and MAGs are interchangeable terms). The *FASTA* file of the selected MAGs were uploaded to *KBase* (Arkin et al. 2018) and annotated using *Prokka v1.12*. After annotation, lipase genes were searched with their enzyme commission number (EC number: 3.1.1.3). Bins that had at least one (putative) lipase gene were specified as putative lipolytic bins. Specific EC numbers of other hydrolytic enzymes like phosphatases, proteases, esterases and carbohydrate degraders were searched (Appendix B). Further analysis like the taxonomic classification was performed by *GTDB-Tk v0.3.2* (Chaumeil et al. 2019) on putative lipolytic bins at KBase. To find the relative abundance of the microorganisms existing at each reactor condition, reads from both biofilm and liquid phase of the replicate reactors at each temperature and treatment set-up were merged with KBase apps and were analysed with *GOTTCHA2 v2.1.5*. All statistical analysis was performed using Minitab 18. All metagenomics data are accessible at European Nucleotide Archive (ENA) under the accession number PRJEB47041.

3.3. Results and discussion

3.3.1. Reads, Contigs and MAGs

The number of sequenced reads only depends on the quantity and quality of the extracted DNA and the sequencing platform. The highest and lowest number of reads belonged to liquid phase of the sterile feed at 4°C (100 million) and 15°C (64 million), respectively (Table 3-1).

Table 3-1. Reads generated after sequencing the DNA extractes of both liquid and biofilm phase in AnMBRs.

Sample	Number of generated reads	
	Liquid phase	Biofilm phase
4°C-Nster 1	78,901,230	85,281,044
4°C-Nster 2	78,902,998	82,749,706
4°C-Ster 1	100,225,412	83,599,124
4°C-Ster 2	86,566,622	65,788,040
15°C-Nster 1	73,358,304	92,205,090
15°C-Nster 2	71,669,572	81,414,448
15°C-Ster 1	76,591,032	82,766,216
15°C-Ster 2	64,788,634	89,683,802

About 1 million (M) contigs with a total length of nearly 1.5 billion base pair (bp) were found. The largest contig was about 1 Mbp long (Table 3-2). The N50 and L50 were 1,490 bp and 186,044, respectively. N50 and L50 are two statistical terms required to compare the quality of different assemblers. The best assemblers usually give fewer longer contigs, which means that higher N50s are preferred. If we sort all contigs from largest to smallest and calculate the total length, N50 is the length of the contig at which half of the total length of the assembly is ranked and L50 is the rank of that contig. In other words, the 186,044th contig that was 1490 bp long ranked half of the total length of the assembly.

Table 3-2. Contigs statistics obtained from the co-assembly of the reads of the AnMBRs.

Contigs information	Statistics
Total number of contigs	1,109,690
# contigs (>= 0 bp)	1,109,690
# contigs (>= 1,000 bp)	352,375
# contigs (>= 10,000 bp)	9,103
# contigs (>= 100,000 bp)	142
# contigs (>= 1,000,000 bp)	1
Largest contig (bp)	1,226,853
Total length (bp)	1,428,194,318
Total length (>= 0 bp)	1,428,194,318
Total length (>= 1000 bp)	913,963,731
Total length (>= 10000 bp)	210,083,869
Total length (>= 100000 bp)	25,407,583
Total length (>= 1000000 bp)	1,226,853
N50 (bp)	1,490
N75 (bp)	782
L50	186,044
L75	531,442
GC (%)	52.49

We recovered about 1519 MAGs. However, only 40 MAGs had at least one putative lipase gene and met the accepted quality threshold (genome completeness ≥ 90 % and contamination ≤ 10 %). These MAGs were selected as putative lipolytic MAGs (Appendix C.).

3.3.2. Lipolytic potential: Whole metagenome vs MAGs

A total of 31,570,310 protein sequences in the whole metagenome were found, but only 6,710,896 had known functions. Among the proteins with known functions, there were only 903 sequences with (putative) lipolytic activity (EC number of 3.1.1.3). The putative lipolytic MAGs contained 78 different classes of the total lipase genes (Figure 3-2).

By contrast, there were numerous genes coding for the extracellular enzymes that degrade proteins, carbohydrates, short-chain lipids, and phosphates in both the whole metagenome and MAGs, respectively (Table 3-3).

Table 3-3. Comparison between the number of extracellular hydrolytic enzymes in the whole metagenome and MAGs.

Enzyme class	Number in the whole metagenomes	Number in the MAGs
Proteases	135,456	1,272
Phosphatases	91,147	764
Carbohydrate degraders	47,893	663
Esterases/phospholipases	6,189	200
Lipases	903	78

Three most abundant genes for degrading sugars in the whole metagenome were *β-galactosidase*, *β-glucosidase*, and *α-galactosidase*. However, in the putative lipolytic MAGs, *β-glucosidase*, *β-hexosaminidase*, *cellulase*, *α-amylase*, *α-galactosidase*, and *endo-β-xylanase* were the most abundant (Appendix B). The large difference in the number of the genes can indicate that cells might have various alternative gene regulatory systems for expressing the genes which are involved in degrading sugars than lipases. Bacteria have a global regulatory mechanism known as carbon catabolite repression (CCR). In the presence of easily accessible carbon sources like sugars, CCR inhibits the expression of genes that allow cells to use a secondary carbon source (Görke and Stülke, 2008). One of the key genes in this process is catabolite repression resistance gene, known as the *phosphotransferase system sugar specific EII component (PTS-EII)* or *putative sugar kinases*. These genes were present in all putative lipolytic MAGs (Appendix D).

The CCR regulatory system for selecting the most suitable carbon source is aligned with economic theories (Allison and Vitousek, 2005). In the presence of simple substrates, cells do not invest carbon (C) and nitrogen (N) for producing extracellular enzymes that decompose complex substrates. However, where C and N resources exist in complex form, producing the relevant enzymes becomes inexpensive (Allison and Vitousek, 2005). For lipases, where glucose is abundant, CCR depresses its production (Boekema *et al.*, 2007). In addition, the

expression of proteases depresses the lipase production (Andersson, 1980; Black and DiRusso, 2003). In *Bacillus subtills,* the accumulation of amino acids induced the cells to produce more proteases and depress the lipase expression (Eggert *et al.*, 2003).

Moreover, about 20% of lipases were putative (pattern-filled in Figure 3-2), which means some of the lipases could in fact be esterases. The only way to determine the activity of these putative genes would be to express them synthetically. In terms of class, the most abundant lipase in the whole metagenome was "Lipase 1", while "Lipase 2" dominated the good bins. In fact, there was no consistent patterns in terms of the abundance of different classes between the MAGs and the whole metagenome. It is worth mentioning that Prokka uses several databases, e.g., *ISfinder*, *UniProtKB* and *National Center for Biotechnology Information* (*NCBI) Bacterial Antimicrobial Resistance Reference Gene Database*, for annotation and lipase genes might have different names in each database. The meaningful functional annotation of genes is challenging since the enzymes are predicted based on the homology of the sequences rather than biochemical features. For instance, annotation tools assigned lipolytic functions to genes due to the presence of consensus sequences like *GXSXSXXG* (G: glycine, X: any amino acids, S: serine) while no lipase with this sequence has shown lipolytic activity in lab yet (Ali *et al.*, 2012). The challenges of lipase classification with different tools based on the motifs they possess is discussed in Chapter 5.

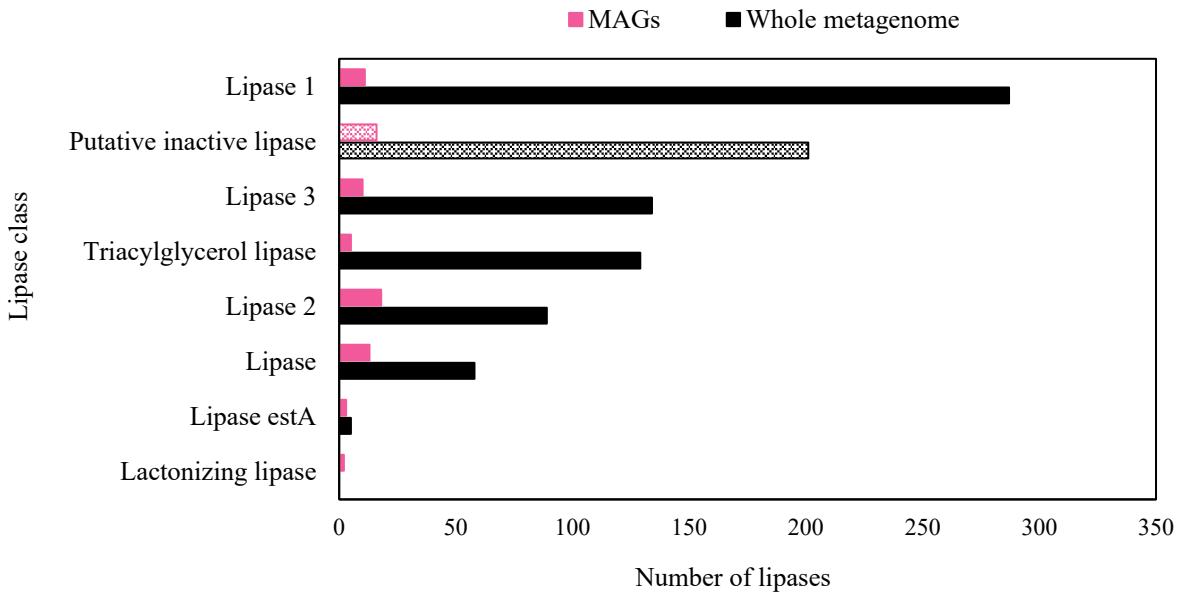


Figure 3-2. Abundance of different classes of the lipases: Comparison of the whole metagenome data and putative lipolytic MAGs.

3.3.3. MAGs taxonomical assignment vs classification

At both phylum and class level, *GTDB-Tk* assigned all 40 putative lipolytic MAGs to taxa. At lower taxonomic ranks (e.g., order, family, genus, species); however, a few MAGs remained unassigned (Appendix E). Particularly, at species level, where except for one MAG, all MAGs were unassigned. Unassignment to a taxon according to the *GTDB-Tk* at a certain level means that either the genome represents a novel species or that a species assignment could not be reliably established (cerebis, 2017).

Notwithstanding that *GTDB-Tk* assigned all MAGs at phylum and class level to taxa, some of them remained unclassified even at those two levels. The only taxon-assigned and classified MAG at all levels was *Bin 481*, a sulphate-reducing bacterium (Appendix E). Curiously, two of the MAGs, *Bin 684*, and *Bin 820*, did not have a classification at phylum level. However, within the *NCBI* database, these two MAGs are in the class *Deltaproteobacteria* and the phylum *Proteobacteria*. *Deltaproteobacteria* have been proved to be polyphyletic (Yarza *et al.*, 2014) and need to be reclassified (Parks *et al.*, 2018). Modern databases like the *GTDB* unlike the *NCBI*, standardize the MAG's classification by forming a tree from a large number (~120) ubiquitous single-copy proteins and calculating the relative evolutionary divergence (*RED*) values (Parks *et al.*, 2018). Therefore, within the *GTDB*, *Deltaproteobacteria* is no longer in the phylum *Proteobacteria* nor classified yet as a distinct phylum. Details about the count of ubiquitous proteins and *RED* values in each MAG are presented in Appendix C. and Appendix F, respectively.

3.3.4. Linking the lipolytic MAGs to the taxa

Putative lipolytic MAGs belonged to 14 distinct phyla (mostly from the *Actinobacteria*, *Proteobacteria* and *Bacteroidota*), with two unclassified MAGs only to the phyla level (Figure 3-3).

In the phylum *Actinobacteria*, except for the Bin 205 which was not assigned to any genera, 5 distinct classified (*Mycolicibacterium, Corynebacterium, Propionicimonas, Austwickia* and *Rhodoluna*) and 3 unclassified (67-14, IMCC26207, UBA10799) genera existed. All of these genera may have facultatively anaerobic species. Komatsu *et al.* (2019) isolated *Mycolicibacterium peregrinum* from a pig farm and showed that this species has an anaerobic respiration with genes involved in lipid and fatty acid metabolisms. Another facultative anaerobe in this genus is *Mycolicibacterium toneyamachuris* (Kuge *et al.*, 2020). Based on

Bergey's Manual of Systematics of Archaea and Bacteria, both *Corynebacterium (Bernard and Funke, 2015),* and *Propionicimonas* (Ueki *et al.*, 2015) have several facultatively anaerobic species. Most species of *Austwickia*, and *Rhodoluna* are still unknown. But the former may be facultatively anaerobe. (Kagia and Liu, 2014).

Some of the putative lipase genes were found from genera which were not expected to be lipolytic or indeed in anaerobic reactors (such as aerobic autotrophs). The putative lipolytic MAGs were classified into three categories: i) a possible MAG with a lipase gene but no *fadL* gene to transport long-chain fatty acids; ii) a true lipid degrader: a MAG with both lipase and *fadL* genes; and iii) a miscellaneous lipid degrader: a MAG that degrades lipids for other purposes like denitrification, polyhydroxyalkanoates (PHAs) accumulation/degradation or invasion of other bacteria's outer membrane (Appendix G). The fourth possibility is that these are mis-assemblies or mis-annotations (Kunin *et al.*, 2008). Even high-quality MAGs can be subject to these misinterpretations.

None of the MAGs were labelled with certainty as a possible or true lipid degrader due to both non-universality of *fadL* gene and mis-assembly/mis-annotation possibility. For Gram-positive bacteria, still no universal known long-chain fatty acid transporter protein like the *fadL* in Gram-negatives, is characterised (Salvador López and Van Bogaert, 2021). Hence, it was not possible to decide which of the putative Gram-positive lipolytic MAGs (13 from the phylum *Actinobacteria* and 2 from the *Firmicutes_A*) are a true lipid degrader. Also, in putative Gram-negative lipolytic MAGs, only 2 out of 18 (*Bin 967* and *Bin 1501,* respectively, represented *Rhodoferax* and an unclassified genus from *Syntrophorhabdia* class in *Desulfobacterota* phylum) had both lipase and *fadL* gene. The absence of *fadL* in the rest of the 16 MAGs might be because of the mis-assembly and mis-annotation.

Additionally, the co-presence of lipases and other genes in the MAG, like the essential denitrification genes, or genes required for synthesizing or degrading PHAs, was assumed to be a sign of miscellaneous lipid degrader.

One of the most curious lipolytic MAGs was *Bin 22*, a possible *Nitrosomonas*. The presence of a lipase gene in this genome seemed redundant as *Nitrosomonas* are aerobic nitrifiers, and classically utilize carbon dioxide as a carbon source (Cheremisinoff, 1995; Brandt *et al.*, 2017). However, some species like *Nitrosomonas europaea* are facultative anaerobes (Abeliovich and Vonshak, 1992) and some have even shown denitrification activity under anaerobic conditions

(Ward, 2008). The link between lipolysis and denitrification has been shown in some studies. Denitrifying bacteria utilize long-chain fatty acids in the absence of light in anaerobic reactors, (Mackie *et al.*, 1991) and anaerobic denitrifiers like *Acidovorax caeni sp. nov.* have lipase activity (Heylen *et al.*, 2008).

Besides, PHA production/degradation is linked to lipolysis as well. Bacteria that accumulate PHA, either produce lipases to degrade oily substrates and obtain carbon to store PHA (Tufail *et al.*, 2017) or degrade the intracellular PHA when the carbon is limited (Mitra *et al.*, 2020). *Nitrosomonas* has been proposed as a PHA-producing bacterium (Yang *et al.*, 2013; Yin *et al.*, 2018). Previous reports have suggested that many bacteria, including denitrifiers, produce lipases rather than polymerases to degrade PHAs, though the reason is not known (Jaeger *et al.*, 1995; Muhammadi *et al.*, 2015; Wang and Chu, 2016; Chu and Wang, 2017; Sharma *et al.*, 2019). Therefore, either anaerobic condition or the presence of PHA or other bacteria might induce the lipase expression. The assimilation of long-chain fatty acids, such as palmitic acid, in anaerobic conditions represses ammonia-oxidation activity of nitrifiers (Juliette *et al.*, 1995). The presence of global nitrogen regulatory gene (*ntcA*), which existed in *Bin 22,* can activate the assimilation of other nitrogen sources if ammonium/NH_4^+ is absent (Lee *et al.*, 1999). Also, when *Nitrosomonas sp. Is79* was co-cultured with *Nitrobacter winogradskyi*, the abundance of periplasmic lipases in its proteome increased (Sedlacek *et al.*, 2016).

One possible explanation for the presence of lipase in the *Bin 22* therefore might be that it represents an uncharacterised facultative *Nitrosomonas* species that use the lipase for denitrification and PHA production/degradation.

Topologically the closest species to *Bin 22* was *Nitrosomonas sp003201565* (Appendix F) deposited in the protein database of *NCBI* as *Nitrosomonas sp. Nm84* (accession number: QJJP01000015). This genome, from a pure culture, not only had the lipase and *fadL* genes, but like *Bin 22*, it contained the essential denitrification genes including *nirK* (Copper-containing nitrite reductase), *norB* and *norC* (nitric oxide subunit B and C) (Braker *et al.*, 2000; Torregrosa-Crespo *et al.*, 2017). However, they both lacked the PHA synthesising genes. On the other hand, 13 putative lipolytic MAGs from phyla *Proteobacteria* and *Actinobacteria* had either only PHA synthesizing genes (e.g., *PhaC*) or both PHA synthesizing and denitrification genes (Appendix G). Therefore, *Bin 22* might use the lipase for degrading the PHA produced by other bacteria from these two phyla for denitrification.

Similarly, the other 10 MAGs from several phyla that only had denitrification and lipase genes (no PHA synthesising genes) might use the lipase for degrading the PHA produced by others.

Furthermore, potential genes involved in the export of lipases to the extracellular medium was searched for *Bin 22* to validate the presence of lipase genes. Gram-negative bacteria use both *Type I* and *Type II secretion system* for exporting lipases (Ahn *et al.*, 1999; Hausmann and Jaeger, 2010). *Type I* secretion pathway usually involves the expression of *ATP-binding Cassette (ABC)* transporters consisted of *ABC* proteins, membrane fusion proteins (MFP) and outer membrane proteins (OMP) at the upstream of the lipase gene. In addition to this, the lipase gene itself should contain several conserved glycine-rich motifs of *GGXGXD* (G, glycine, X, any amino acid, D, Aspartic acid) known as *LARD*/lipase *ABC* transporter recognition domain at the C-terminal (Chung *et al.*, 2009). Nonetheless, none of the aforementioned export genes or motifs were found in *Bin 22* or in the associated public genome of *Nitrosomonas sp. Nm84*. Only one of the related lipases (accession number PXW86082) in the public genome had the motifs at the C-terminal.

There were also 16 lipase containing MAGs (those with known genus were all facultative anaerobes) that had no denitrification nor PHA synthesizing genes. For most of them it is not known what the exact role of lipases are. For example. *Chlorobium* in *Bin 803* are photosynthetic green sulphur-reducing bacteria. This MAG, however, had both dark-operative protochlorophyllide reductase (*BChl*) and light-harvesting antenna/chlorosomes (*csmA*) genes that enable *Chlorobium* to survive at extremely low light conditions (Frigaard *et al.*, 2003). Two *Chlorobium* species in *NCBI* had also lipase genes but no *fadL* genes including *Chlorobium limicola* (accession number KUL20464) and *Chlorobium phaeobacteroides DSM 26* (accession number ABL66324). *Desulfobacter postgatei* (*Bin 481*), a sulphate reducing bacteria in *NCBI* had *fadL* gene but no lipase gene. It is not known whether or not this bacterium is a cheater, but uptake of long-chain fatty acids and improved lipid degradation have been confirmed for other sulphate reducers (Alves *et al.*, 2020; Florentino *et al.*, 2020)

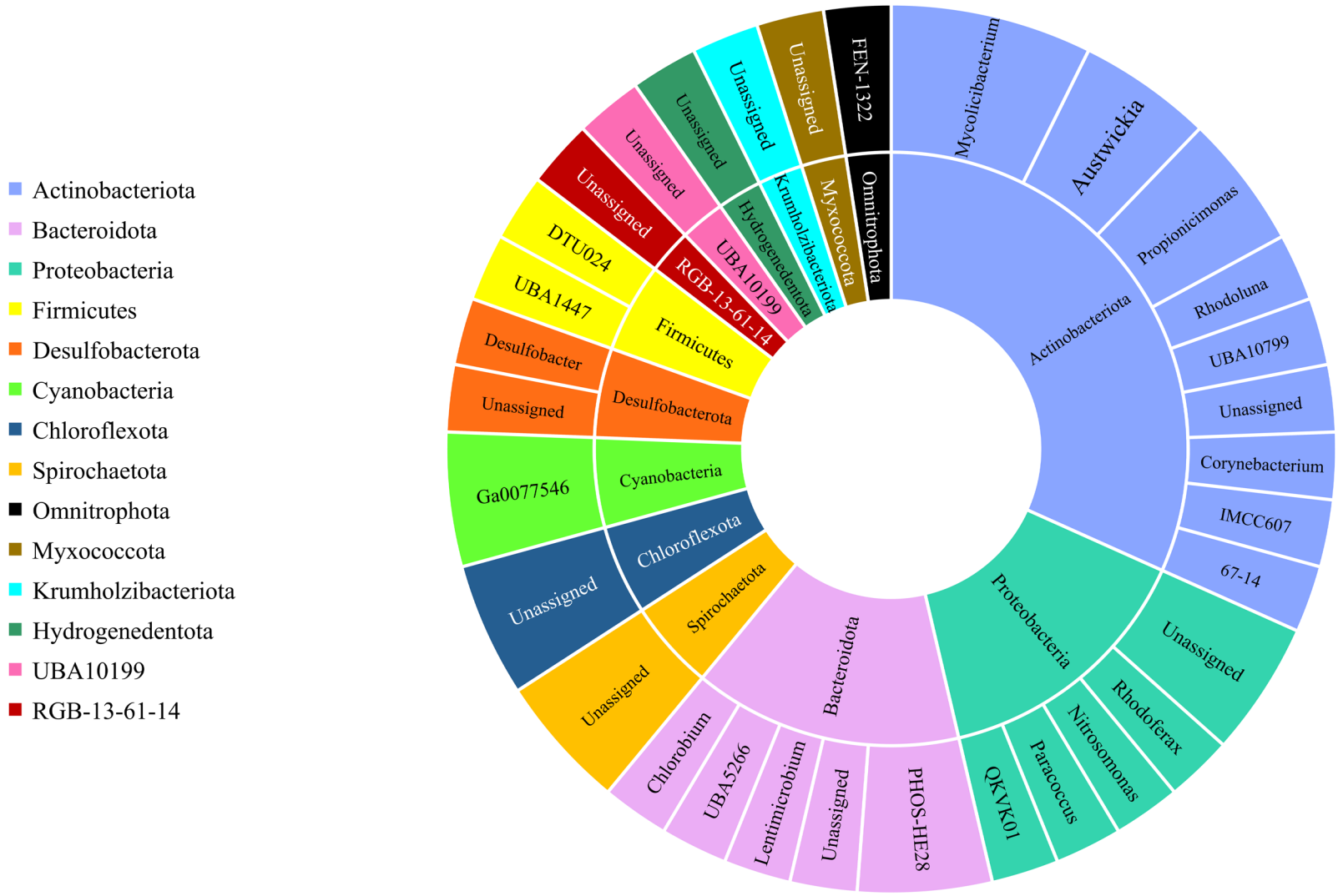


Figure 3-3. Taxonomic classification of putative lipolytic MAGs at phylum and genus level using GTDB-Tk (Size of each wedge presents number of identified MAGs in each phylum).

3.3.5. Linking the putative lipolytic MAGs to reactor conditions and lipases

How each MAG is associated with different reactor conditions is presented in Appendix H. For most putative lipolytic MAGs, the number of mapped reads per reactor conditions did not vary significantly. However, for a few MAGs, statistically significant differences were observed (Appendix I-L). For instance, for temperature, at 4°C, only *Bin 803* (*Chlorobium*) and at 15°C, *Bin 328* (Unclassified *Ga0077546* from Cyanobacteria), *Bin 231* (Unassigned from *Chloroflexota*), *Bin 154* (Unassigned from *Hydrogenedentota*), and *Bin 609* (Unclassified *FEN-1322* from *Omnitrophota*) had noticeably higher number of mapped reads.

Whereas, considering only the effect of feed treatment, *Bin 22* (*Nitrosomonas*), *Bin 367* (*Lentimicrobium*), *Bin 428* (*Austwickia*), and *Bin 231* (*Unassigned* from *Chloroflexota*) had significantly higher number of reads mapped to the sterile condition, while *Bin 803* (*Chlorobium*), and *Bin 328* (Unclassified *Ga0077546* from *Cyanobacteria*) to the Non-sterile. In case of the phase of sampling, except for the *Bin 790* (Unclassified UBA10799 from *Actinobacteriota*) which was statistically higher in the liquid phase, *Bin 1001,* and *Bin 328* (Unclassified Ga0077546 from *Cyanobacteria*), *Bin 481* (*Desulfobacter postgatei*), and *Bin 609* (Unclassified FEN-1322 from *Omnitrophota*) were higher in the biofilm.

About 55% of the lipases were in MAGs from the phylum *Actinobacteriota* of which half distributed within two genera, *Mycolicibacterium* and *Corynebacterium*. Both genera existed at both temperatures, treatment, and phase, though the latter was slightly (but not statistically significant) higher in the liquid phase (Appendix H).

Regardless of their class/taxonomy lipases from the different MAGs were significantly different in length (pairwise Tukey test, P-value = 0.002). One-way Analysis of variance, ANOVA, (pairwise Tukey test, P-value = 0.467) on the length of individual lipases per phylum showed that *Actinobacteriota* had both the largest (819 aa, amino acid) and the shortest (180 aa) lipases. In addition, the highest and the lowest average length of the lipases were within the phyla *Actinobacteriota* (399 aa) and *Omnitrophota* (220 aa), respectively (Figure 3-4).

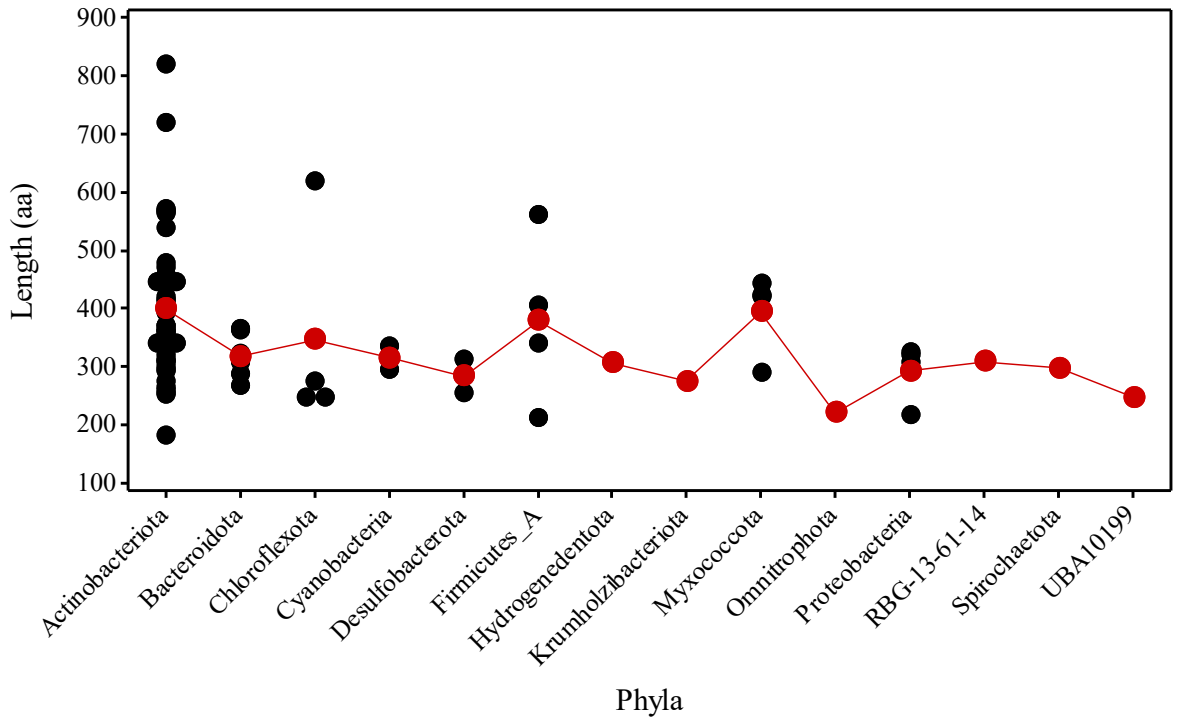


Figure 3-4. Distribution of the length for individual lipases per phylum (● show the length of individual lipases and (● shows the average length of all lipases in a certain phylum (One-way ANOVA, Minitab 18, P-value= 0.467).

3.3.6. Can temperature affect the length of the lipases?

Proteins produced by extremophiles are expected to have a shorter or longer length (Riley *et al.*, 2008). At extreme conditions, cells minimize their investment in C and N resources for protein synthesis such that they are stable in that condition. Kananavičiūtė *et al.* (2020) discussed that collagen-like proteins that thermophilic bacteria produce have shorter length than their mesophilic counterparts. One-way ANOVA on the length of the lipases from the significant putative lipolytic MAGs (MAGs with the highest mapped reads, but not statistically, from either 4°C or 15°C reactors) showed that there is no correlation between the size of the lipases and the temperature of the reactors (Appendix M). Lipases from the 4°C reactor had higher average length than the 15°C reactor though the difference was not statistically significant (Figure 3-5). Protein size is mostly associated to biochemical structure and biological function (Tiessen *et al.*, 2012). For instance, the core hydrophobicity of amino acids which affects the protein folding is temperature dependent; the lower the temperatures, the lower the hydrophobicity of amino acids (van Dijk *et al.*, 2015). One study has shown that the membrane proteins in cold-adapted bacteria are not different in terms of protein length with their mesophilic counterparts (Kahlke and Thorvaldsen, 2012). Riley *et al.* (2008) compared

the length of thermophilic enzymes with that of the mesophilic counterparts and found no difference.

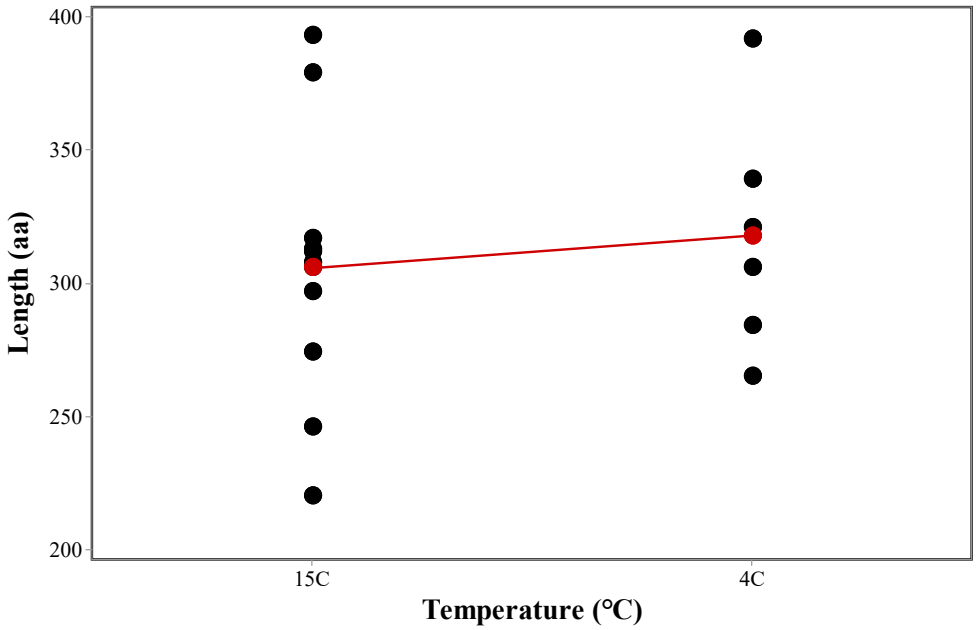


Figure 3-5. Comparison of the lipase length in significant putative lipolytic MAGs from 4 °C and 15 °C (One-way ANOVA, Pairwise Tukey test, P-value=0.637, Minitab 18) the list of the selected MAGs is in Appendix M.

3.3.7. Who is abundant in each reactor?

In all reactor conditions, bacteria were dominant and constituted between 81-90% of the microbial community. Archaea and viruses respectively had the relative abundance of 2-8 % and 3-13% (Figure 3-6). Viruses had their highest abundance at Sterile-15°C and the lowest at Non-sterile- 4°C. By contrast, the archaea were the highest at Non-sterile-15°C and the lowest at Sterile-4°C. However, statistically temperature and treatment did not have a significant effect on the relative abundance of the viruses or archaea (Two-way ANOVA, P_{value}~1).

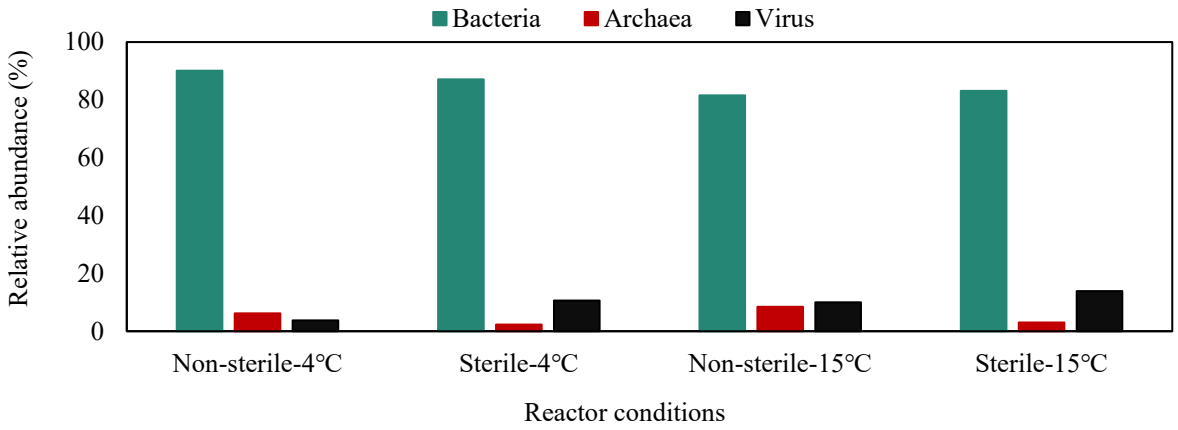


Figure 3-6. The relative abundance of three kingdom (●) Bacteria, (●) Archaea and (●) Viruses in each reactor condition.

Also, the richness and evenness of microbial community (Figure 3-7) were not statistically different per reactor conditions (Table 3-4 and Figure 3-8).

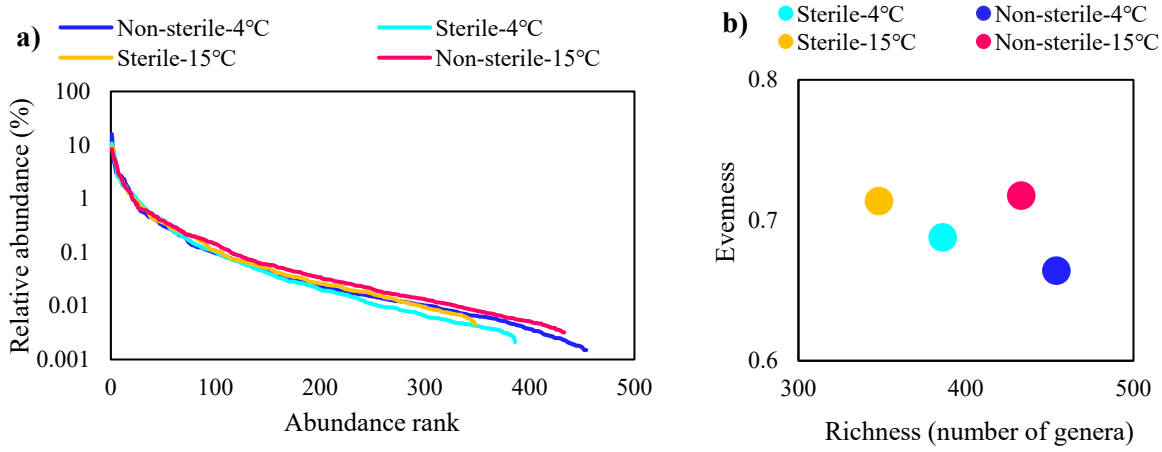


Figure 3-7. a) Rank abundance curve (Whittaker plot) for genera at different reactor conditions b) Richness and evenness of genera at different reactor conditions calculated based on the Shannon diversity index.

Table 3-4. P-values for the ANOVA (Minitab 18) on richness and evenness of genera in all reactors considering the effect of temperature and treatment.

Parameters	P-value	
	Treatment	Temperature
Richness	0.07	0.18
Evenness	0.60	0.21

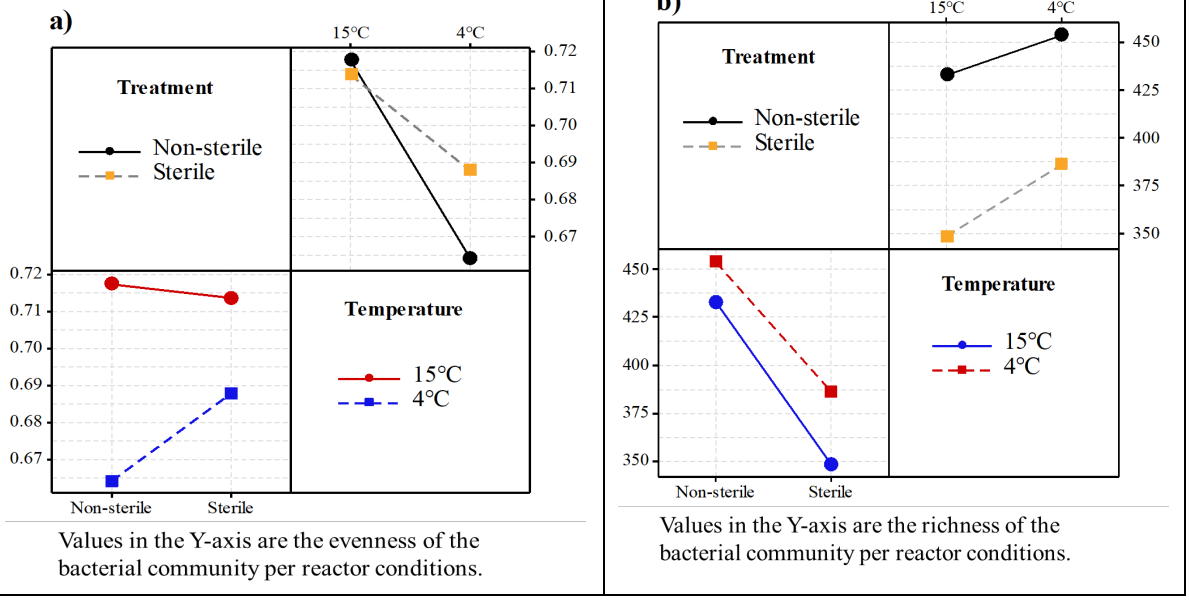


Figure 3-8. Interaction plot: Effect of temperature and treatment on a) evenness and b) richness of the microbial community in all reactor conditions. The Y-axis values in the 'plot a' are the evenness values of the bacterial community and the Y-axis values in the 'plot b' are the richness of the bacterial community per reactor conditions.

There were 32 common bacterial genera with relative abundance of more than 1% in at least one reactor conditions (Figure 3-9). For instance, *Acinetobacteria* had only 1% relative abundance at Non-sterile-4°C and in other conditions they were less than 1%. Only ten of the

common genera had the relative abundance of more than 1% at all conditions. Also, 7 common species (≥1%) were present in all reactors (Figure 3-10).

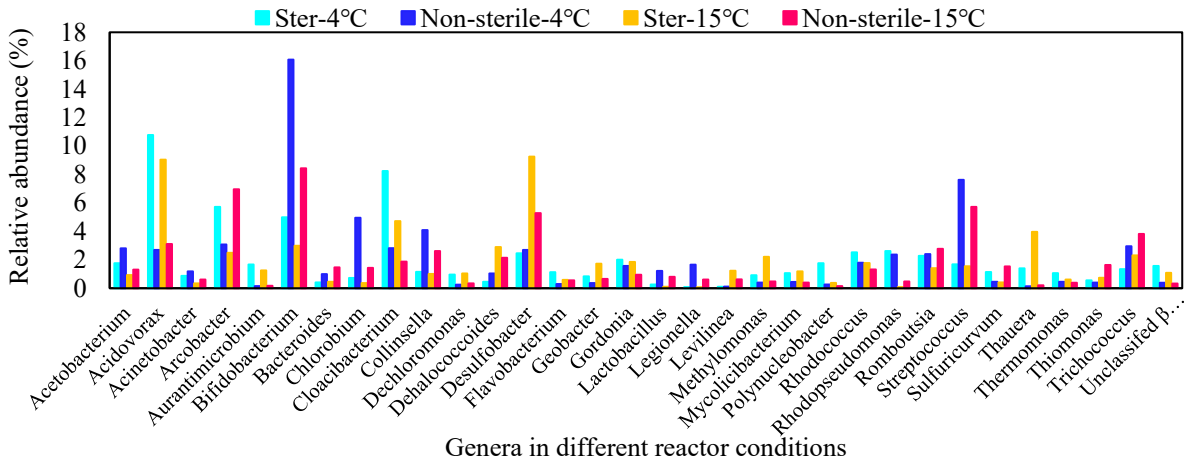


Figure 3-9. Common genera with more than 1% relative abundance in at least one of the reactor conditions identified by GOTTCHA2.

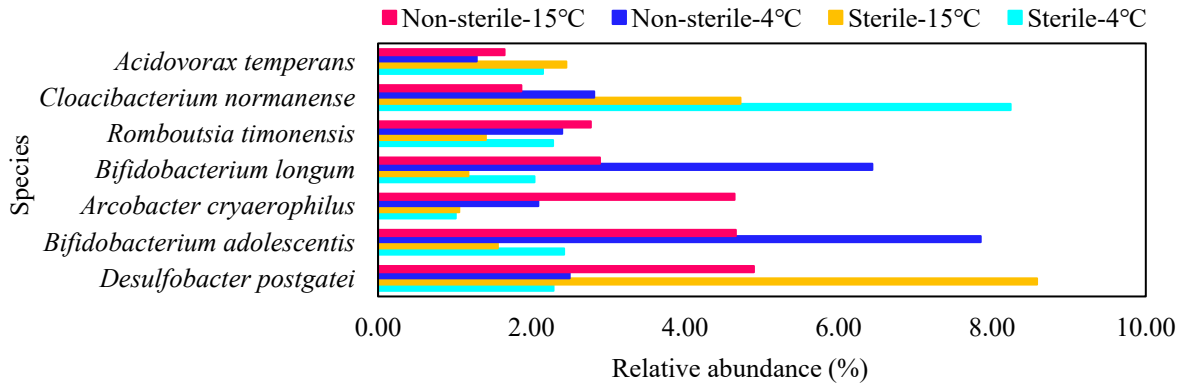


Figure 3-10. Common species with more than 1% relative abundance in all reactor conditions, identified by GOTTCHA2.

For most common genera (≥1%), the effect of temperature and treatment on relative abundance was insignificant (Appendix N). However, *Bifidobacterium* and *Desulfobacter* were more abundant at 4°C and 15°C, respectively. Similarly, a significant effect of treatment was noticeable among the Sterile and Non-sterile fed reactors for *Bifidobacterium*, *Streptococcus*, *Acidovorax* and *Cloacibacterium*. The first two were higher in Non-sterile conditions whereas the second two were the highest at the Sterile conditions.

Three of the genera recovered in MAGs had more than 1% relative abundance in reactors but not in all conditions. Except for *Desulfobacter*, only *Chlorobium* and *Mycolicibacterium* had more than 1% abundance at Non-sterile and Sterile conditions, respectively (Figure 3-9). The rest of the genera identified in lipase containing MAGs had very low relative abundance (*Corynebacterium*, *Lentimicrobium*, *Nitrosomonas*, *Paracoccus* and *Rhodoferax*). Also,

GOTTCHA2 found no relative abundance for three of the MAGs (*Austwickia, Propionicimonas, and Rhodoluna*) in any reactors (Figure 3-11) due to bioinformatics tools limitation. Compared to the *GTDB-Tk*, used for the MAGs taxonomic classification, *GOTTCHA2* might have used another database for taxonomic classification that lacked the genome of these three genera.

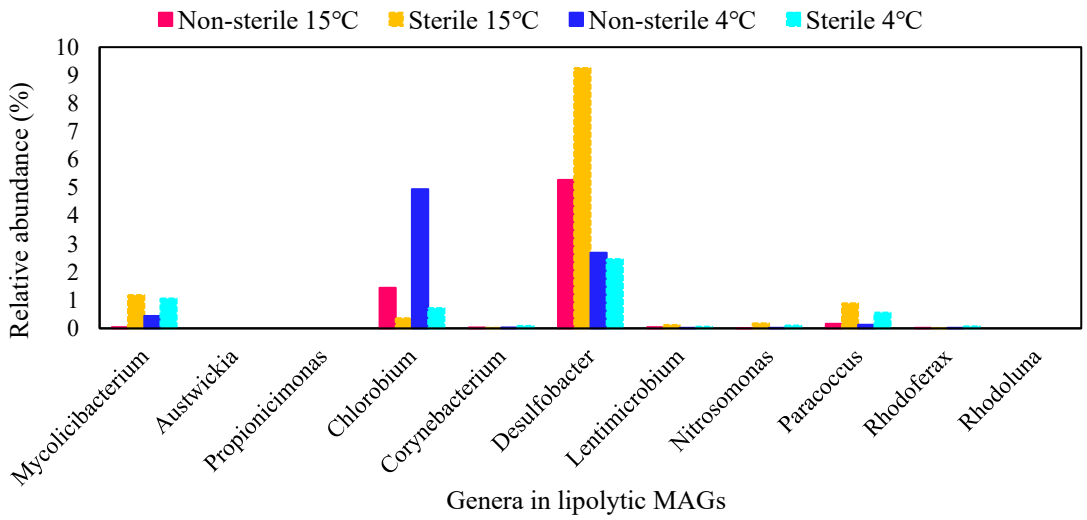


Figure 3-11. Relative abundance of the genera recovered in MAGs in the reactors.

As the empirical cumulative distribution function plot (3-parameter loglogistic distribution) illustrated, more than 95% of the genera at all reactor conditions had a relative abundance below 1% (Figure 3-12). Given that only three of the lipolytic MAGs, of which only *Mycolicibacterium* had more than one lipase genes, were among the 5% most abundant genera inside the reactors, we can infer that the potential lipase producers were not the dominant population.

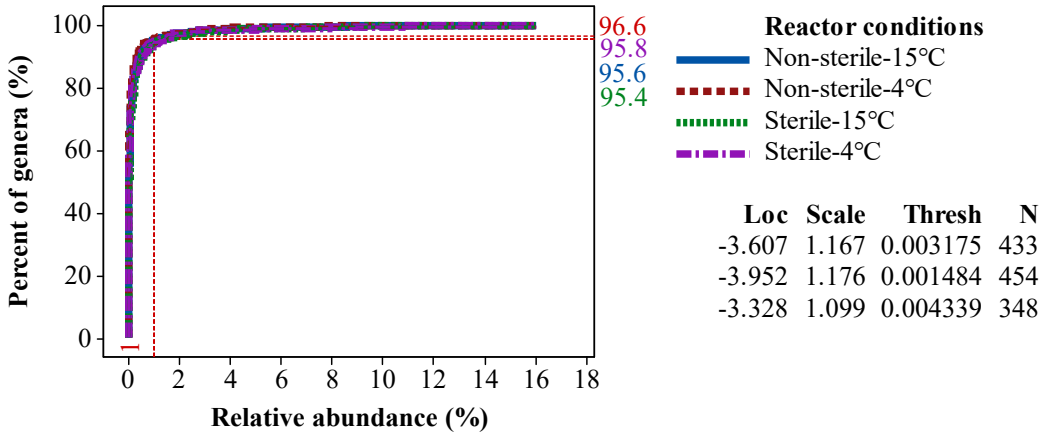


Figure 3-12. The empirical cumulative distribution function plot (3-parameter loglogistic distribution) for the abundance of genera at different reactor conditions. Loc: Location parameter, Thresh: Threshold parameter, N: number of data (genera).

3.4. Conclusion

Lipolysis is not always associated with exogenous lipid degradation. PHA accumulation/ degradation and invasion of other bacterial outer membrane might be linked to lipid degradation and possessing lipase genes on the genome. Lipases compared to other hydrolytic extracellular enzymes were lower in numbers in both whole metagenomic data and putative lipolytic MAGs. Most lipases in the recovered putative lipolytic MAGs, belonged to the phyla *Actinobacteria* and genera *Mycolicibacterium* and *Corynebacterium*. The only lipolytic MAG with known classification at all levels was a sulphate reducing bacteria, *Desulfobacter postgatei*. The relative abundance of most genera (95%) in all reactors was below 1% and *Desulfobacter* along with *Chlorobium*, and *Mycolicibacterium* were the only recovered lipolytic MAGs that were present at all reactor conditions with more than 1% relative abundance. This indicates that the population of bacteria that have the potential to ferment lipids is much lower than other fermentative bacteria.

With few exceptions, there was no significant correlation between the reactor conditions and the number of reads mapped to the MAGs. Also, temperature had no significant role on lipase length.

Chapter 4 : Can we find expressed lipases by metaproteomics?

4.1. Introduction

Proteins are products of gene expression and are responsible for all functions that (micro) organisms do. Proteomics and metaproteomics are molecular biology tools that allow us to study the proteome or metaproteome of a single microbe or a microbial community and know their actual function at a certain time.

The general steps involved in bottom-up or shotgun metaproteomics are: i) protein extraction and downstream processing; ii) digestion or cleavage of proteins to peptides; iii) separation by liquid chromatography; iv) mass spectrometry; and v) bioinformatics.

Despite recent advances, metaproteomics is still in its infancy with important bottlenecks in protein extraction and computational data analysis. Protein extraction methods vary depending on the nature of the sample and protein location. Proteins from samples that do not contain impurities like humic substances (e.g., fresh water) that require specific extraction procedures (e.g. using phenol) are easier to extract (Heyer *et al.*, 2019). Intracellular proteins need harsher conditions (strong acids or bases, mechanical methods) for lysing the cells and releasing them. By contrast, for extracellular proteins milder extraction procedures should be employed to avoid cell lysis and yet maintain a high yield (Speda *et al.*, 2017).

For environmental samples, the main target of metaproteomics are usually the extracellular proteins/enzymes. Extracellular enzymes hydrolyse large impermeable organic molecules and allow the cells to take up the constituents as food. These enzymes can be part of the EPS that some members of the community excrete into their extracellular medium. Therefore, all the methods developed for the EPS extraction are applicable for the extraction of extracellular enzymes too. Yet, none of the suggested protocols for EPS extraction is unbiased (Seviour *et al.*, 2019). In wastewater samples, extracellular enzymes attach to microbial flocs by hydrophobic and ionic forces. Ionic agents like cation exchange resins (CER) and hydrophobic agents like Triton can break these forces and release the enzymes with minimum cell disturbance (Frølund *et al.*, 1996; Gessesse *et al.*, 2003).

After extraction, the proteins can be further purified via a precipitation step or subjected to gel electrophoreses for further fractionation prior to mass spectrometric analysis. There are two main approaches for mass spectrometry. The main difference between them lies in proteolytic cleavage of proteins into peptides through enzymatic digestion. In the bottom-up approach, the

mass spectrometry is performed on peptides. By contrast, in top-down approaches, intact proteins are subjected to mass spectrometry. This latter method usually suffers from the unknown mass of the intact proteins due to post-translational and degradation processes. Hence, top-down protocols are not suitable for complex samples and the bottom-up approach in combination with scanning methods (every peptide above an intensity threshold gets fragmented) is preferred for environmental samples. Yet assembling peptides and assigning them into a certain protein is also challenging; redundant, homologous or isobaric peptides might belong to several proteins or even different species in a given metagenome (Hettich *et al.*, 2013).

Peptides get separated via liquid chromatography, typically according to their hydrophobicity (e.g., using reverse phase chromatography), before entering a mass spectrometer. All mass spectrometers have three main components: an ion source that converts peptides to ions; a mass analyser that selects ions based on their mass-over-charge ratio (m/z); and a detector that measures the number of ions at each m/z ratio (Han *et al.*, 2008). Each of these components are available in different models. For instance, two most common ion sources are electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI). The ESI turns peptides to positively charged ions by forcing them through an orifice while the MALDI ionize peptides with the aid of a laser. Most frequently used mass analysers are also quadrupole, times-of-flight (TOF), ion trap and Orbitrap that can be combined in tandem mass spectrometry for more accurate measurement (Schuchardt and Sickmann, 2007).

Tandem mass spectrometry or MS/MS employs two mass spectrometers that perform the scanning in two modes. In the data dependent acquisition (DDA) mode, the first spectrometer selects the peaks with the highest signal which belong to the most abundant peptides called precursors. Precursors are further fragmented in the collision cell and scanned in the second mass analyser. By contrast, in the data independent acquisition (DIA) mode, there is no precursor prioritization, and all the peptides get fragmented (Canterbury *et al.*, 2014). Fragmentation process is necessary because some peptides are chemically different but have similar molecular weight or m/z which make the identification difficult. Each fragmentation approaches would generate different pairs of ions that either retain N-terminus (labelled as a, b, c ions) or C-terminus (labelled as x, y, z ions) end. The presence of different ion pairs can make the peptide mass calculation and result interpretation complex.

The ideal fragmentation is to break the amide/peptide bonds between the carboxyl and amine groups of amino acids (C-N cleavage) and generate b/y ion pairs. These types of ions are dominant in low energy dissociation processes like collision induced dissociation (CID) or very similarly higher-energy collisional dissociation (HCD) or collision-activated dissociation (CAD). In electron-based activation methods like electron capture dissociation (ECD) or electron transfer dissociation (ETD), the cleavage can occur between the alpha carbon and amine groups (C_α-N) too, resulting in generation of c/z ions. By contrast, in high energy activation methods like ultraviolet photodissociation (UVPD), the bond between the carboxyl group and the alpha carbon (C_α-C) breaks and a/x ions form in addition to b/y and c/z ions. Besides, there is a possibility for cleavage of multiple bonds or secondary fragmentation of ions which results in formation of internal ions too (R Julian, 2017).(Julian, 2017). Conventionally, internal fragment ions (e.g., a/x, a/y, a/z, b/x, b/y, b/z, c/x, c/y, and c/z) were regarded as disturbance and excluded from the data analysis as they could not be reliably assigned to mass spectra. Newer research though is trying to include them to increase the protein sequence coverage (Zenaidee *et al.*, 2020).

The output format of mass spectra varies for each instrument and can be both open and proprietary. Nonetheless the data are represented either as continuous (profile-mode) or centroided/peak-picked (peak list) spectra, containing the intensity and m/z of ions for each scan. Some instruments like AB SCIEX provide the raw data as .wiff and .Wiff.Scan including the metadata and spectra, respectively (Deutsch, 2012). The different formats are convertible by free tools like ProteoWizard.

The computational analysis of mass spectra and identification of peptides and proteins is one of the most challenging part of metaproteomics. Unlike proteomics that deals with proteomes of single species cultures, in metaproteomics many proteins from complex microbial communities are present. Most popular tools like MaxQuant work well for single species proteomics but when installed on common desktop computers those tools may struggle to analyse metaproteomics data with very large metagenomics sequence databases. Moreover, using very large databases requires multi-round search or pre-filtering approaches and often suffers from reduced sensitivity (leaving many false negatives). Unfortunately, there is still no standardized metaproteomics processing pipeline for analysing complex microbial communities (Kleikamp *et al.*, 2020).

Current bioinformatics approaches in (meta) proteomics are classed as database search, de novo sequencing, and a combination of both. In the database search approaches, search engines, like Andromeda (in MaxQuant), SEQUEST, Mascot, X! Tandem, and MS-GF, are used for correlating the theoretical and experimental masses of peptides. Theoretical peptide mass estimations are usually obtained from an *in silico* digested target database (i.e., metagenomics). However, many of these peptides are spurious and unlikely to be produced *in vivo/vitro* (Li *et al.*, 2016). Usually, a decoy database is also generated from the target database, i.e., by inverting the peptides. The decoy database controls the false positive hits and contains all the peptides that cannot exist in the sample *in vivo/vitro*. Peptide matching for large target databases usually results in lower number of significant hits. In larger databases, more spurious peptides are present and since the decoy database is larger too, it is more likely to find high-scored false positive hits at a fixed false discovery rate (FDR) (Jeong *et al.*, 2012; Kumar *et al.*, 2017).

By contrast, in *de novo* sequencing, amino acid sequences are directly extracted from the MS/MS spectra either by using the graph theory or considering the fragmented ions without using any target databases. These methods are particularly advantageous for finding novel proteins that do not have known sequences or post translational modifications. However, *de novo* sequencing only yields good results with the high-resolution spectra and issues like poor peptide fragmentation, peptide ion series directionality, and cleavage abnormalities in spectra can make the data analysis challenging (Hughes *et al.*, 2010).

This chapter contains an analysis of the metaproteome of a cold-adapted microbial community from the AnMBRs for which metagenomes were discussed in Chapter 3. The aim of the chapter is to find all expressed extracellular lipases or other marker proteins (like long-chain fatty acid transporters) at different reactor conditions.

4.2. Material and Methods

4.2.1. Protein extraction, precipitation, and separation

Wastewater samples were taken from both biofilm and bulk liquid of the AnMBRs as described in chapter 3. Before protein extraction, volatile suspended solid (VSS) was measured following the standard method of the American Public Health Association (Clesceri *et al.*, 1996). Proteins were extracted from the EPS using the protocol suggested for the extraction of extracellular lipases by both Gessesse *et al.* (2003) and Frølund *et al.* (1996). In brief, the combination of

CER and Triton X-100 was used, details of which is described in Appendix O. The extracted proteins were quantified by Pierce™ Modified Lowry Protein Assay Kit, Thermo Fisher Scientific prior to precipitation by the phenol/chloroform method (Wessel and Flügge, 1984). Precipitated proteins were solubilized and reduced in Laemmli buffer and β-mercaptoethanol, sonicated (20 min, cool temperature) and heated (5 min, 60 °C) before being run on one-dimensional sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), for 5 min at 120 V (Bio-Rad Mini-PROTEAN®). The gel was stained following the protocol of Bio-Safe Coomassie Brilliant Blue G-250 and was destained overnight. All the downstream processes details as well as the gels pictures are also included in Appendix O. In-gel digestion and mass spectrometry were done at NUPPA, Newcastle University Protein & Proteome Analysis centre following the protocol detailed in Appendix P.

4.2.2. Data analysis

Mass spectrometric raw data were converted to mgf files using *MSConvert* and analysed as a single group using *PEAKS Studio X* using a High-Performance Computing Windows workstation. The metagenomics protein sequence database was cleaned for sequence redundancy and annotation errors using *CD-hit* and *notepad++*. Furthermore, the database search using the cleaned metagenomics constructed database was performed using a two-round search strategy. The initial search allowed 50 ppm parent ion and 0.1 Da fragment mass error tolerance and carbamidomethylation as fixed modification. Protein matches of the initial search with a -10lgP protein score greater or equal to 20 were collected, which resulted in a preliminary search output of 11814 protein groups. The second-round search, using the refined database from the first-round search, allowed up to 3 missed cleavages, 50 ppm parent ion and 0.1 Da fragment mass error tolerance, carbamidomethylation as fixed modification, oxidation and deamidation as variable modifications and employed a decoy fusion database for determining false discovery rates. Peptide spectrum matches were filtered against 1% or 5% FDR, and protein identifications with 2 or more unique peptides across the group were considered as significant matches. Processing of metadata was done using *MATLAB 2017b*. Additional taxonomic and *Kyoto Encyclopedia of Genes and Genomes (KEGG)* number annotations was performed using *GhostKOALA (V. 2.2)*.

4.3. Results and discussion

4.3.1. VSS concentration

VSS varied significantly among samples from different reactor conditions (Table 4-1). Generally, samples from Non-sterile, 4 ℃ and Biofilm conditions had significantly higher VSS. The P-values of two-way ANOVA on VSS data are presented in Appendix Q. Interaction plots showed that for samples taken from the liquid phase, the VSS concentration did not vary considerably at both treatments and temperatures. By contrast, for biofilm samples, the VSS was significantly higher at 4 ℃ and Non-sterile conditions. However, at 15 ℃, the VSS of both Sterile and Non-sterile conditions were not significantly different (Figure 4-1).

Table 4-1. Average concentration of volatile suspended solids at different reactor conditions, reported errors are standard error of measurement from three replicates.

Conditions	VSS (g/l)
Sterile-4 ℃-Biofilm	58.65 ± 12.92
Sterile-4 ℃-Liquid	19.13 ± 2.97
Non-sterile-4 ℃-Biofilm	96.08 ± 7.02
Non-sterile-4 ℃-Liquid	10.85 ± 3.35
Sterile-15 ℃-Biofilm	44.48 ± 0.42
Sterile-15 ℃-Liquid	18.33 ± 9.30
Non-sterile-15 ℃-Biofilm	50.47 ± 10.07
Non-sterile-15 ℃-Liquid	18.88 ± 6.65

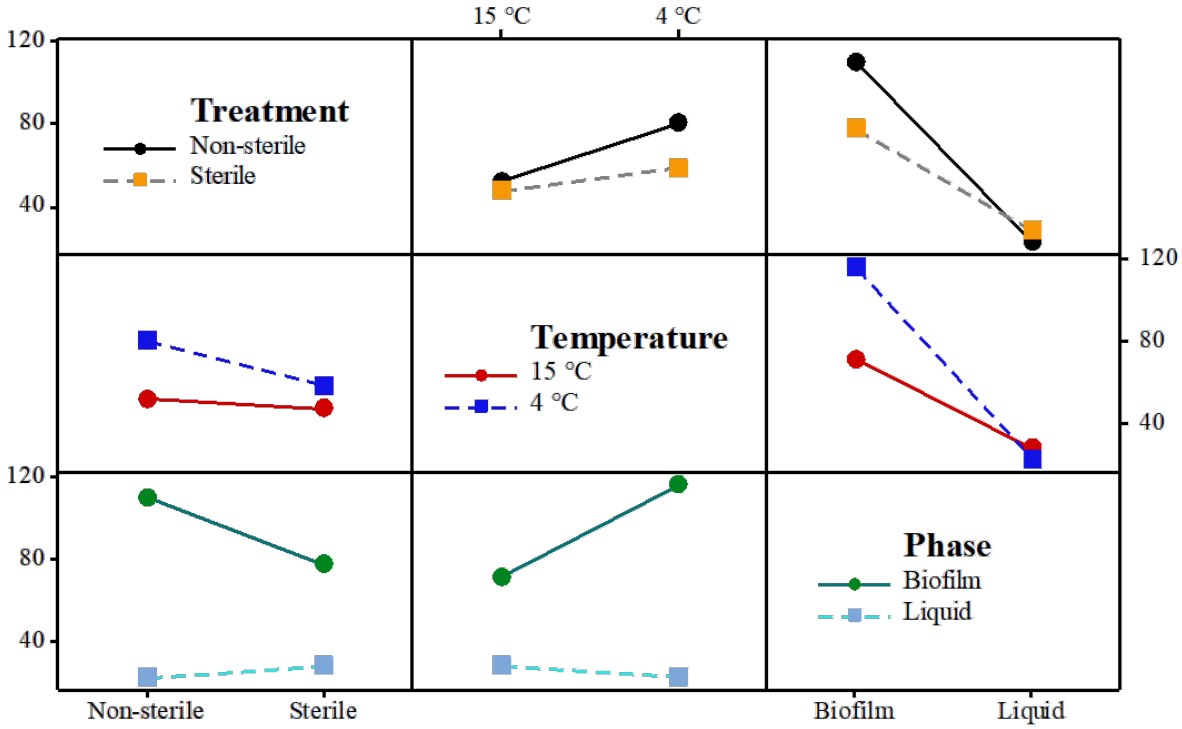


Figure 4-1. Interaction plot for VSS concentration (g/l) at different reactor conditions (treatment, phase, and temperature), Minitab 18. The Y-axis values are VSS concentration (mg/l).

4.3.2. Protein quantification

The average concentration of proteins in the extracts varied from 749 µg/ml to 1161 µg/ml (Table 4-2). Even though samples taken from the 15 ℃ and Non-sterile conditions had higher concentrations, the difference was not statistically significant (Appendix R).

Table 4-2. Concentration of extracted proteins in supernatant for different reactors.

Conditions	Average concentration of proteins (µg/ml)
Sterile-4 ℃-Biofilm	803 ± 310
Sterile-4 ℃-Liquid	992 ± 86
Non-sterile-4 ℃-Biofilm	1066 ± 280
Non-sterile-4 ℃-Liquid	749 ± 200
Sterile-15 ℃-Biofilm	799 ± 110
Sterile-15 ℃-Liquid	942 ± 337
Non-sterile-15 ℃-Biofilm	1161 ± 286
Non-sterile-15 ℃-Liquid	1134 ± 129

4.3.3. Expressed proteins: Are there any lipases?

A total of 93 and 117 distinct protein classes were found at FDR 1% and 5%, respectively as listed in Appendix S, using the complete metagenomics constructed database. However, proteins of the same class had different accession numbers (coming from different genes in the target database) and therefore the actual number of identified protein groups at both FDR were 256 and 329, respectively.

At FDR 5%, there were 24 new protein classes compared to FDR 1% though neither of the new or common hits were significantly different in number (P-value=0.514, one-way ANOVA, Minitab 18). Not only were none of the hits lipases, but also none were other hydrolytic enzymes. Jachlewski *et al.* (2015) have also reported that for *Archaea*, EPS extraction and subsequent mass spectrometry, did not result in identification of extracellular enzymes like *lipases*, *proteases*, *glucosidases*, *esterases*, and *phosphatases* though enzymatic assays had confirmed their activity. This might be due to the low concentration of these enzymes in the extracellular medium which is still not detectable through SDS-PAGE.

About 75% of the identified proteins were involved in processing the genetic information, signalling and cellular processes, processing environmental information and energy metabolism. Further 4%, 2%, and 1% of the proteins were related to carbohydrate, amino acids, and lipid metabolism, respectively (Figure 4-2).

In terms of class, *outer membrane porin proteins* (omp32) outnumbered the rest of the classes (25 %) and after them in descending order there were *vitamin B12 transporters (btuB), TonB-dependent starch-binding receptors (susC)* and *major outer membrane proteins P. IA (porA).*

The results further revealed the presence of several *porins*, ABC transporters like *lamB* (*Maltoporin*) and *fadL* (long-chain fatty acid transporters), of which the latter is particularly of interest. The expression of *FadL* might be related to the expression of lipases. It was assumed that cells would only invest on expressing *fadL* genes when expressed lipases had already released *long-chain fatty acid transporters* from the lipidic molecules.

Also, *cytoplasmic proteins* were present including *groEL (60 KDa chaperonin), tufA (elongation factor Tu), fusA (elongation factor G), rpsA (30S ribosomal protein S1), rpsC (30S ribosomal protein S3), rpsE (30S ribosomal protein S5), rpsG (30S ribosomal protein S7)* and *rpsP (30S ribosomal protein S16).* The presence of these proteins in the EPS is not odd and is related to either the presence of extracellular vesicles in the EPS or cell lysis that happens during the biofilm maturation (Lee *et al.*, 2008; Jachlewski *et al.*, 2015).

Among the proteins profiled, there were several proteins that are typically found in the extracellular vesicles including *outer membrane proteins* and *porins* like *ompA, ompW, ompX ompF, porA* and *porB*. Other proteins like *acrA (Multidrug efflux pump subunit)* release toxic compounds and attack the competing bacteria. *ABC transporters (fadL, lamB, btuB)* and *TonB-dependent receptors (susC)* act as nutrient sensors and transporters under nutrient limited conditions (Lee *et al.*, 2008).

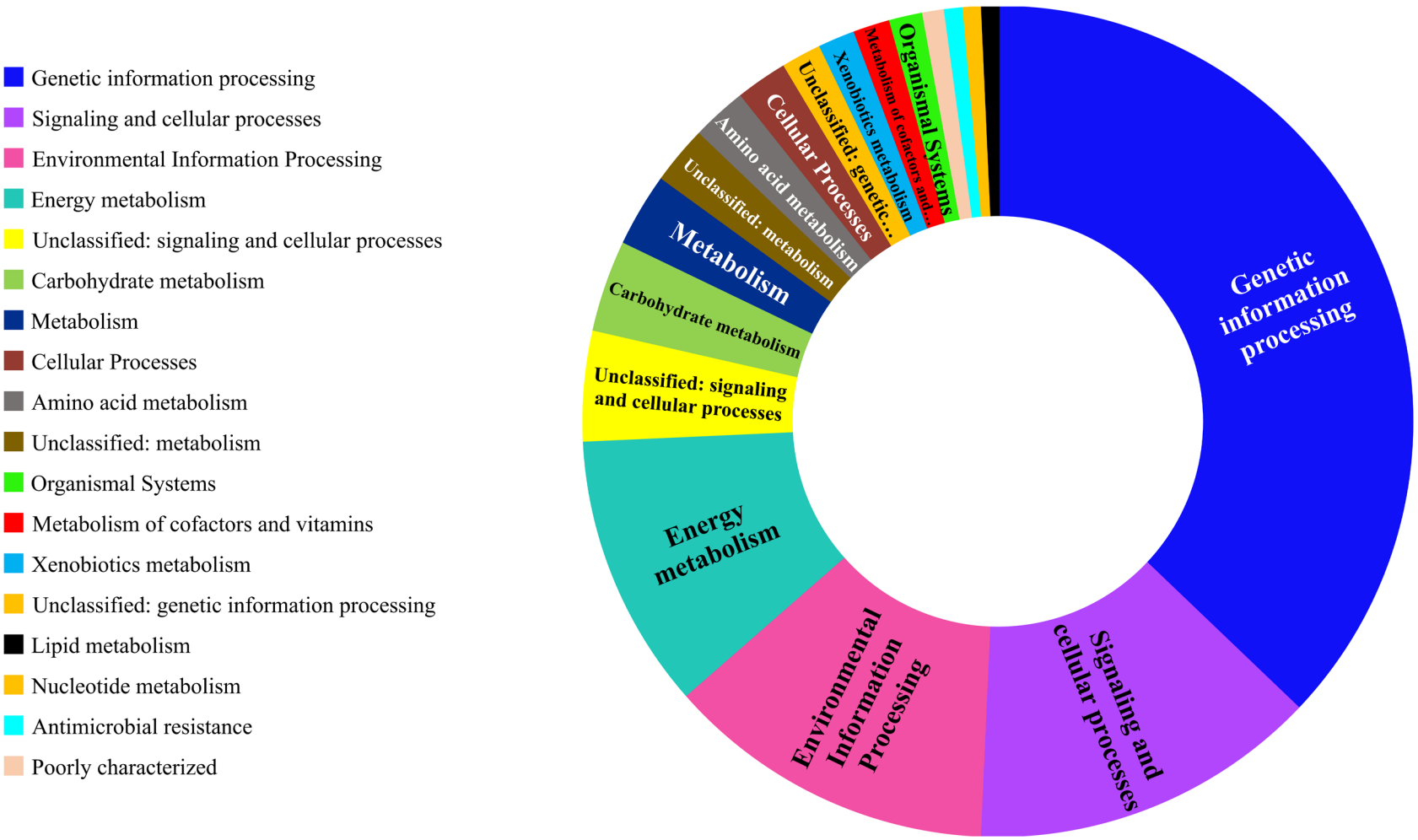


Figure 4-2. Functional classification of identified proteins at FDR 5% based on KEGG database.

4.3.4. Taxonomical distribution of identified proteins by metaproteomics

About 97% of the expressed genes (FDR=5%) were related to the bacterial domain and at least from 19 distinct class (Figure 4-3-a), among which *Betaproteobacteria* had the greatest share (57 %). The top-ranked identified genera with expressed proteins were all from class *Betaproteobacteria* including *Paucimonas*, *Dechloromonas*, *Acidovorax*, *Azoarcus* and *Thauera*, respectively (Figure 4-3-b). The full list of all genera associated to the expressed proteins is presented in Appendix T. Among the top-ranked, all genera except for *Paucimonas* have been formerly identified by *GOTTCHA2* (see Chapter 3) and their relative abundance in each reactor was known (Figure 4-4). Comparatively, *Azoarcus*, was the only low-abundant genera with no abundance at Non-steril-4°C.

Although *Paucimonas* was absent from the reactors based on *GOTTCHA2*, it had the highest number of related expressed proteins (Appendix U). Most of them were ribosomal proteins or were involved in energy metabolism. One *porin* and one *outer membrane protein* were present too.

Notably, putative lipases identified by metagenomics were dominantly distributed among the Gram-positive bacteria. By contrast, metaproteomics mostly identified proteins that belonged to the Gram-negative bacteria. However, this is not curious, for the metaproteomics data only reveals those proteins which were extracted at the time of sampling.

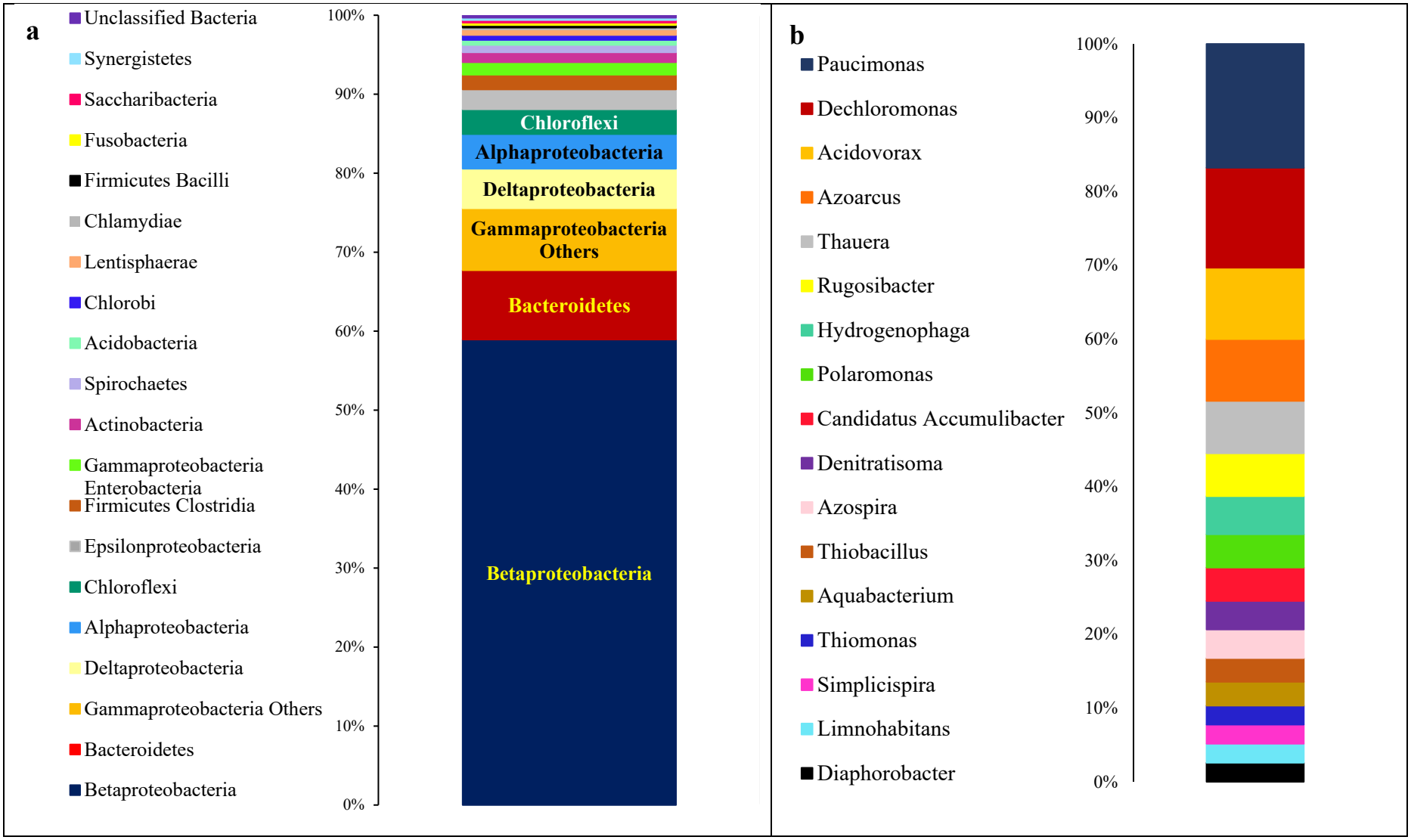
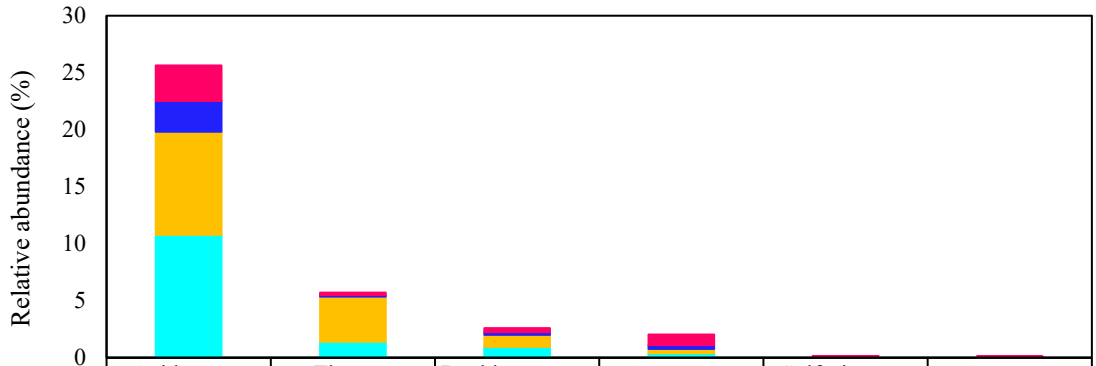


Figure 4-3 a) Taxonomic distribution of expressed proteins at class level, b) list of genera that had more than three expressed proteins (FDR=5 %).



	Acidovorax	Thauera	Dechloromonas	Aeromonas	Sulfurimonas	Azoarcus
Non-sterile-15°C	3.11	0.21	0.34	0.93	0.01	0.01
Non-sterile-4°C	2.70	0.14	0.25	0.34	0.03	0.00
Sterile-15°C	9.04	3.97	1.04	0.38	0.06	0.07
Sterile-4°C	10.78	1.40	0.96	0.38	0.05	0.04

Figure 4-4. Relative abundance of top-ranked genera per reactors

By contrast, 68% of the related proteins to other genera were *porins* and *outer membrane proteins* (Appendix V). Additionally, both *Dechloromonas* and *Azoarcus* had long-chain fatty acid transporters *(fadL)* and thus were potentially lipolytic. Expressed *fadL* was also found in two other genera (not among the top-ranked), *Aeromonas* and *Sulfurimonas*. Their relative abundance is presented in Figure 4-4. *Azoarcus* and *Sulfurimonas* were low-abundant, in all conditions, whereas *Aeromonas* had higher relative abundance ~ (1%) at Non-sterile-15°C.

The expression of *fadL* in *Dechloromonas*, *Azoarcus*, *Aeromonas* and *Sulfurimonas* implies the presence of long-chain fatty acids in the system and therefore can be a proxy for lipolysis performed by these genera or others. However, none of these four genera were recovered as putative lipolytic MAGs by metagenomics. The absence of lipases along with the presence of *fadL* genes in a genome might be indicative of cheating mechanisms. Nonetheless, the complete genome of these four genera in *NCBI* had both the *fadL* and lipase genes. While this might remove the "cheating label", from these genera, it does not necessarily make them true lipase producers either. We do not know whether or not *fadL* and lipases are coregulated, but we do know that both can be exported through extracellular vesicles in Gram-negative and Gram-positive bacteria (Galka *et al.*, 2008; Lee *et al.*, 2008; Lee *et al.*, 2009a; Lee *et al.*, 2016b; Hong *et al.*, 2019). The presence of both *fadL* and lipases in the extracellular vesicles might have an entirely different reason than the lipolysis of exogenous lipid molecules. For instance, Galka *et al.* (2008) have shown that pathogens transport lipases as a virulence factor through extracellular vesicles to attack the lipidic membrane of the host cell and deliver lipids to them. The same scenario might apply to bacterial cells interaction, but no study has shown this yet.

Moreover, about 3% of the expressed genes (FDR=5%) were from *Archaea*. The presence of archaeal proteins within the extracted EPS of bacteria is not surprising as the same extraction procedures can be applied for both (i.e. CER) and most biofilms contain *Archaea* as well (Jachlewski *et al.*, 2015).

The *Archaea* were all from the phylum *Euryarchaeota,* and the identified proteins (80%) were mostly related to the genus *Methanothrix* (Figure 4-5, a). *Methanoregula* and *Pyrococcus* were the other two genera, and both had only one associated protein. The identified protein for *Pyrococcus* was the *tubulin-like protein (CetZ)* (Figure 4-5, b)) that controls the shape of archaeal cells (Duggin *et al.*, 2015) and for *Methanoregula* was *5,10-methylenetetrahydromethanopterin reductase (mer)* which is involved in methane metabolism (K00320) pathways. Also, all eight *Methanothrix* related proteins (Figure 4.5, b)) were either involved in energy metabolism (i.e., *V-type ATP synthase subunit C*) pathways or methane metabolism (i.e., *Carbon monoxide dehydrogenase/acetyl-CoA synthase subunit alpha*).

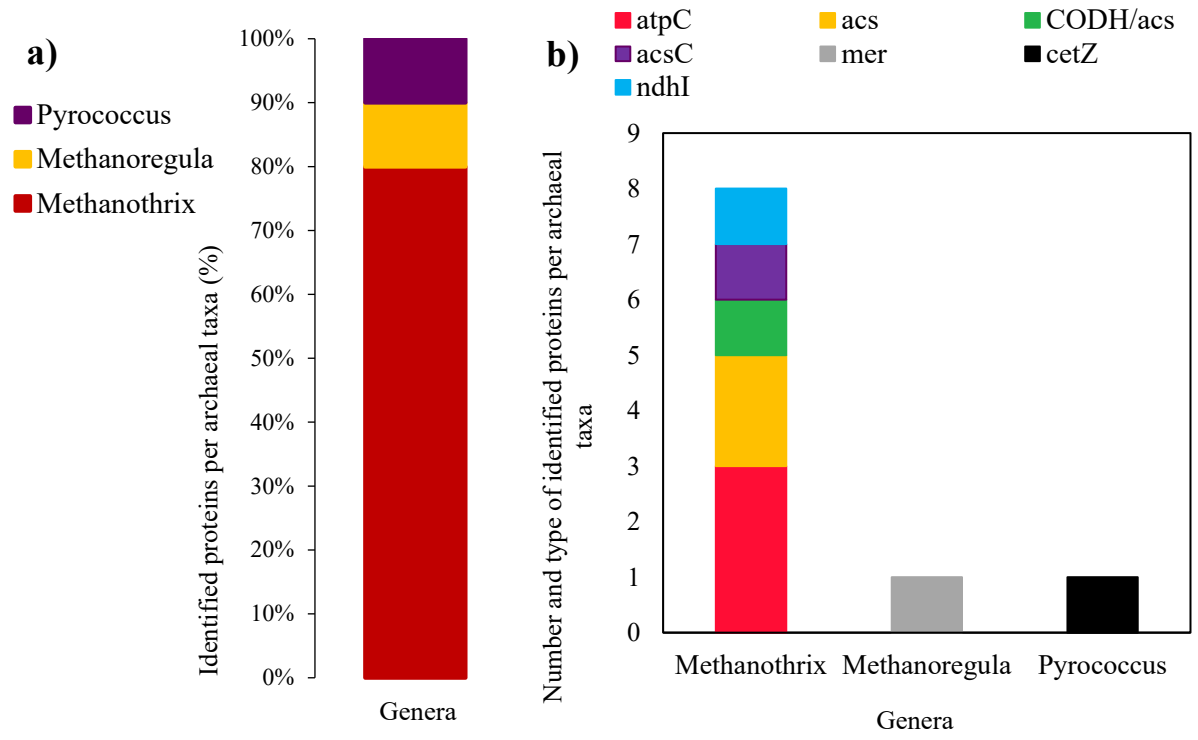


Figure 4-5. Archaeal expressed proteins (FDR= 5%) a) Taxonomic distribution at genus level (percentage) b) Associated genes/proteins, atpC= V-type ATP synthase subunit C, acs= Acetyl-coenzyme A synthetase, CODH/acs= Carbon monoxide dehydrogenase/acetyl-CoA synthase subunit alpha, acsC= Corrinoid/iron-sulfur protein large subunit, mer= 5,10-methylenetetrahydromethanopterin reductase, cetZ= Tubulin-like protein, ndhI= NAD(P)H-quinone oxidoreductase subunit I chloroplastic.

4.3.5. Identified proteins of abundant genera

Out of the 32 common bacterial genera with relative abundance of more than 1% (Figure 3-9), through metaproteomics, we have identified proteins expressed by 15 of them (Table 4-3). More than half (55%) of the proteins were outer membrane proteins and porins. Curiously, some of these genera accumulate lipids, e.g., PHAs. Lipid-accumulation is a barrier for lipid degradation in wastewater systems (Chipasa and Mdrzycka, 2008). Cold temperature is a stimulator for PHA accumulation (Srivastava *et al.*, 2020).

At least six of the identified genera including *Acinetobacter* (Hauschild *et al.*, 2017), *Cloacibacterium* (Ram *et al.*, 2018), *Dechloromonas* (Oshiki *et al.*, 2008), *Rhodopseudomonas* (Carlozzi and Sacchi, 2001), *Thauera* (Oshiki *et al.*, 2008; Singleton *et al.*, 2021), and *Thermomonas* (Coats *et al.*, 2016) are involved in PHA accumulation. However, *Dechloromonas* was the only genera that had expressed *fadL* gene (no lipases). This genus, in the activated sludge plants, have been identified as an anaerobic denitrifier too (Singleton *et al.*, 2021). The expression of *norC (Nitric oxide reductase subunit C)* and *actP* (Cation/acetate symporter) confirms its denitrification activity and competition with methanogens to assimilate acetate (Table 4-3). Other denitrifiers like *Thauera* and *Acidovorax* which were previously found by metagenomics, were present. *Thauera* enter the anaerobic digester (in the activated sludge plants) from the biofilms formed on walls of sewers (Cyprowski *et al.*, 2018).

Also, the presence of sulphur-reducing bacteria, *Sulfuricurvum* (Table 4-3) along with the sulphate-reducers, e.g. *Desulfobacter*, has been associated to the occurrence of internal sulphur cycle in the system (St. James and Richardson, 2020). Although *Desulfobacter* was not identified by metaproteomics, it was recovered as a good lipolytic MAG and had high abundance at all reactor conditions (Figure 3-9). Sulphate reduction limits PHA-accumulation, and sulphate-reducers in the absence of sulphate can switch to syntrophic and fermentative metabolisms.

Table 4-3. Expressed proteins found from the common genera (≥1% relative abundance) per reactor conditions in Figure 3-11.

Class	Genus	Name	Function	Quantity
Actinobacteria	Aurantimicrobium	rpoD	RNA polymerase sigma factor	1
Alphaproteobacteria	Rhodopseudomonas	omp2b	Porin	1
Bacteroidetes	Cloacibacterium	susC	TonB-dependent receptor	1
		Putative Omp	Putative outer membrane protein	1
Bacteroidetes	Flavobacterium	Putative Omp	Putative outer membrane protein	1
Betaproteobacteria	Polynucleobacter	ompW	Outer membrane protein W	2
Betaproteobacteria	Thauera	omp32	Outer membrane porin protein 32	7
		rplA	50S ribosomal protein L1	2
		pckG	Phosphoenolpyruvate carboxykinase [GTP]	1
		dmdC	3-methylmercaptopropionyl-CoA dehydrogenase	1
Betaproteobacteria	Thiomonas	omp32	Outer membrane porin protein 32	3
		ilvC	Ketol-acid reductoisomerase (NADP(+))	1
Betaproteobacteria	Unclassified betaproteobacterium CB	omp32	Outer membrane porin protein 32	1
		ompW	Outer membrane protein W	1
Betaproteobacteria	Dechloromonas	Putative Omp	Putative outer membrane protein	2
		atpA	ATP synthase subunit alpha	1
		atpD	ATP synthase subunit beta 1	1
		atpF	ATP synthase subunit b	2
		actP	Cation/acetate symporter	1
		ompA	Outer membrane protein A	1
		ompP1	Outer membrane protein P1	1
		omp 47KDa	47 kDa outer membrane protein	2
		porA	Major outer membrane protein P.IA	3
		fadL	Long-chain fatty acid transport protein	1
		norC	Nitric oxide reductase subunit C	1
		gltA	Citrate synthase	2
		sdhA	Succinate dehydrogenase flavoprotein subunit	1
		fusA	Elongation factor G	1
Betaproteobacteria	Acidovorax	omp32	Outer membrane porin protein 32	10
		groL1	60 kDa chaperonin	1
		ompW	Outer membrane protein W	2
		SODB	Superoxide dismutase [Fe]	1

Class	Genus	Name	Function	Quantity
		fusA	Elongation factor G	1
Deltaproteobacteria	Geobacter	MDH	Malate dehydrogenase	1
Epsilonproteobacteria	Sulfuricurvum	btuB	Vitamin B12 transporter	
Gammaproteobacteria - Others	Acinetobacter	pagN	Outer membrane protein	1
		omp38	Outer membrane protein	2
Gammaproteobacteria - Others	Methylomonas	pmoB1	Particulate methane monooxygenase alpha subunit	2
Gammaproteobacteria - Others	Thermomonas	oar	Protein oar	1

4.4. Conclusion

This chapter aimed to correlate potential lipolytic genes found through metagenomics in Chapter 3 to the expressed lipases found by metaproteomics. Nonetheless, no expressed lipases or other hydrolytic enzymes were identified by metaproteomics.

Top-ranked protein classes were either outer membrane porins like *omp32* and *porA* or transporters such as *btuB* and *susC*. Taxonomically, most proteins were associated to genera *Paucimonas*, *Dechloromonas*, *Acidovorax*, *Azoarcus* and *Thauera* from the class *Betaproteobacteria*. Except for *Paucimonas*, the other four genera have been already profiled by metagenomics in Chapter 3 in all reactor conditions. Overall, metaproteomics identified 15 out of the 32 abundant (≥1%) common genera found by metagenomics per reactors. Interestingly, 6 of these genera can accumulate lipids/PHA that can limit lipid degradation.

Although no *lipase* was found, *fadL*, transporters that carry long-chain fatty acids through outer membrane of Gram-negative bacteria, were present and associated to genera like *Dechloromonas*, *Azoarcus*, *Sulfurimonas* and *Aeromonas* which were present in all reactor conditions. However, since complete genomes of these genera in *NCBI* had both lipase and *fadL* genes, we assumed them as potential lipase producers rather than cheaters. Moreover, some bacteria export *fadL* through their extracellular vesicles which might be independent of extracellular lipase regulation. Even lipases have been found in extracellular vesicles.

Metaproteomics is highly dependent on the accuracy, completeness, and size of the constructed metagenomics database. By using de novo approaches, this dependency can be reduced. Developing universal protein extraction protocols is also a game changing step in the future of metaproteomics. This is particularly important for extracellular hydrolytic enzymes. Furthermore, developing better computational tools that match mass spectra to peptide sequences more efficiently can improve metaproteomics data analysis notably.

Chapter 5 : On classifying lipases

5.1. Introduction

In this chapter different protein classification tools and databases for classifying the bacterial lipases identified in the putative lipolytic MAGs in Chapter 3 are discussed and evaluated.

Environmental microbiology research has progressed rapidly with the introduction of second and third generation sequencing technologies. (Meta)genome sequencing has generated millions of protein sequences that are now available in public databases. The major challenge is now to associate functions to these protein sequences. Most such sequences have not been characterized experimentally (Blum *et al.*, 2021).

Classifying such proteins experimentally is expensive and slow. There is therefore a need for automated classification tools to predict the attributes of a protein. Conventionally, automated tools like BLAST and FASTA annotated protein sequences based on sequence similarity searches. However, the functionality of these tools is limited by the search algorithms and the databases they use to search against. Newer tools use protein signature databases and multiple sequence alignments to find the highly conserved residues. This newer approach is more likely to identify divergent homologues (McDowall and Hunter, 2011). At present there are several protein signature databases that classify proteins with different approaches, including sequence clustering, regular expression, profiles, and hidden Markov models (HMM), as presented in Table 5-1.

Table 5-1. Databases which use protein signature for classification

Database name	Protein classification methods	Content	Latest version and update	Reference
PRODOM	Sequence clustering	Protein domains	2012.1/CG1803 Dec 2nd 2015	(Servant *et al.*, 2002)
PROSITE	Regular expression/ Profiles	Protein domains, families, and functional sites	2021_03 Jun 2nd 2021	(Sigrist *et al.*, 2013)
PRINTS	Fingerprints	Composite conserved motifs	v. 42.0 Feb 2nd 2012	(Attwood *et al.*, 2003)
Pfam	Hidden Markov models (HMM)	Protein families	v. 33.1 March 2021	(Mistry *et al.*, 2020)
TIGRFAMs	Hidden Markov models (HMM)	Protein families	v. 15.0 Sep 16th 2014	(Haft *et al.*, 2001)
PANTHER	Hidden Markov models (HMM)	Protein families and functionality	v. 16.0 Dec 18th 2020	(Mi *et al.*, 2020)
SUPERFAMILY	Hidden Markov models (HMM)	Structural protein domains with evolutionary relationship	v. 2 2019	(Pandurangan *et al.*, 2018)

For example, *PRODOM* clusters the proteins that have highly similar regions (homologous). This classification approach is good for detecting new domains within the uncharacterized proteins. (McDowall and Hunter, 2011). The *PROSITE* database (Sigrist *et al.*, 2013) uses regular expression or patterns. These are short and highly conserved motifs corresponding to residues with important functions or structures like the enzyme's active sites or substrate binding sites and exclude the less-conserved regions or whole domains (Hulo *et al.*, 2007; Sigrist *et al.*, 2013). *PROSITE* also uses profiles to reduce the high rate of false positive and false negative matches. Profiles are scoring matrices giving weight to amino acids and their positions. They are more tolerant of amino acid changes and sequence length differences, and hence can identify both conserved and divergent regions (Attwood and Mitchell, 2019). Nonetheless, only a limited number of proteins have a profile in *PROSITE*.

PRINTS database uses fingerprints, a group of motifs with unique inter-relationships that together could be used for diagnosing a protein family (McDowall and Hunter, 2011; Attwood and Mitchell, 2019).

Databases that use HMMs include *Pfam*, *TIGRFAMs*, *PANTHER*, and *SUPERFAMILY*. HMMs are similar to profiles and model both divergent and conserved regions (McDowall and Hunter, 2011) except that they use probabilities rather than absolute scores for amino acid's position (Attwood and Mitchell, 2019). Hence, they provide a better quality and a rapid access for protein classification.

Of those databases that use HMMs *Pfam* is the most popular for annotating the novel genomes and metagenomes. It covers about 75.1% of the *UniProtKB* (the universal protein knowledgebase) reference proteome and 49.4% of its residues. The latest update (version 33.1) has 18,259 families (Mistry *et al.*, 2020).

The *European Bioinformatics Institute* (*EBI*) integrated some of the signature databases into one by introducing *InterPro*.

InterPro gathers information from 13 member databases including *Pfam*, *PROSITE*, *SUPERFAMILY*, *PANTHER*, *PRINTS*, and *TIGERFAMs* (Blum *et al.*, 2021). By uploading/pasting the *FASTA* format of protein sequences in the search box, *InterProScan* (protein scanning software) searches the query proteins against *InterPro* and reports the existing biological information from each member database. *UniProtKB* also uses *InterPro* to annotate its protein sequences (UniProt, 2021).

Most of the aforementioned databases have little information about conserved lipolytic motifs (short, conserved patterns with a distinct function in a protein sequence) and families. However, lipases share three folds in their structure which has been used to classify them into three superfamilies: i) *alpha/beta*, ii *alpha/beta/alpha*, and iii) *beta-lactamase* (Kovacic *et al.*, 2018).

The classic catalytic triad of *serine* (S), *aspartic acid* (D)/ *glutamic acid* (E) and *histidine* (H) in lipases exist within the structure of all three superfamilies. However, each of these residues appear in a certain motif.

In the *alpha/beta* lipase superfamily, the active site *serine* is either present in a conserved pentapeptide motif of *GXSXG* (*Glycine*, Any amino acid, *Serine*, Any amino acid, *Glycine*) with the form of a nucleophilic elbow or as a *GDS* (*Glycine*, *Aspartic acid*, *Serine*) motif.

In the lipases of the superfamily *alpha/beta/alpha* (known as *SGNH* lipases), the active site *serine* exists in the *GDSL* (*Glycine*, *Aspartic acid*, *Serine*, *leucine*) motif. The name *SGNH* refers to four residues of *serine*, *glycine*, *asparagine*, and *histidine*. Each of these residues are conserved in four motifs or blocks.

Block I contain the active site *serine* and appears in a certain *GDS* (*Glycine*, *Aspartic acid*, *Serine*) motif. Block II has a *glycine* residue for donating a hydrogen to the oxyanion hole. The oxyanion hole is a small region in the active site of an enzyme which lowers the activation energy and promote the catalysis reaction. Block III is also involved in donating a hydrogen bond to the oxyanion hole with the typical *GXND* motif (*Glycine*, Any amino acid, *Asparagine*, *Aspartic acid*). The last block, block V, contain the catalytic *aspartic acid* and *histidine* as *DXXH* (*Aspartic acid*, 2 Any amino acid, *Histidine*) (Mølgaard *et al.*, 2000).

Little information is currently available about the lipases with the *beta-lactamase* fold. In this superfamily, the *GXSXG* motif is still present (in the C-terminal). However, the *serine* residue in this motif is no longer the active site. Instead, the active site *serine* usually appears in the N-terminal part of the protein and in a conserved motif of *SXXK* (*Serine*, 2 Any amino acids, *Lysine*). This motif is followed by a *tyrosine* that has a crucial role in the enzymatic activity (Kovacic *et al.*, 2018).

The most recent classification of bacterial lipolytic enzymes, including both triacylglyceride lipases (EC 3.1.1.3) and carboxylesterases (EC 3.1.1.1), is produced by Kovacic *et al.* (2018).

They have classified lipases into 19 families based on the similarity of amino acid sequences and physiological properties.

Lipase families that belong to *alpha/beta* superfamilies are archived in the *ESTHER* database (http://bioweb.supagro.inra.fr/ESTHER/Arpigny_Jaeger.table). Therefore, *Family II* and *Family VIII* which represent *SGNH* hydrolases and *beta-lactamases,* respectively are excluded from the *ESTHER* database.

The *ESTHER* database offers both protein blast (*BLASTp*) and alignment (*ClustalOmega*) for query protein sequences against sub-databases that only contain sequences of lipolytic families as classified by the Kovacic *et al.* (2018). However, only a few lipase sequences are present in each sub-database. Furthermore, the *ESTHER* database is the only lipolytic database that gets updated regularly. Other lipase databases like MELDB (Kang *et al.*, 2006) and LIPABASE (Messaoudi *et al.*, 2011), no longer exist.

UniProtKB also uses the *ESTHER* database for annotation and classification of proteins and contains 4669 sequences of bacterial lipases (EC. 3.1.1.3). Out of these, only 43 are manually curated and 38 have family classification based on the *ESTHER* database too (retrieved on 21/06/2021).

5.2. Materials and Methods

All 78 putative lipolytic sequences obtained from the anaerobic metagenome in Chapter 3, were uploaded/ pasted either as one *FASTA* file or individually in the search box of tested protein databases presented in Table 5-2.

All 'jobs' were submitted by selecting default options. For the *ESTHER* database different classified lipolytic families, *Family I* [*I.1- I.3*, *I.5*, *I.6*, *I.8*], *Family I.4* (*Lipase_2*), *Family XI* (*Lipase_3*), *Family X* and *Family XII*, were selected as a reference sub-database individually. For protein blast in the *ESTHER* database, apart from uploading one *FASTA* file with all sequences, an individual sequence, "*Lipase 3*" from *Bin 403*, was blasted as a test.

Table 5-2. Tested protein classification databases and their search engine.

Database	Search engine	Reference
InterPro	*InterProScan*	(Jones *et al.*, 2014; Blum *et al.*, 2021)
PROSITE	*ScanProsite*	(de Castro *et al.*, 2006; Sigrist *et al.*, 2013)
SUPERFAMILY	-	(Gough *et al.*, 2001; Wilson *et al.*, 2009)
PANTHER	*grafting*	(Thomas *et al.*, 2003)
PRINTS		(Attwood *et al.*, 1994)
CDD	*SPARCLE*	(Lu *et al.*, 2020)
Pfam	*HMMER*	(Mistry *et al.*, 2020)
ESTHER	*BLASTp*	(Lenfant *et al.*, 2013)

All lipase sequences were copied into a document file and checked against the four lipolytic patterns of *PROSITE* (Appendix W) manually. Potential motifs were recorded in an Excel file for further comparison.

One lipase (*Family I.3*) belonging to *Psychrobacter sp. PR-Wf-1* with an accession number of *A5WGV1* (Kovacic *et al.*, 2018) was selected as a test to evaluate the performance of *ScanProsite* on sequences already recorded as lipases. The accession number was searched in *UniProtKB* and the sequence was downloaded as a *FASTA* file for scanning by *ScanProsite*.

For comparison of some lipases, protein blast by *BLASTp* (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) and alignment with the hits by *ClustalOmega* (https://www.ebi.ac.uk/Tools/msa/clustalo/) with the default settings was performed for i) putative lipases of *Bin 744* (446 aa) and *Bin 1111* (450 aa), in which the alignment was done with the first two hits; ii) "*Lipase 2*" (362 aa) and "*Triacylglyceride lipase*" (562 aa) in *Bin 1020* (alignment with the first hit); and iii) lipases in *Bin 583*, *Bin 820* and *Bin 1001* (313 aa) (alignment with the first hit). It is worth mentioning that bins and MAGs are interchangeable words.

5.3. Results and discussion

The analysis revealed major discrepancies between different tools for predicting a protein family for a sequence. This is a major barrier to the reliable labelling of protein sequences.

5.3.1. InterPro and member databases

InterProScan failed to predict any family membership for more than half of the putative lipolytic sequences (41 of the total 78). The scanning tools of the member databases like *Pfam*, *PANTHER*, *SUPERFAMILY* and *CDD* classified most of the lipolytic sequences (Appendix

X). However, different tools made inconsistent predictions and some predictions were non-relevant to lipolytic families. For instance, *InterProScan* placed four sequences in "*Palmitoyl-protein thioesterase*" family while *Prokka* had already annotated them as either "*Lipase*" or "*Lactonizing lipase*". "*Palmitoyl-protein thioesterase*" family has an EC number (3.1.2.22) that is different from that of lipases (3.1.1.3). More importantly "*Palmitoyl-protein thioesterase*" family is not among the recognized list of the *InterPro* lipase family https://www.ebi.ac.uk/interpro/entry/InterPro/?page_size=100&search=lipase&type=family#table, retrieved on 08/08/2021.

InterProScan final prediction for these four sequences was based on the *Pfam* database. Both *SUPERFAMILY* and *CDD* found an *alpha/beta hydrolase fold*, and *PRINTS* found no hits for any of the four sequences. *ScanProsite* though identified two of these lipase sequences (*Bin 684* & *Bin 967*) as *LIPASE_SER (PS00120)* and *PANTHER* labelled both as "*SLL1969 Protein*" which also represents the palmitoyl hydrolase activity. By contrast, the other two lipase sequences (*Bin 631* & *Bin 1111*) had distinct placements by *PANTHER* and no hits in the *PROSITE* (Table 5-3).

Palmitoyl-protein thioesterases remove thioester-linked long chain fatty acids (e.g., palmitate) from the cysteine residues in proteins (Won *et al.*, 2018), which is a distinct activity from the lipolysis. Nonetheless, some members of this family might show esterase/lipase activity (Wang *et al.*, 2013). Yet, none of the predictions that each tool made determines which is the true activity. In other words, it is not clear whether the sequence is a bifunctional lipase/palmitoyl thioesterase or is a lipase with a similar domain to palmitoyl thioesterases. This ambiguity is also observable among lipases tagged as "*SLL1969 Protein*" representing the palmitoyl thioesterase activity by *PANTHER* but placed in a different family by *InterProScan* and *Pfam*. For instance, the "*Lipase*" in *Bin 744* (348 aa) is in the *GPI inositol-deacylase PGAP1-like* family and was identified as a lipase by the *ScanProsite;* the "*EstA*" in *Bin 617* (289 aa) was labelled as "*Lipase class 2*" by *Pfam* and as "*Lipase EstA/EstB*" by *InterProScan* (Table 5-3).

It is not clear how *InterPro* assigned a certain family/feature to a sequence from multiple predictions carried out by different tools. This ambiguity was mostly related to the Expect-values (or *E-values*) that describes the number of random hits for a database with a certain size. An E-value of 1 for a hit means that 1 match with the similar score can be found by chance within the particular database size. This means that lower *E-values* are more desirable.

For example, "*Lipase 3*" in *Bin 737* and "*Lipase 1*" in *Bin 484* (246 aa) had the same predictions by different tools except in *InterPro* where the former was labelled as "*Epoxide hydrolase-like*" and the latter as "*None predicted*" (Appendix X). The epoxide feature was assigned to "*Lipase 3*" in *Bin 737* by *PRINTS* with the *E-value* of 1.79×10^{-10}. However, the "*Abhydrolase*" fingerprint with *E-value* of 6.6×10^{-10} with the same tool was not assigned to "*Lipase 1*" by *InterPro*. Similarly, although all sequences with "*Abhydrolase*" prediction from *PRINTS* database had an *E-values* of the order between 10^{-8} to 10^{-14} (compared to the 10^{-5} to 10^{-10} *E-values* of the four "*Epoxide hydrolase-like*" sequences), none were picked by *InterPro*. We know that *E-values* depend on the size of each database, and *E-value* of the same order within different databases might not serve similar. But how *InterPro* filter *E-values* and select one for function assignment is not clear.

Furthermore, the feature/family prediction for a certain lipase sequence differed when that sequence was searched with *InterProScan* and when it was searched in the source database that *InterProScan* made its prediction based on that. For instance, for "*Lipase 2*" in *Bin 1111* (273 aa), *InterProScan* predicted that this lipase sequence belongs to the "*Streptomyces scabies esterase-like*" family (source database was *PANTHER*). However, used independently, *PANTHER* search box labelled the sequence as "*Lipase 2*", the same prediction that *Prokka* had already made. By contrast, *Pfam*, *SUPERFAMILY* and *CDD* placed it as either "*GDSL-Like lipase/acylhydrolase*" or "*SGNH hydrolase*", respectively. Based on *ESTHER* database, "*Lipase 2*" in *Bin 1111* is an *Alpha/beta hydrolase* which conflicts the prediction of *SGNH* hydrolase fold. *ScanProsite* also did not find any lipolytic pattern in the sequence though a motif like the *PS00120* pattern was observed (Appendix Y). The observed motif was "***Y****VA*LG***S****SM****A****A*" in which 4 amino acids had been substituted (**in bold**).

Table 5-3. Family membership prediction by all tools for lipase sequences in selected MAGs.

MAG ID	Prokka classification	Length (aa)	Prosite	IntherPro	Pfam	Panther	Superfamily	CDD	PRINTS
684	Lipase	247	PS00120	Palmitoyl protein thioesterase	Palmitoyl protein thioesterase	SLL1969 Protein	Abhydrolase	Abhydrolase/ EstA	No hit
967	Lipase	306	PS00120	Palmitoyl protein thioesterase	Palmitoyl protein thioesterase	SLL1969 Protein	Abhydrolase	Abhydrolase/ EstA	No hit
631	Lactonizing	297	None	Palmitoyl protein thioesterase	Palmitoyl protein thioesterase	Lecithin-Cholesterol Acyltransferase-Related	Abhydrolase	Abhydrolase/ EstA	No hit
1111	Lipase	339	None	Palmitoyl protein thioesterase	Palmitoyl protein thioesterase	Fasting Induced Lipase	Abhydrolase	Abhydrolase/ EstA	No hit
744	Lipase	348	PS00120	GPI inositol-deacylase PGAP1-like	PGAP1-like Protein	SLL1969 Protein	Abhydrolase	Abhydrolase/ EstA	No hit
617	EstA	289	None	Lipase EstA/EstB	Lipase class 2	SLL1969 Protein	Abhydrolase	Abhydrolase/ EstA	No hit

5.3.2. ScanProsite does not recognize test lipase

Searching the lipolytic sequences with the *ScanProsite* tool and comparing the result with the *InterProScan* revealed two important points. First, except for one sequence (the lipase in *Bin 265*), *InterProScan* did not report other lipolytic patterns of *PROSITE*. *ScanProsite* tool found 8 sequences with lipolytic patterns. Among the sequences *ScanProsite* identified as lipases (Table 5-4), all but two possessed the pentapeptide *GXSXG* motif (*PS00120*). One of those two was "*Lipase 2*" in *Bin 583* that had the *HGG* (Histidine, Glycine, Glycine) pattern (*PS01173*) with histidine as the active site. The other was "*Lipase 1*" in *Bin 265* which had the *GDSL* motif (*PS01098*) with the serine as the active site. The latter lipase was the only *SGNH* hydrolase, and the rest were all *alpha/beta* hydrolases. This was confirmed by prediction of other databases too, particularly the *SUPERFAMILY* (Appendix Y).

The second important point was that *ScanProsite* tool did not find lipolytic patterns in most lipases (69 out of 78) probably due to substitutions of a few amino acid in the sequences. Yet, this does not mean that those sequences cannot be lipases. A test lipase sequence from *Psychrobacter sp. PR-Wf-1* (Accession number: *A5WGV1*), which is a member of *Family I.3* (Kovacic *et al.*, 2018) had a *GY**SA**GA* motif. *PROSITE* did not recognize this motif as a lipolytic pattern though *UniProt* (https://www.uniprot.org/uniprot/A5WGV1) has archived it as a "*triacylglyceride lipase*". In this motif the first A, representing *alanine*, sits beside the *serine* which is not allowed in the *PS00120* lipolytic pattern (Appendix W) and hence *ScanProsite* shows no hits for it. The test lipase clearly showed that *ScanProsite* can mistakenly exclude lipases because lipolytic patterns of *PROSITE* cannot capture divergent sequence groups. Rather than patterns, profiles should be employed as they can provide more in-depth analysis. However, profiles are not currently available for most proteins including lipases in *PROSITE*.

Table 5-4. Details of lipolytic patterns in sequences identified as lipases by ScanProsite.

MAG ID	Prokka name	Length (aa)	Prosite ID	Prosite name	Lipolytic pattern	Phyla	Genus
583	Lipase 2	274	PS01173	Lipase_GDXG_His	MILI**HGG**GFKEEDKSG	Krumholzibacteria	NID
684	Lipase	247	PS00120	Lipase_Ser	VVII**GHSKGG**	Unclassified	NID
1359	Est A	215	PS00120	Lipase_Ser	IHFV**GHSLGG**	Proteobacteria	NID
265	Lipase 1	325	PS01098	Lipase_GDSL_Ser	VVFF**GDS**LSDTG	Proteobacteria	NID
967	Lipase	306	PS00120	Lipase_Ser	LVLV**GHSMGG**	Proteobacteria	Rhodoferax
336	Lipase	319	PS00120	Lipase_Ser	VDLV**GHSQGG**	Actinobacteriota	NID
744	Lipase	348	PS00120	Lipase_Ser	VDLV**GHSMGG**	Actinobacteriota	Mycolicibacterium
768	Lipase	353	PS00120	Lipase_Ser	VDLV**GHSMGG**	Actinobacteriota	Mycolicibacterium
768	Lipase	352	PS00120	Lipase_Ser	VDLV**GHSNGG**	Actinobacteriota	Mycolicibacterium

5.3.3. About putative lipases and classification tools

All lipases annotated as "*Putative*" by *PROKKA* were labelled as "*Lipase*, *secreted*" and "*Secretory lipase*" by *InterProScan* and *Pfam*, respectively. *PANTHER* also predicted either "*Lipase 5*" or "*Family Not Named*" for all the 16 sequences while *SUPERFAMILY* and *CDD* identified them as "*alpha/beta hydrolase fold*" and "*Secretory lipase*". There were no hits within the *PRINTS* and *PROSITE* (Appendix Y) for "*Putative*" lipases either. The two most dominant conserved residues among "*Putative*" lipases were "*GYSQGG*" (*Glycine, Tyrosine, Serine, Glutamine, Glycine, Glycine*) and "*GHSQGG*" (*Glycine, Histidine, Serine, Glutamine, Glycine, Glycine*), of which the latter is conserved within lipases of *Family I.2* (Appendix Z).

Another important point is that two of the "*Putative*" lipases, *Bin 744* (446 aa) and *Bin 1111* (450 aa), did not have the common lipase box (Appendix W) in their sequence. However, after the protein blast and alignment with the first two hits that had the best *E-values* (Appendix AA), two motifs of *GWLTGG* (*Glycine, Tryptophan, Leucine, Threonine, Glycine, Glycine*) and *GIAGGG* (*Glycine, Isoleucine, Alanine, Glycine, Glycine, Glycine*) were observed in them. Unlike the actual *GYSGGG* (*Glycine, Tyrosine, Serine, Glycine, Glycine, Glycine*) motif of the first two hits, they both lacked the active site serine. Also, the first two hits for both lipases were of the same taxa but with different *E-values*. These new motifs might represent potential lipolytic residues though we can only confirm it after gene cloning and further activity tests.

5.3.4. Tools in ESTHER database

When a *FASTA* file containing all 40 lipase sequences (from the putative lipolytic MAGs), was blasted in the *ESTHER* database only one hit within the *Family I.1* was found. Not only was the *E-value* poor (8.2), but also it was not clear which sequence of the file had been matched (Figure 5-1). By contrast, numerous hits were found for the query sequence (*Lipase 3* from *Bin 403*) within different databases of lipolytic families (Table 5-5). Yet, these hits were not helpful in deciding which databases of lipolytic families corresponds to the query sequence. The number of sequences in a database represents its size and can impact the *E-values*. For instance, the database of *lipase class 2* had 172 sequences with *E-value* of 9×10^{-4} for the hit. By contrast, *Family I.2* with only 18 sequences, showed a comparable *E-value* (5×10^{-4}) which made it impossible to understand which family best represents the query sequence.

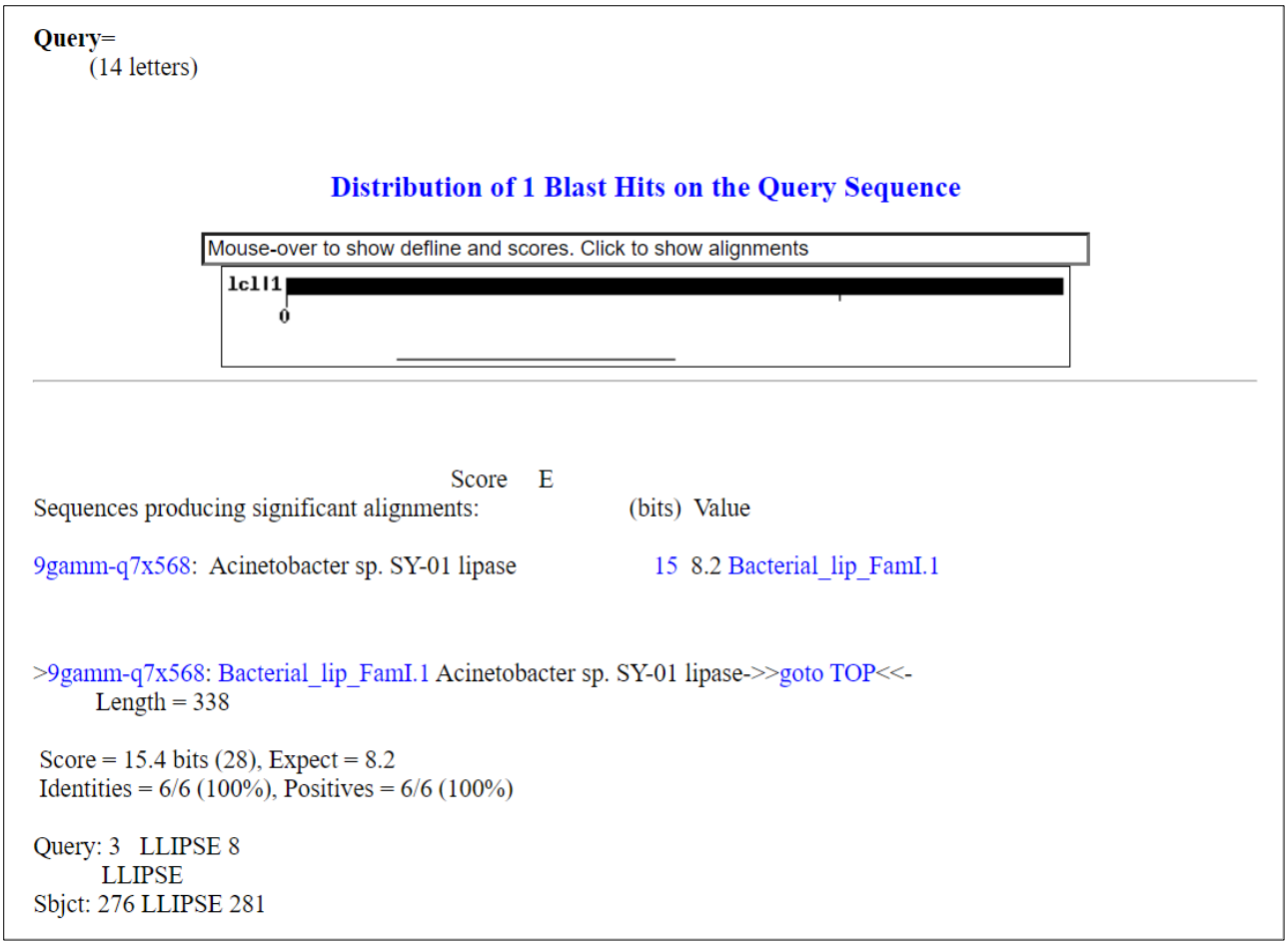


Figure 5-1. BLASTp results for the FASTA file of all lipase sequences.

Table 5-5. Details of BLASTp hits for Lipase 3 of MAG 403 within classified lipolytic families of ESTHER database.

Database	Number of sequences in ESTHER database	Number of BLASTp hits	Minimum E-value	Maximum E-value
Family I.1	47	17	0.022	6.0
Family I.2	18	15	5×10^{-4}	5.9
Family I.3	29	2	0.62	4.0
Family I.5	17	1	9.2	9.2
Family I.6	13	5	2×10^{-8}	9.0
Family I.8	111	22	7×10^{-5}	4.9
Family I.4 or Lipase class 2	172	49	9×10^{-4}	9.8
Family XI or Lipase class 3	428	18	0.002	8.6
Family X	80	14	0.008	8.5
Family XII	7	7	2.1	7.9

5.3.5. How to deal with multi motif cases?

"*Lipase 2*" (362 aa) and "*Triacylglyceride lipase*" (562) in *Bin 1020* had two motifs similar to the *PROSITE* lipolytic patterns specified as *PS01174* and *PS01173* in Appendix W. However, none of the tools classified them in any lipolytic groups unless that they had *alpha/beta hydrolase fold* (Appendix Y). Results from *BLASTp* and alignment with the first hit showed that the hits also possessed both motifs (Figure 5-2). Fortunately, the hits were manually curated entries in *UniProt* and hence, their active site was already identified as serine (position at 216 & 309, respectively in each hits) in the *GDS* motif along with aspartic acid (316 & 383) and histidine (346 & 413). Therefore, the histidine in the *HGG* motif did not have any catalytic role. Similarly, we can infer that in lipases of the *Bin 1020*, *GDS* motif represent the right lipolytic motif. In other words, the *HGG* motif (*PS01173*) should only be picked as a lipolytic pattern when the other patterns like *GXSXG* (*PS00120*) and *GDS* (*PS01174*) are absent.

```
sp|P95125|LIPN_MYCTU                LLVFYHGGGWTLGDLDTHDALCRLTCRDADIQVLSIDYRLAPEHPAPAAVEDAYAAFVWA   195
Lipase2_Bin1020_362                 ALVFFHGGGYVLGDLDSYDAVCRLLCRDAGVHVFAVDYRLAPEHPAPAALDDCLAAFRWV   187
sp|I6Y2J4|LIPY_MYCTU                YVVAIHGGAFILPPSIFHWLNYSVTAYQTGATVQVPIYPLVQEGGTAGTVVPAMAGLI--   291
Triacylglycerol_lipase_Bin1020_562  RVIALHGGGFITETSMFTFLTYSSLATNTGATVVVPVYPVVSKGGTARTVVPVTANLI--   390
                                     ::  ***.:               . ::.  *    * :. :  :  ::    * :

sp|P95125|LIPN_MYCTU                HEHASDEFGALPGRVAVGGDSAGGNLSAVVCQLARDKARYEGGPTP----VLQWLLYPRT   251
Lipase2_Bin1020_362                 ADHA-AEFGVDAGRIGVGGDSAGGGLAAAVAQCTRADT-----VAP----AGQLLVYPWT   237
sp|I6Y2J4|LIPY_MYCTU                ---STQIAQHGVSNVSVVGDSAGGNLALAAAQYMVSQ----GNP----VPSSM-------   333
Triacylglycerol_lipase_Bin1020_562  ---RSEVLTYGADNVSVLGDSAGGNIGLAALELLATRIRN-GDIAPESMPGRLVLLSSGL   446
                                              ..:.* ******.:. .. :
```

Figure 5-2. Alignment of two lipases from Bin 1020 (Lipase 2, 362 aa and Triacylglycerol lipase, 562 aa) with their first BLASTp hit using ClustalOmega.

Analysis of lipase sequences from *Bin 583*, *Bin 820,* and *Bin 1001* (313 aa) that possessed potential *PS01173* and *PS00120* patterns, returned similar results. Of the three MAGs, *ScanProsite* only reported on *Bin 583* for having the *PS01173* pattern though it had also a *PS00120* pattern (*FGAR**G**S**S**A**G**G*) in its sequence. The *BLASTp* hit for the lipase of *Bin 583*, on the other hand, had *ITIT**G**G**S**A**G**A* with its serine assigned to catalytic role in the *UniProt* (manually curated entry). However, instead of *HGG*, it had a *PGG* motif. By contrast, lipases of *Bin 820* and *Bin 1001*, additional to *HGG*, had *VAVA**G**H**S**A**G**A* and *IGVW**G**V**S**A**G**G* motifs, respectively. They even had the same hit as the *Bin 583* and likewise, the *PS00120* pattern seems more likely represent a conserved lipolytic motif than the *PS01173*.

Overall, these analyses showed once more that the present patterns in *ScanProsite* are not comprehensive, and they lead us to false results.

5.3.6. Common grounds between lipases

In general, there was no consensus lipolytic motif between the lipases annotated with similar names and different amino acids appeared as *X* in the common *GXSXG* motifs. Nevertheless, in case of lipases annotated as "*Triacylglycerol lipases*", *GDSAGG* was present in all five MAGs. While all belonged to the phylum *Actinobacteriota*, they were from two distinct genera of *Mycolicibacterium* and *Austwickia*. Also, none of the classification tools assigned them to any specific family other than *alpha/beta hydrolases*.

5.4. Conclusion

In recent years, protein databases have evolved significantly both in number and content. Various protein scanning and classifying tools are developed too. However, assigning functions to most proteins is still a challenging task.

Particularly for lipases, both conventional and newer databases like *PROSITE* and *Pfam*, respectively, lack adequate and accurate lipolytic patterns and profiles. Although *PROSITE* has started to use profiles rather than patterns to involve a less permissive and more selective approach for classifying some proteins, it is still using patterns for lipases. The limitation of using patterns was clearly reflected in our results where *ScanProsite* failed to identify a classified lipase sequence which was manually curated in *Uniprot*. In addition to this, when multi lipolytic patterns were present in the sequences, *ScanProsite* failed to find the true pattern. In these cases, blasting the protein sequence and aligning it with the high-scored hits can be helpful. For some lipases particularly, the presence of *GXSXG/GDS* pattern was superior to the *HGG*. This technique worked too for those lipases that were annotated by *Prokka* as "*putative lipase*" but did not have any lipase box.

Pfam could not adequately classify most of lipolytic sequences that had *alpha/beta* hydrolase fold. Presumably because there are about 64,110 *GDSL* lipase sequences in *pfam* whereas only half of this value (33,381) are the lipase sequences with *alpha/beta* hydrolase fold.

The only dedicated database for lipases that updates regularly is *ESTHER*, but it only contains those lipases that possess the *alpha/beta* hydrolase fold and not the *SGNH* hydrolases and *beta lactamases.* The number of lipolytic sequences included in various lipase families in *ESTHER* is remarkably low such that a protein blast of query protein does not return a reliable result.

Although we expected that *InterPro* would represents the member databases best and return the most reliable classification for lipolytic sequences, this did not happen. Typically, *InterPro* did not have consensus family prediction results with its member databases for a particular lipase sequence and for some lipase sequences whilst the search box of member databases assigned the sequence to a family, *InterProScan* assigned no family for the same sequence. Of the current 37,000 entries that exist in *InterPro* from all member databases, only 17 family entries belong to lipases. However, not all of these lipolytic entries are related to bacteria. Only 6 families, 1 domain and 3 active sites among those entries are for bacterial lipases.

Overall, none of the current tools can be used for sensible lipase classification as they do not show consistent results even for a certain known lipase sequence. Better automated tools along with synthetic molecular biology approaches that can check true activity of lipases are required for extracting meaningful and consistent lipolytic motifs.

Chapter 6 : Concluding remarks

This study aimed to screen for potential and actual extracellular lipases produced by a cold-adapted microbial community using molecular biology techniques such as metagenomics and metaproteomics.

Eight lab-scale An-MBRs were developed to treat domestic wastewater at 4 and 15 °C. The inoculum used in the reactors was collected from the *Arctic* and the feed was primary influent from a full-scale activated sludge plant. For some reactors the feed was treated with UV.

Both DNA and protein were extracted from the biofilm and bulk liquid. Purified extracts were sent for sequencing and mass-spectrometry and further data analysis was performed using various bioinformatics tools detailed in Chapter 3 and Chapter 4. Identified lipolytic sequences recovered in putative lipolytic MAGs were evaluated with several protein classification tools in Chapter 5.

A summary of some of the most important findings is:

1. Within the metagenomics data lipases had significantly lower number of genes compared to other hydrolytic extracellular enzymes
2. Of the 32 common abundant genera in all reactors (relative abundance ≥1%) only three (*Chlorobium*, *Desulfobacter*, and *Mycolicibacterium*) were recovered as putative lipolytic MAGs.
3. Most lipases were from the phyla *Actinobacteria* and genera *Mycolicibacterium* and *Corynebacterium* that accumulate PHAs.
4. Lipolytic activity may not always be directed at degrading exogenous lipidic molecules and may be linked to PHA accumulation/degradation, denitrification, and invasion of other bacteria's outer membrane.
5. With few exceptions, there was no significant correlation between the reactor conditions and the number of reads mapped to the putative lipolytic MAGs.
6. Temperature had no significant role on lipase length.
7. Metaproteomics did not provide sufficient proteome coverage for less abundant proteins such as extracellular enzymes including lipases.
8. Out of the 32 common genera profiled by metagenomics, 15 were identified by metaproteomics too; at least 6 of them were involved in lipid/PHA accumulation.

9. Metaproteomics identified *fadL* genes for four genera (*Dechloromonas*, *Azoarcus*, *Aeromonas* and *Sulfurimonas*), but did not identify any associated lipases. None of these four genera were recovered as putative lipolytic MAGs by metagenomics.
10. The proteins identified by metaproteomics were mainly porins and outer membrane proteins and some cytoplasmic proteins were identified too that might enter the EPS through extracellular vesicles.
11. A newer generation of protein databases like *pfam* that use profiles rather than conventional patterns in databases like *PROSITE*, are generally better for protein classifications. However, for lipases both of these databases lack adequate and accurate patterns and profiles.
12. *ScanProsite* failed to identify a classified lipase sequence which was manually curated in *UniProtKB.*
13. Protein blast and further alignment of the blasted sequence with the high-scored hits is more useful for identifying the lipolytic motifs in a sequence than relying on protein classification tools.
14. *The ESTHER* database is a good archive of lipases of the *alpha/beta hydrolase* family, but it should not be used as a reference database for protein blast and classification since it contains only a low number of lipase sequences.
15. *InterProScan* is still not a reliable tool for identification or classification of lipolytic sequences.
16. No consisted results obtained for a certain lipase sequence with different protein classification tools.

Despite the interesting results, this study had some limitations as well. One of the limitations of the current study is related to initial sampling and further DNA and protein extractions. No metagenomics and metaproteomics was done on the reactor feed or inocula for comparison with the samples from the liquid and biofilm phase of the reactors. Therefore, it was not possible to identify which of the lipolytic MAGs came from the wastewater treatment plant, or inocula (soils and sediments from the Arctic) and which grew in the reactors. Besides, metagenomics does not show that the extracted DNA necessarily belonged to the active bacteria. The DNA of dead microbes can also be extracted along with the viable microbes.

The other limitation was that molecular biology techniques and tools are still potentially biased by the extraction steps to the sequencing, the mass spectrometry and final bioinformatics data analysis.

There is lack of data about bacterial extracellular lipase sequences in public databases too. In *UniProtKB* for example, of the current 4669 lipase sequences (EC: 3.1.1.3), only 43 are manually curated and only two have recorded mass spectra data. Also, only 6 of those sequences are cross-referenced to *PRIDE* database, an archive for proteomics data. This lack of information about bacterial lipases and their mass spectra limits the true identification of bacterial lipase sequences, their annotation (assigning function to them) and classification.

Chapter 7 : Future works

The fate of lipids in different wastewater treatment systems is still not well understood. Particularly, in low-temperature anaerobic treatment of domestic wastewater by psychrophiles a sound grasp of the barriers to lipolysis could have a significant impact on the development of such systems at full-scale.

The first step towards the rational engineering of lipolysis in any system is to identify the lipolytic bacteria. Molecular biology tools like metagenomics and metaproteomics either individually or combined can, in principle, provide such information. However, despite all the advances in this area in the past decade these tools can only illuminate one facet of the puzzle. Many factors such as the presence of different metabolites, inhibitors and enzymes can affect the gene regulation but are overlooked by metagenomics and metaproteomics. These tools cannot be used in isolation.

Some of the suggested future works for understanding the fate of the lipids and identifying the lipolysis potential among the microbes are:

1. Performing both metagenomics and metaproteomics for the feed and the inocula to identify which bacteria are grown in the reactor, which are related to inocula and which are from the wastewater treatment plant.
2. Optimizing and developing unbiased extraction methods for both DNA and proteins. Particularly extracellular enzymes including lipases are more sensitive to the extraction substances and protocols.
3. Integrating short and long reads to improve the quality of assembly in metagenomics and hence reducing the occurrence of mis-assembly and mis-annotation.
4. Developing tools for assigning function to protein sequences, classifying proteins accurately, particularly for bacterial lipases the database coverage is poor.
5. Employing high-resolution mass spectrometers to quantify and identify the composition of microbial community.
6. Developing better tools for identifying the mass spectra and matching them to protein groups.
7. Enriching public databases with the mass spectra of classified bacterial lipases.
8. Reducing the dependency of metaproteomics to metagenomics databases and using de novo metaproteomics instead.

9. Integrating the analysis of extracellular vesicles, separating them from the EPS during the extraction procedure, characterising their protein content and demystifying their role in the EPS and relative to members of the bacterial community.
10. Employing synthetic biology techniques to determine the gene regulation mechanisms and extracellular excretion pathways for identified lipase sequences and enriching the public databases.
11. Developing biosensors that can detect lipases in real time (e.g., by detecting free long-chain fatty acids).

References

Abeliovich, A. and Vonshak, A. (1992) 'Anaerobic metabolism of Nitrosomonas europaea', *Archives of Microbiology*, 158(4), pp. 267-270.

Ahn, J.H., Pan, J.G. and Rhee, J.S. (1999) 'Identification of the tliDEF ABC transporter specific for lipase in *Pseudomonas fluorescens SIK W1*', *Journal of bacteriology*, 181(6), pp. 1847-1852.

Ali, Y.B., Verger, R. and Abousalham, A. (2012) 'Lipases or Esterases: Does It Really Matter? Toward a New Bio-Physico-Chemical Classification', in Sandoval, G. (ed.) *Lipases and Phospholipases: Methods and Protocols*. Totowa, NJ: Humana Press, pp. 31-51.

Allison, S.D. (2005) 'Cheaters, diffusion and nutrients constrain decomposition by microbial enzymes in spatially structured environments', *Ecology Letters*, 8(6), pp. 626-635.

Allison, S.D., Lu, L., Kent, A.G. and Martiny, A.C. (2014a) 'Extracellular enzyme production and cheating in Pseudomonas fluorescens depend on diffusion rates', *Frontiers in Microbiology*, 5(169).

Allison, S.D., Lu, L., Kent, A.G. and Martiny, A.C. (2014b) 'Extracellular enzyme production and cheating in Pseudomonas fluorescens depend on diffusion rates', *Frontiers in Microbiology*, 5, p. 169.

Allison, S.D. and Vitousek, P.M. (2005) 'Responses of extracellular enzymes to simple and complex nutrient inputs', *Soil Biology and Biochemistry*, 37(5), pp. 937-944.

Alves, J.I., Salvador, A.F., Castro, A.R., Zheng, Y., Nijsse, B., Atashgahi, S., Sousa, D.Z., Stams, A.J.M., Alves, M.M. and Cavaleiro, A.J. (2020) 'Long-Chain fatty acids degradation by *Desulfomonile* species and proposal of "*Candidatus Desulfomonile Palmitatoxidans*"', *Frontiers in Microbiology*, 11(3227).

Andersson, R.E. (1980) 'Microbial lipolysis at low temperatures', *Applied and environmental microbiology*, 39(1), pp. 36-40.

APHA (2006) *Standard methods for the examination of water and wastewater*. 19 edn. Wasington DC: American Public Health Association.

Aquino, S.F., Araújo, J.C., Passos, F., Curtis, T.P. and Foresti, E. (2019) 'Fundamentals of anaerobic sewage treatment', in de Lemos Chernicharo, C.A. and Bressani-Ribeiro, T. (eds.) *In anaerobic reactors for sewage treatment: design, construction and operation*. IWA Publishing, p. 0.

Arkin, A.P., Cottingham, R.W., Henry, C.S., Harris, N.L., Stevens, R.L. and Maslov, S. (2018) 'KBase: The United States department of energy systems biology knowledgebase', *Nature Biotechnology*, 36(7), pp. 566-569.

Arnosti, C. (2011) 'Microbial extracellular enzymes and the marine carbon cycle', *Annual Review of Marine Science*, 3(1), pp. 401-425.

Arpigny, J.L. and Jaeger, K.E. (1999) 'Bacterial lipolytic enzymes: classification and properties', *Biochemical Journal*, 343(Pt 1), pp. 177-183.

Attwood, T.K., Beck, M.E., Bleasby, A.J. and Parry-Smith, D.J. (1994) 'PRINTS--a database of protein motif fingerprints', *Nucleic acids research*, 22(17), pp. 3590-3596.

Attwood, T.K., Bradley, P., Flower, D.R., Gaulton, A., Maudling, N., Mitchell, A.L., Moulton, G., Nordle, A., Paine, K., Taylor, P., Uddin, A. and Zygouri, C. (2003) '*PRINTS* and its automatic supplement, prePRINTS', *Nucleic acids research*, 31(1), pp. 400-402.

Attwood, T.K. and Mitchell, A.L. (2019) 'The evolution of protein family databases', in Ranganathan, S., Gribskov, M., Nakai, K. and Schönbach, C. (eds.) *In encyclopedia of bioinformatics and computational biology*. Oxford: Academic Press, pp. 34-45.

Baltar, F. (2018) 'Watch out for the "Living Dead": cell-free enzymes and their fate', *Frontiers in Microbiology*, 8(2438).

Baltar, F., Arístegui, J., Gasol, J.M., Sintes, E., van Aken, H.M. and Herndl, G.J. (2010) 'High dissolved extracellular enzymatic activity in the deep Central Atlantic Ocean', *Aquatic Microbial Ecology*, 58(3), pp. 287-302.

Becker, P. (2010) 'Chapter 42 - Understanding and Optimizing the Microbial Degradation of Olive Oil: A Case Study with the Thermophilic Bacterium Geobacillus thermoleovorans IHI-91', in Preedy, V.R. and Watson, R.R. (eds.) *In olives and olive oil in health and disease prevention*. San Diego: Academic Press, pp. 377-386.

Becker, P., Köster, D., Popov, M.N., Markossian, S., Antranikian, G. and Märkl, H. (1999) 'The biodegradation of olive oil and the treatment of lipid-rich wool scouring wastewater under aerobic thermophilic conditions', *Water Research*, 33(3), pp. 653-660.

Becker, P. and Märkl, H. (2000) 'Modeling of olive oil degradation and oleic acid inhibition during chemostat and batch cultivation of *Bacillus thermoleovorans IHI-91*', *Biotechnology and Bioengineering*, 70(6), pp. 630-637.

Bernard, K.A. and Funke, G. (2015) 'Corynebacterium', in *Bergey's Manual of Systematics of Archaea and Bacteria*. pp. 1-70.

Biller, S.J. (2020) 'Extracellular Vesicles in the Environment', in Kaparakis-Liaskos, M. and Kufer, T.A. (eds.) *Bacterial Membrane Vesicles: Biogenesis, Functions and Applications*. Cham: Springer International Publishing, pp. 75-99.

Black, P.N. (1988) 'The fadL gene product of *Escherichia coli* is an outer membrane protein required for uptake of long-chain fatty acids and involved in sensitivity to *bacteriophage T2*', *Journal of bacteriology*, 170(6), pp. 2850-2854.

Black, P.N. and DiRusso, C.C. (2003) 'Transmembrane movement of exogenous long-chain fatty acids: proteins, enzymes, and vectorial esterification', *Microbiology and Molecular Biology Reviews*, 67(3), pp. 454-472.

Bligh, E.G. and Dyer, W.J. (1959) 'A rapid method of total lipid extraction and purification', *Can J Biochem Physiol*, 37(8), pp. 911-7.

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G.A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D.H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Bateman, A. and Finn, R.D. (2021) 'The InterPro protein families and domains database: 20 years on', *Nucleic acids research*, 49(D1), pp. D344-D354.

Boekema, B.K.H.L., Beselin, A., Breuer, M., Hauer, B., Koster, M., Rosenau, F., Jaeger, K.-E. and Tommassen, J. (2007) 'Hexadecane and Tween 80 stimulate lipase production in *Burkholderia glumae* by different mechanisms', *Applied and environmental microbiology*, 73(12), pp. 3838-3844.

Braker, G., Zhou, J., Wu, L., Devol, A.H. and Tiedje, J.M. (2000) 'Nitrite reductase genes (nirK and nirS) as functional markers to investigate diversity of denitrifying bacteria in pacific northwest marine sediment communities', *Applied and environmental microbiology*, 66(5), pp. 2096-2104.

Brandt, M.J., Johnson, K.M., Elphinston, A.J. and Ratnayaka, D.D. (2017) 'Chapter 10 - Specialized and Advanced Water Treatment Processes', in Brandt, M.J., Johnson, K.M., Elphinston, A.J. and Ratnayaka, D.D. (eds.) *Twort's Water Supply (Seventh Edition)*. Boston: Butterworth-Heinemann, pp. 407-473.

Bressani-Ribeiro, T., Chamhum-Silva, L.A. and Chernicharo, C.A.L. (2019) 'Constraints, performance and perspectives of anaerobic sewage treatment: lessons from full-scale sewage treatment plants in Brazil', *Water Science and Technology*, 80(3), pp. 418-425.

Bryson, S., Li, Z., Pett-Ridge, J., Hettich, R.L., Mayali, X., Pan, C. and Mueller, R.S. (2016) 'Proteomic Stable Isotope Probing Reveals Taxonomically Distinct Patterns in Amino Acid Assimilation by Coastal Marine Bacterioplankton', *mSystems*, 1(2), pp. e00027-15.

Canterbury, J.D., Merrihew, G.E., MacCoss, M.J., Goodlett, D.R. and Shaffer, S.A. (2014) 'Comparison of data acquisition strategies on quadrupole ion trap instrumentation for shotgun proteomics', *Journal of the American Society for Mass Spectrometry*, 25(12), pp. 2048-2059.

Carlozzi, P. and Sacchi, A. (2001) 'Biomass production and studies on *Rhodopseudomonas palustris* grown in an outdoor, temperature controlled, underwater tubular photobioreactor', *J Biotechnol*, 88(3), pp. 239-49.

cerebis, M.D. (2017) *GTDB-Tk README.md: Classification summary file*. Available at: https://github.com/cerebis/GtdbTk#classification-summary-file (Accessed: 6/04/2021).

Cheremisinoff, P.N. (1995) 'Chapter 8 - Treatment of Effluent Fertilizer Industry Example', in Cheremisinoff, P.N. (ed.) *Waste minimization and cost reduction for the process industries*. Park Ridge, NJ: William Andrew Publishing, pp. 285-324.

Chipasa, K.B. and Mdrzycka, K. (2008) 'Characterization of the fate of lipids in activated sludge', *Journal of Environmental Sciences*, 20(5), pp. 536-542.

Chróst, R.J. (1991) 'Environmental Control of the Synthesis and Activity of Aquatic Microbial Ectoenzymes', in Chróst, R.J. (ed.) *Microbial Enzymes in Aquatic Environments*. New York, NY: Springer New York, pp. 29-59.

Chu, L. and Wang, J. (2017) 'Denitrification of groundwater using a biodegradable polymer as a carbon source: long-term performance and microbial diversity', *RSC Advances*, 7(84), pp. 53454-53462.

Chung, C.W., You, J., Kim, K., Moon, Y., Kim, H. and Ahn, J.H. (2009) 'Export of recombinant proteins in *Escherichia coli* using ABC transporter with an attached lipase ABC transporter recognition domain (LARD)', *Microbial cell factories*, 8, pp. 11-11.

Clark, D.P. and Cronan, J.E. (2005) 'Two-Carbon Compounds and Fatty Acids as Carbon Sources', *EcoSal Plus*, 1(2).

Claverie, J.-M. and Notredame, C. (2006) *Bioinformatics for dummies*. John Wiley & Sons.

Clesceri, L.S., Greenberg, A.E. and Eaton, A.D. (1996) 'Standard methods for the examination of water and wastewater', *APHA, AWWA and WPCF, Washington DC*.

Coats, E.R., Watson, B.S. and Brinkman, C.K. (2016) 'Polyhydroxyalkanoate synthesis by mixed microbial consortia cultured on fermented dairy manure: Effect of aeration on process rates/yields and the associated microbial ecology', *Water Research*, 106, pp. 26-40.

Cookney, J., McLeod, A., Mathioudakis, V., Ncube, P., Soares, A., Jefferson, B. and McAdam, E.J. (2016) 'Dissolved methane recovery from anaerobic effluents using hollow fibre membrane contactors', *Journal of Membrane Science*, 502, pp. 141-150.

Cyprowski, M., Stobnicka-Kupiec, A., Ławniczek-Wałczyk, A., Bakal-Kijek, A., Gołofit-Szymczak, M. and Górny, R.L. (2018) 'Anaerobic bacteria in wastewater treatment plant', *International Archives of Occupational and Environmental Health*, 91(5), pp. 571-579.

de Castro, E., Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A. and Hulo, N. (2006) 'ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins', *Nucleic acids research*, 34(Web Server issue), pp. W362-W365.

Demir, Ö. and Yapıcıoğlu, P. (2019) 'Investigation of GHG emission sources and reducing GHG emissions in a municipal wastewater treatment plant', *Greenhouse Gases: Science and Technology*, 9(5), pp. 948-964.

Deutsch, E.W. (2012) 'File formats commonly used in mass spectrometry proteomics', *Molecular & cellular proteomics : MCP*, 11(12), pp. 1612-1621.

Dueholm, T.E., Andreasen, K.H. and Nielsen, P.H. (2001) 'Transformation of lipids in activated sludge', *Water Science and Technology*, 43(1), pp. 165-172.

Duggin, I.G., Aylett, C.H.S., Walsh, J.C., Michie, K.A., Wang, Q., Turnbull, L., Dawson, E.M., Harry, E.J., Whitchurch, C.B., Amos, L.A. and Löwe, J. (2015) 'CetZ tubulin-like proteins control archaeal cell shape', *Nature*, 519(7543), pp. 362-365.

Eggert, T., Brockmeier, U., Dröge, M.J., Quax, W.J. and Jaeger, K.-E. (2003) 'Extracellular lipases from Bacillus subtilis: regulation of gene expression and enzyme activity by amino acid supply and external pH', *FEMS Microbiology Letters*, 225(2), pp. 319-324.

Elmitwalli, T. (2000) *Anaerobic treatment of domestic sewage at low temperature*. Wageningen Univ.

Flemming, H.-C., Neu, T.R. and Wozniak, D.J. (2007) 'The EPS Matrix: The "House of Biofilm Cells"', *Journal of Bacteriology*, 189(22), pp. 7945-7947.

Flemming, H.-C. and Wingender, J. (2001) 'Relevance of microbial extracellular polymeric substances (EPSs) - Part I: Structural and ecological aspects', *Water Science and Technology*, 43(6), pp. 1-8.

Florentino, A.P., Costa, R.B., Hu, Y., O'Flaherty, V. and Lens, P.N.L. (2020) 'Long-chain fatty acid degradation coupled to biological sulfidogenesis: a prospect for enhanced metal recovery', *Frontiers in Bioengineering and Biotechnology*, 8(1218).

Frias, A., Manresa, A., de Oliveira, E., López-Iglesias, C. and Mercade, E. (2010) 'Membrane Vesicles: A Common Feature in the Extracellular Matter of Cold-Adapted Antarctic Bacteria', *Microbial Ecology*, 59(3), pp. 476-486.

Frigaard, N.-U., Chew, A.G.M., Li, H., Maresca, J.A. and Bryant, D.A. (2003) '*Chlorobium Tepidum*: Insights into the structure, physiology, and metabolism of a green sulfur bacterium derived from the complete genome sequence', *Photosynthesis Research*, 78(2), pp. 93-117.

Fritts, R.K., McCully, A.L. and McKinlay, J.B. (2021) 'Extracellular Metabolism Sets the Table for Microbial Cross-Feeding', *Microbiology and Molecular Biology Reviews*, 85(1), pp. e00135-20.

Frølund, B., Palmgren, R., Keiding, K. and Nielsen, P.H. (1996) 'Extraction of extracellular polymers from activated sludge using a cation exchange resin', *Water Research*, 30(8), pp. 1749-1758.

Galka, F., Wai, S.N., Kusch, H., Engelmann, S., Hecker, M., Schmeck, B., Hippenstiel, S., Uhlin, B.E. and Steinert, M. (2008) 'Proteomic characterization of the whole secretome of *Legionella pneumophila* and functional analysis of outer membrane vesicles', *Infection and Immunity*, 76(5), pp. 1825-1836.

Georlette, D., Blaise, V., Collins, T., D'Amico, S., Gratia, E., Hoyoux, A., Marx, J.-C., Sonan, G., Feller, G. and Gerday, C. (2004) 'Some like it cold: biocatalysis at low temperatures', *FEMS Microbiology Reviews*, 28(1), pp. 25-42.

Gessesse, A., Dueholm, T., Petersen, S.B. and Nielsen, P.H. (2003) 'Lipase and protease extraction from activated sludge', *Water Research*, 37(15), pp. 3652-3657.

Görke, B. and Stülke, J. (2008) 'Carbon catabolite repression in bacteria: many ways to make the most out of nutrients', *Nature Reviews Microbiology*, 6(8), pp. 613-624.

Gough, J., Karplus, K., Hughey, R. and Chothia, C. (2001) 'Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure11Edited by G. Von Heijne', *Journal of Molecular Biology*, 313(4), pp. 903-919.

Griffiths, R.I., Whiteley, A.S., O'Donnell, A.G. and Bailey, M.J. (2000) 'Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and

rRNA-based microbial community composition', *Applied and Environmental Microbiology*, 66(12), pp. 5488-5491.

Haft, D.H., Loftus, B.J., Richardson, D.L., Yang, F., Eisen, J.A., Paulsen, I.T. and White, O. (2001) '*TIGRFAMs*: a protein family resource for the functional identification of proteins', *Nucleic acids research*, 29(1), pp. 41-43.

Hamann, E., Gruber-Vodicka, H., Kleiner, M., Tegetmeyer, H.E., Riedel, D., Littmann, S., Chen, J., Milucka, J., Viehweger, B., Becker, K.W., Dong, X., Stairs, C.W., Hinrichs, K.U., Brown, M.W., Roger, A.J. and Strous, M. (2016) 'Environmental Breviatea harbour mutualistic Arcobacter epibionts', *Nature*, 534(7606), pp. 254-8.

Han, X., Aslanian, A. and Yates, J.R., 3rd (2008) 'Mass spectrometry for proteomics', *Current opinion in chemical biology*, 12(5), pp. 483-490.

Hauschild, P., Röttig, A., Madkour, M.H., Al-Ansari, A.M., Almakishah, N.H. and Steinbüchel, A. (2017) 'Lipid accumulation in prokaryotic microorganisms from arid habitats', *Applied Microbiology and Biotechnology*, 101(6), pp. 2203-2216.

Hausmann, S. and Jaeger, K.-E. (2010) 'Lipolytic Enzymes from Bacteria', in Timmis, K.N. (ed.) *Handbook of hydrocarbon and lipid microbiology*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1099-1126.

Heidrich, E.S., Curtis, T.P. and Dolfing, J. (2011) 'Determination of the Internal Chemical Energy of Wastewater', *Environmental Science & Technology*, 45(2), pp. 827-832.

Hettich, R.L., Pan, C., Chourey, K. and Giannone, R.J. (2013) 'Metaproteomics: Harnessing the Power of High Performance Mass Spectrometry to Identify the Suite of Proteins That Control Metabolic Activities in Microbial Communities', *Analytical Chemistry*, 85(9), pp. 4203-4214.

Heyer, R., Schallert, K., Büdel, A., Zoun, R., Dorl, S., Behne, A., Kohrs, F., Püttker, S., Siewert, C., Muth, T., Saake, G., Reichl, U. and Benndorf, D. (2019) 'A Robust and Universal Metaproteomics Workflow for Research Studies and Routine Diagnostics Within 24 h Using Phenol Extraction, FASP Digest, and the MetaProteomeAnalyzer', *Frontiers in Microbiology*, 10(1883).

Heylen, K., Lebbe, L. and De Vos, P. (2008) '*Acidovorax caeni sp. nov.*, a denitrifying species with genetically diverse isolates from activated sludge', *International Journal of Systematic and Evolutionary Microbiology*, 58(1), pp. 73-77.

Hong, J., Dauros-Singorenko, P., Whitcombe, A., Payne, L., Blenkiron, C., Phillips, A. and Swift, S. (2019) 'Analysis of the Escherichia coli extracellular vesicle proteome identifies markers of purity and culture conditions', *Journal of Extracellular Vesicles*, 8(1), p. 1632099.

Hrudey, S.E. (1981) 'Activated sludge response to emulsified lipid loading', *Water Research*, 15(3), pp. 361-373.

Hugenholtz, P. and Tyson, G.W. (2008) 'Metagenomics', *Nature*, 455, p. 481.

Hughes, C., Ma, B. and Lajoie, G.A. (2010) 'De Novo Sequencing Methods in Proteomics', in Hubbard, S.J. and Jones, A.R. (eds.) *Proteome Bioinformatics*. Totowa, NJ: Humana Press, pp. 105-121.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuche, B.A., de Castro, E., Lachaize, C., Langendijk-Genevaux, P.S. and Sigrist, C.J.A. (2007) 'The 20 years of PROSITE', *Nucleic Acids Research*, 36(suppl_1), pp. D245-D249.

Ilshadsabah, A. and Suchithra, T.V. (2019) 'Bacterial Nanowires: An Invigorating Tale for Future', in Prasad, R. (ed.) *Microbial Nanobionics: Volume 2, Basic Research and Applications*. Cham: Springer International Publishing, pp. 77-88.

Jachlewski, S., Jachlewski, W.D., Linne, U., Bräsen, C., Wingender, J. and Siebers, B. (2015) 'Isolation of extracellular polymeric substances from biofilms of the thermoacidophilic archaeon *sulfolobus acidocaldarius*', *Frontiers in Bioengineering and Biotechnology*, 3(123).

Jaeger, K.-E., Ransac, S., Dijkstra, B.W., Colson, C., van Heuvel, M. and Misset, O. (1994) 'Bacterial lipases', *FEMS Microbiology Reviews*, 15(1), pp. 29-63.

Jaeger, K.E., Steinbüchel, A. and Jendrossek, D. (1995) 'Substrate specificities of bacterial polyhydroxyalkanoate depolymerases and lipases: bacterial lipases hydrolyze poly(omega-hydroxyalkanoates)', *Applied and environmental microbiology*, 61(8), pp. 3113-3118.

Jaimes-Estévez, J., Zafra, G., Martí-Herrero, J., Pelaz, G., Morán, A., Puentes, A., Gomez, C., Castro, L.d.P. and Escalante Hernández, H. (2021) 'Psychrophilic Full Scale Tubular Digester Operating over Eight Years: Complete Performance Evaluation and Microbiological Population', *Energies*, 14(1), p. 151.

Jehmlich, N., Vogt, C., Lünsmann, V., Richnow, H.H. and von Bergen, M. (2016) 'Protein-SIP in environmental studies', *Curr Opin Biotechnol*, 41, pp. 26-33.

Jeon, E.-Y., Song, J.-W., Cha, H.-J., Lee, S.-M., Lee, J. and Park, J.-B. (2018) 'Intracellular transformation rates of fatty acids are influenced by expression of the fatty acid transporter FadL in Escherichia coli cell membrane', *Journal of Biotechnology*, 281, pp. 161-167.

Jeong, K., Kim, S. and Bandeira, N. (2012) 'False discovery rates in spectral identification', *BMC bioinformatics*, 13 Suppl 16(Suppl 16), pp. S2-S2.

Jimenez-Diaz, L., Caballero, A. and Segura, A. (2017) 'Pathways for the Degradation of Fatty Acids in Bacteria', in Rojo, F. (ed.) *Aerobic Utilization of Hydrocarbons, Oils and Lipids*. Cham: Springer International Publishing, pp. 1-23.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R. and Hunter, S. (2014) 'InterProScan 5: genome-scale protein function classification', *Bioinformatics (Oxford, England)*, 30(9), pp. 1236-1240.

Joseph, B., Ramteke, P.W., Thomas, G. and Shrivastava, N. (2007) 'Standard review cold-active microbial lipases: A versatile tool for industrial applications', *Biotechnol Mol Biol Rev*, 2(2), pp. 39-48.

Julian, R.R. (2017) 'The Mechanism Behind Top-Down UVPD Experiments: Making Sense of Apparent Contradictions', *Journal of the American Society for Mass Spectrometry*, 28(9), pp. 1823-1826.

Juliette, L.Y., Hyman, M.R. and Arp, D.J. (1995) 'Roles of bovine serum albumin and copper in the assay and stability of ammonia monooxygenase activity in vitro', *Journal of bacteriology*, 177(17), pp. 4908-4913.

Kagia, K. and Liu, W.-T. (2014) 'The Family Dermatophilaceae', in Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E. and Thompson, F. (eds.) *The Prokaryotes: Actinobacteria*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 317-325.

Kahlke, T. and Thorvaldsen, S. (2012) 'Molecular characterization of cold adaptation of membrane proteins in the Vibrionaceae core-genome', *PloS one*, 7(12), pp. e51761-e51761.

Kananavičiūtė, R., Kvederavičiūtė, K., Dabkevičienė, D., Mackevičius, G. and Kuisienė, N. (2020) 'Collagen-like sequences encoded by extremophilic and extremotolerant bacteria', *Genomics*, 112(3), pp. 2271-2281.

Kang, H.Y., Kim, J.F., Kim, M.H., Park, S.H., Oh, T.K. and Hur, C.G. (2006) 'MELDB: a database for microbial esterases and lipases', *FEBS Lett*, 580(11), pp. 2736-40.

Kim, J.H., Lee, J., Park, J. and Gho, Y.S. (2015) 'Gram-negative and Gram-positive bacterial extracellular vesicles', *Seminars in Cell & Developmental Biology*, 40, pp. 97-104.

Kim, S.-H. and Shin, H.-S. (2010) 'Enhanced lipid degradation in an upflow anaerobic sludge blanket reactor by integration with an acidogenic reactor', *Water Environment Research*, 82(3), pp. 267-272.

Kleikamp, H.B.C., Pronk, M., Tugui, C., da Silva, L.G., Abbas, B., Lin, Y.M., van Loosdrecht, M.C.M. and Pabst, M. (2020) 'Quantitative profiling of microbial communities by de novo metaproteomics', *bioRxiv*, p. 2020.08.16.252924.

Kleiner, M. (2019) 'Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities', *mSystems*, 4(3), pp. e00115-19.

Kleiner, M., Dong, X., Hinzke, T., Wippler, J., Thorson, E., Mayer, B. and Strous, M. (2018) 'Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities', *Proceedings of the National Academy of Sciences*, 115(24), pp. E5576-E5584.

Kleiner, M., Thorson, E., Sharp, C.E., Dong, X., Liu, D., Li, C. and Strous, M. (2017) 'Assessing species biomass contributions in microbial communities via metaproteomics', *Nature Communications*, 8(1), p. 1558.

Kleiner, M., Wentrup, C., Lott, C., Teeling, H., Wetzel, S., Young, J., Chang, Y.-J., Shah, M., VerBerkmoes, N.C., Zarzycki, J., Fuchs, G., Markert, S., Hempel, K., Voigt, B., Becher, D., Liebeke, M., Lalk, M., Albrecht, D., Hecker, M., Schweder, T. and Dubilier, N. (2012) 'Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use', *Proceedings of the National Academy of Sciences*, 109(19), pp. E1173-E1182.

Komatsu, T., Ohya, K., Sawai, K., Odoi, J.O., Otsu, K., Ota, A., Ito, T., Kawai, M. and Maruyama, F. (2019) 'Draft genome sequences of Mycolicibacterium peregrinum isolated from a pig with lymphadenitis and from soil on the same Japanese pig farm', *BMC Research Notes*, 12(1), p. 341.

Kong, Z., Wu, J., Rong, C., Wang, T., Li, L., Luo, Z., Ji, J., Hanaoka, T., Sakemi, S., Ito, M., Kobayashi, S., Kobayashi, M., Qin, Y. and Li, Y.-Y. (2021) 'Large pilot-scale submerged anaerobic membrane bioreactor for the treatment of municipal wastewater and biogas production at 25°C', *Bioresource Technology*, 319, p. 124123.

Kovacic, F., Babic, N., Krauss, U. and Jaeger, K.-E. (2018) 'Classification of Lipolytic Enzymes from Bacteria', in Rojo, F. (ed.) *Aerobic Utilization of Hydrocarbons, Oils and Lipids*. Cham: Springer International Publishing, pp. 1-35.

Kuge, T., Fukushima, K., Matsumoto, Y., Abe, Y., Akiba, E., Haduki, K., Saito, H., Nitta, T., Kawano, A., Kawasaki, T., Matsuki, T., Kagawa, H., Motooka, D., Tsujino, K., Miki, M., Miki, K., Kitada, S., Nakamura, S., Iida, T. and Kida, H. (2020) 'Pulmonary disease caused by a newly identified mycobacterium: Mycolicibacterium toneyamachuris: a case report', *BMC Infectious Diseases*, 20(1), p. 888.

Kumar, D., Yadav, A.K. and Dash, D. (2017) 'Choosing an Optimal Database for Protein Identification from Tandem Mass Spectrometry Data', in Keerthikumar, S. and Mathivanan, S. (eds.) *Proteome Bioinformatics*. New York, NY: Springer New York, pp. 17-29.

Kunau, W.H., Dommes, V. and Schulz, H. (1995) 'beta-oxidation of fatty acids in mitochondria, peroxisomes, and bacteria: a century of continued progress', *Prog Lipid Res*, 34(4), pp. 267-342.

Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K. and Hugenholtz, P. (2008) 'A bioinformatician's guide to metagenomics', *Microbiology and molecular biology reviews : MMBR*, 72(4), pp. 557-578.

Lee, E.-Y., Choi, D.-S., Kim, K.-P. and Gho, Y.S. (2008) 'Proteomics in gram-negative bacterial outer membrane vesicles', *Mass Spectrometry Reviews*, 27(6), pp. 535-555.

Lee, E.-Y., Choi, D.-Y., Kim, D.-K., Kim, J.-W., Park, J.O., Kim, S., Kim, S.-H., Desiderio, D.M., Kim, Y.-K., Kim, K.-P. and Gho, Y.S. (2009a) 'Gram-positive bacteria produce membrane vesicles: Proteomics-based characterization of *Staphylococcus aureus*-derived membrane vesicles', *PROTEOMICS*, 9(24), pp. 5425-5436.

Lee, E.Y., Choi, D.Y., Kim, D.K., Kim, J.W., Park, J.O., Kim, S., Kim, S.H., Desiderio, D.M., Kim, Y.K., Kim, K.P. and Gho, Y.S. (2009b) 'Gram-positive bacteria produce membrane vesicles: proteomics-based characterization of *Staphylococcus aureus*-derived membrane vesicles', *Proteomics*, 9(24), pp. 5425-36.

Lee, H.-K., Ahn, M.-J., Kwak, S.-H., Song, W.-H. and Jeong, B.-C. (2003) 'Purification and characterization of cold active lipase from psychrotrophic Aeromonas sp. LPB 4', *Journal of Microbiology*, 41(1), pp. 22-27.

Lee, H.-M., Vázquez-Bermúdez, M.F. and Marsac, N.T.d. (1999) 'The gobal nitrogen regulator *NtcA* regulates transcription of the signal transducer PII (*GlnB*) and influences its

phosphorylation level in response to nitrogen and carbon supplies in the *Cyanobacterium Synechococcus sp. Strain PCC 7942*', *Journal of Bacteriology*, 181(9), pp. 2697-2702.

Lee, J., Kim, O.Y. and Gho, Y.S. (2016a) 'Proteomic profiling of Gram-negative bacterial outer membrane vesicles: Current perspectives', *PROTEOMICS – Clinical Applications*, 10(9-10), pp. 897-909.

Lee, J., Kim, O.Y. and Gho, Y.S. (2016b) 'Proteomic profiling of Gram-negative bacterial outer membrane vesicles: Current perspectives', *Proteomics– Clinical Applications*, 10(9-10), pp. 897-909.

Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P. and Chatonnet, A. (2013) 'ESTHER, the database of the α/β-hydrolase fold superfamily of proteins: tools to explore diversity of functions', *Nucleic Acids Research*, 41(D1), pp. D423-D429.

Lettinga, G., Pol, L.W.H., Koster, I.W., Wiegant, W.M., De Zeeuw, W.J., Rinzema, A., Grin, P.C., Roersma, R.E. and Hobma, S.W. (1984) 'High-rate anaerobic waste-water treatment using the uasb reactor under a wide range of temperature conditions', *Biotechnology and Genetic Engineering Reviews*, 2(1), pp. 253-284.

Lettinga, G., Rebac, S. and Zeeman, G. (2001) 'Challenge of psychrophilic anaerobic wastewater treatment', *Trends in Biotechnology*, 19(9), pp. 363-370.

Li, H., Joh, Y.S., Kim, H., Paek, E., Lee, S.-W. and Hwang, K.-B. (2016) 'Evaluating the effect of database inflation in proteogenomic search on sensitive and reliable peptide identification', *BMC Genomics*, 17(13), p. 1031.

Li, W.-W. and Yu, H.-Q. (2016) 'Advances in Energy-Producing Anaerobic Biotechnologies for Municipal Wastewater Treatment', *Engineering*, 2(4), pp. 438-446.

Lim, K., Evans, P.J. and Parameswaran, P. (2019) 'Long-term performance of a pilot-scale gas-sparged anaerobic membrane bioreactor under ambient temperatures for holistic wastewater treatment', *Environmental Science & Technology*, 53(13), pp. 7347-7354.

Liu, Y., Defourny, K.A.Y., Smid, E.J. and Abee, T. (2018) 'Gram-Positive Bacterial Extracellular Vesicles and Their Impact on Health and Disease', *Frontiers in Microbiology*, 9(1502).

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C.J. and Marchler-Bauer, A. (2020) 'CDD/SPARCLE: the conserved domain database in 2020', *Nucleic acids research*, 48(D1), pp. D265-D268.

Mackie, R.I., White, B.A. and Bryant, M.P. (1991) 'Lipid metabolism in anaerobic ecosystems', *Critical Reviews in Microbiology*, 17(6), pp. 449-479.

Maktabifard, M., Zaborowska, E. and Makinia, J. (2018) 'Achieving energy neutrality in wastewater treatment plants through energy savings and enhancing renewable energy production', *Reviews in Environmental Science and Bio/Technology*, 17(4), pp. 655-689.

Maktabifard, M., Zaborowska, E. and Makinia, J. (2019) 'Evaluating the effect of different operational strategies on the carbon footprint of wastewater treatment plants – case studies from northern Poland', *Water Science and Technology*, 79(11), pp. 2211-2220.

Maleki, E., Bokhary, A., Leung, K. and Liao, B.Q. (2019) 'Long-term performance of a submerged anaerobic membrane bioreactor treating malting wastewater at room temperature (23 ± 1 °C)', *Journal of Environmental Chemical Engineering*, 7(4), p. 103269.

McAteer, P.G., Christine Trego, A., Thorn, C., Mahony, T., Abram, F. and O'Flaherty, V. (2020) 'Reactor configuration influences microbial community structure during high-rate, low-temperature anaerobic treatment of dairy wastewater', *Bioresource Technology*, 307, p. 123221.

McDowall, J. and Hunter, S. (2011) 'InterPro Protein Classification', in Wu, C.H. and Chen, C. (eds.) *Bioinformatics for Comparative Proteomics*. Totowa, NJ: Humana Press, pp. 37-47.

Messaoudi, A., Belguith, H., Ghram, I. and Hamida, J.B. (2011) 'LIPABASE: a database for 'true' lipase family enzymes', *Int J Bioinform Res Appl*, 7(4), pp. 390-401.

Mhetras, N., Mapare, V. and Gokhale, D. (2021) 'Cold Active Lipases: Biocatalytic Tools for Greener Technology', *Applied Biochemistry and Biotechnology*.

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L.-P., Mushayamaha, T. and Thomas, P.D. (2020) '*PANTHER* version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API', *Nucleic Acids Research*, 49(D1), pp. D394-D403.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, Gustavo A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D. and Bateman, A. (2020) '*Pfam*: The protein families database in 2021', *Nucleic Acids Research*, 49(D1), pp. D412-D419.

Mitchell, A.L., Scheremetjew, M., Denise, H., Potter, S., Tarkowska, A., Qureshi, M., Salazar, G.A., Pesseat, S., Boland, M.A., Hunter, Fiona M I., ten Hoopen, P., Alako, B., Amid, C., Wilkinson, D.J., Curtis, T.P., Cochrane, G. and Finn, R.D. (2017) 'EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies', *Nucleic Acids Research*, 46(D1), pp. D726-D735.

Mitra, R., Xu, T., Xiang, H. and Han, J. (2020) 'Current developments on polyhydroxyalkanoates synthesis by using halophiles as a promising cell factory', *Microbial Cell Factories*, 19(1), p. 86.

Mølgaard, A., Kauppinen, S. and Larsen, S. (2000) 'Rhamnogalacturonan acetylesterase elucidates the structure and function of a new family of hydrolases', *Structure*, 8(4), pp. 373-383.

Muhammadi, Shabina, Afzal, M. and Hameed, S. (2015) 'Bacterial polyhydroxyalkanoates-eco-friendly next generation plastic: Production, biocompatibility, biodegradation, physical properties and applications', *Green Chemistry Letters and Reviews*, 8(3-4), pp. 56-77.

Muller, E.E.L., Sheik, A.R. and Wilmes, P. (2014) 'Lipid-based biofuel production from wastewater', *Current Opinion in Biotechnology*, 30, pp. 9-16.

Neidigh, J.W., Fesinmeyer, R.M. and Andersen, N.H. (2002) 'Designing a 20-residue protein', *Nature Structural Biology*, 9(6), pp. 425-430.

Nielsen, P.H., Roslev, P., Dueholm, T.E. and Nielsen, J.L. (2002) 'Microthrix parvicella, a specialized lipid consumer in anaerobic–aerobic activated sludge plants', *Water Science and Technology*, 46(1-2), pp. 73-80.

Oshiki, M., Satoh, H., Mino, T. and Onuki, M. (2008) 'PHA-accumulating microorganisms in full-scale wastewater treatment plants', *Water Science and Technology*, 58(1), pp. 13-20.

Ozgun, H., Dereli, R.K., Ersahin, M.E., Kinaci, C., Spanjers, H. and van Lier, J.B. (2013) 'A review of anaerobic membrane bioreactors for municipal wastewater treatment: Integration options, limitations and expectations', *Separation and Purification Technology*, 118, pp. 89-104.

Pandurangan, A.P., Stahlhacke, J., Oates, M.E., Smithers, B. and Gough, J. (2018) 'The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver', *Nucleic Acids Research*, 47(D1), pp. D490-D494.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A. and Hugenholtz, P. (2018) 'A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life', *Nature Biotechnology*, 36(10), pp. 996-1004.

Pauli, G., Ehring, R. and Overath, P. (1974) 'Fatty Acid Degradation in Escherichia coli: Requirement of Cyclic Adenosine Monophosphate and Cyclic Adenosine Monophosphate Receptor Protein for Enzyme Synthesis', *Journal of Bacteriology*, 117(3), p. 1178.

Penfield, T. (2017) *Microbial communities in an anaerobic membrane bioreactor (AnMBR) treating domestic wastewater at ambient temperatures in a temperate climate*. Kansas State University.

Petropoulos, E., Dolfing, J., Davenport, R.J., Bowen, E.J. and Curtis, T.P. (2017) 'Developing cold-adapted biomass for the anaerobic treatment of domestic wastewater at low temperatures (4, 8 and 15°C) with inocula from cold environments', *Water Research*, 112, pp. 100-109.

Petropoulos, E., Dolfing, J., Yu, Y., Wade, M.J., Bowen, E.J., Davenport, R.J. and Curtis, T.P. (2018) 'Lipolysis of domestic wastewater in anaerobic reactors operating at low temperatures', *Environmental Science: Water Research & Technology*, 4(7), pp. 1002-1013.

Petropoulos, E., Shamurad, B., Tabraiz, S., Yu, Y., Davenport, R., Curtis, T.P. and Dolfing, J. (2021) 'Sewage treatment at 4 °C in anaerobic upflow reactors with and without a membrane – performance, function and microbial diversity', *Environmental Science: Water Research & Technology*, 7(1), pp. 156-171.

Petropoulos, E., Yu, Y., Tabraiz, S., Yakubu, A., Curtis, T.P. and Dolfing, J. (2019) 'High rate domestic wastewater treatment at 15 °C using anaerobic reactors inoculated with cold-adapted sediments/soils – shaping robust methanogenic communities', *Environmental Science: Water Research & Technology*, 5(1), pp. 70-82.

Pospíšil, J., Vítovská, D., Kofroňová, O., Muchová, K., Šanderová, H., Hubálek, M., Šiková, M., Modrák, M., Benada, O., Barák, I. and Krásný, L. (2020) 'Bacterial nanotubes as a manifestation of cell death', *Nature Communications*, 11(1), p. 4963.

R Julian, R. (2017) 'The Mechanism Behind Top-Down UVPD Experiments: Making Sense of Apparent Contradictions', *Journal of the American Society for Mass Spectrometry*, 28(9), pp. 1823-1826.

Ram, S.K., Kumar, L.R., Tyagi, R.D. and Drogui, P. (2018) 'Techno-economic evaluation of simultaneous production of extra-cellular polymeric substance (EPS) and lipids by *Cloacibacterium normanense NK6* using crude glycerol and sludge as substrate', *Water Science and Technology*, 77(9), pp. 2228-2241.

Rappsilber, J., Mann, M. and Ishihama, Y. (2007) 'Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips', *Nature Protocols*, 2(8), pp. 1896-1906.

Ribera-Pi, J., Campitelli, A., Badia-Fabregat, M., Jubany, I., Martínez-Lladó, X., McAdam, E., Jefferson, B. and Soares, A. (2020) 'Hydrolysis and methanogenesis in UASB-AnMBR treating municipal wastewater under psychrophilic conditions: importance of reactor configuration and inoculum', *Frontiers in bioengineering and biotechnology*, 8, pp. 567695-567695.

Riley, M., Staley, J.T., Danchin, A., Wang, T.Z., Brettin, T.S., Hauser, L.J., Land, M.L. and Thompson, L.S. (2008) 'Genomics of an extreme psychrophile, Psychromonas ingrahamii', *BMC Genomics*, 9(1), p. 210.

Ryu, H.S., Kim, H.K., Choi, W.C., Kim, M.H., Park, S.Y., Han, N.S., Oh, T.K. and Lee, J.K. (2006) 'New cold-adapted lipase from Photobacterium lipolyticum sp. nov. that is closely related to filamentous fungal lipases', *Applied Microbiology and Biotechnology*, 70(3), pp. 321-326.

Salvador López, J.M. and Van Bogaert, I.N.A. (2021) 'Microbial fatty acid transport proteins and their biotechnological potential', *Biotechnology and Bioengineering*, 118(6), pp. 2184-2201.

Sayed, S.Z. and Lettinga, W. (1984) *Agric. Wastes*, 11, p. 197.

Schuchardt, S. and Sickmann, A. (2007) 'Protein identification using mass spectrometry: A method overview', in Baginsky, S. and Fernie, A.R. (eds.) *Plant Systems Biology*. Basel: Birkhäuser Basel, pp. 141-170.

Sedlacek, C.J., Nielsen, S., Greis, K.D., Haffey, W.D., Revsbech, N.P., Ticak, T., Laanbroek, H.J., Bollmann, A. and Vieille, C. (2016) 'Effects of bacterial community members on the proteome of the ammonia-oxidizing bacterium *nitrosomonas sp. strain is79*', *Applied and Environmental Microbiology*, 82(15), pp. 4776-4788.

Servant, F., Bru, C., Carrère, S., Courcelle, E., Gouzy, J., Peyruc, D. and Kahn, D. (2002) '*ProDom*: Automated clustering of homologous domains', *Briefings in Bioinformatics*, 3(3), pp. 246-251.

Seviour, T., Derlon, N., Dueholm, M.S., Flemming, H.-C., Girbal-Neuhauser, E., Horn, H., Kjelleberg, S., van Loosdrecht, M.C.M., Lotti, T., Malpei, M.F., Nerenberg, R., Neu, T.R.,

Paul, E., Yu, H. and Lin, Y. (2019) 'Extracellular polymeric substances of biofilms: Suffering from an identity crisis', *Water Research*, 151, pp. 1-7.

Sharma, P.K., Mohanan, N., Sidhu, R. and Levin, D.B. (2019) 'Colonization and degradation of polyhydroxyalkanoates by lipase-producing bacteria', *Canadian Journal of Microbiology*, 65(6), pp. 461-475.

Sigrist, C.J., de Castro, E., Cerutti, L., Cuche, B.A., Hulo, N., Bridge, A., Bougueleret, L. and Xenarios, I. (2013) 'New and continuing developments at *PROSITE*', *Nucleic Acids Res*, 41(Database issue), pp. D344-7.

Singleton, C.M., Petriglieri, F., Kristensen, J.M., Kirkegaard, R.H., Michaelsen, T.Y., Andersen, M.H., Kondrotaite, Z., Karst, S.M., Dueholm, M.S., Nielsen, P.H. and Albertsen, M. (2021) 'Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing', *Nature Communications*, 12(1), p. 2009.

Smith, A.L., Skerlos, S.J. and Raskin, L. (2013) 'Psychrophilic anaerobic membrane bioreactor treatment of domestic wastewater', *Water Research*, 47(4), pp. 1655-1665.

Smith, A.L., Skerlos, S.J. and Raskin, L. (2015) 'Anaerobic membrane bioreactor treatment of domestic wastewater at psychrophilic temperatures ranging from 15 °C to 3 °C', *Environmental Science: Water Research & Technology*, 1(1), pp. 56-64.

Smith, A.L., Stadler, L.B., Love, N.G., Skerlos, S.J. and Raskin, L. (2012) 'Perspectives on anaerobic membrane bioreactor treatment of domestic wastewater: A critical review', *Bioresource Technology*, 122, pp. 149-159.

Soares, R.B., Memelli, M.S., Roque, R.P. and Gonçalves, R.F. (2017) 'Comparative analysis of the energy consumption of different wastewater treatment plants', *International Journal of Architecture, Arts and Applications*, 3(6), pp. 79-86.

Speda, J., Johansson, M.A., Carlsson, U. and Karlsson, M. (2017) 'Assessment of sample preparation methods for metaproteomics of extracellular proteins', *Analytical Biochemistry*, 516, pp. 23-36.

Srivastava, A., Murugaiyan, J., Garcia, J.A.L., De Corte, D., Hoetzinger, M., Eravci, M., Weise, C., Kumar, Y., Roesler, U., Hahn, M.W. and Grossart, H.-P. (2020) 'Combined methylome, transcriptome and proteome analyses document rapid acclimatization of a bacterium to environmental changes', *Frontiers in Microbiology*, 11(2197).

St. James, A.R. and Richardson, R.E. (2020) 'Ecogenomics reveals community interactions in a long-term methanogenic bioreactor and a rapid switch to sulfate-reducing conditions', *FEMS Microbiology Ecology*, 96(5).

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A. and Narechania, A. (2003) 'PANTHER: A Library of Protein Families and Subfamilies Indexed by Function', *Genome Research*, 13(9), pp. 2129-2141.

Tiessen, A., Pérez-Rodríguez, P. and Delaye-Arredondo, L.J. (2012) 'Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial

species reveals a negative correlation between protein size and protein number, thus providing insight into the evolution of proteomes', *BMC Research Notes*, 5(1), p. 85.

Toledo, A. and Benach, J.L. (2015) 'Hijacking and Use of Host Lipids by Intracellular Pathogens', *Microbiology spectrum*, 3(6), pp. 10.1128/microbiolspec.VMBF-0001-2014.

Torregrosa-Crespo, J., González-Torres, P., Bautista, V., Esclapez, J.M., Pire, C., Camacho, M., Bonete, M.J., Richardson, D.J., Watmough, N.J. and Martínez-Espinosa, R.M. (2017) 'Analysis of multiple haloarchaeal genomes suggests that the quinone-dependent respiratory nitric oxide reductase is an important source of nitrous oxide in hypersaline environments', *Environmental Microbiology Reports*, 9(6), pp. 788-796.

Tsatsaronis, J.A., Franch-Arroyo, S., Resch, U. and Charpentier, E. (2018) 'Extracellular Vesicle RNA: A Universal Mediator of Microbial Communication?', *Trends in Microbiology*, 26(5), pp. 401-410.

Tufail, S., Munir, S. and Jamil, N. (2017) 'Variation analysis of bacterial polyhydroxyalkanoates production using saturated and unsaturated hydrocarbons', *Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology]*, 48(4), pp. 629-636.

Ueki, A., Akasaka, H. and Ueki, K. (2015) 'Propionicimonas', in *Bergey's Manual of Systematics of Archaea and Bacteria*. pp. 1-4.

UniProt, C. (2021) 'UniProt: the universal protein knowledgebase in 2021', *Nucleic acids research*, 49(D1), pp. D480-D489.

van den Berg, B. (2005) 'The FadL family: unusual transporters for unusual substrates', *Current Opinion in Structural Biology*, 15(4), pp. 401-407.

van Dijk, E., Hoogeveen, A. and Abeln, S. (2015) 'The Hydrophobic Temperature Dependence of Amino Acids Directly Calculated from Protein Structures', *PLOS Computational Biology*, 11(5), p. e1004277.

Van Lier, J.B., Mahmoud, N. and Zeeman, G. (2008) 'Anaerobic Wastewater Treatment', in Henze, M., Loosdrecht, M.C.M.v., Ekama, G.A. and Brdjanovic, D. (eds.) *Biological wastewater treatment: principles, modelling and design*. London, UK: IWA publishing.

van Niel, G., D'Angelo, G. and Raposo, G. (2018) 'Shedding light on the cell biology of extracellular vesicles', *Nature Reviews Molecular Cell Biology*, 19(4), pp. 213-228.

Verger, R. (1997) '‘Interfacial activation’ of lipases: facts and artifacts', *Trends in Biotechnology*, 15(1), pp. 32-38.

Verma, S., Meghwanshi, G.K. and Kumar, R. (2021) 'Current perspectives for microbial lipases from extremophiles and metagenomics', *Biochimie*, 182, pp. 23-36.

Wang, J. and Chu, L. (2016) 'Biological nitrate removal from water and wastewater by solid-phase denitrification process', *Biotechnology Advances*, 34(6), pp. 1103-1112.

Wang, R., Borazjani, A., Matthews, A.T., Mangum, L.C., Edelmann, M.J. and Ross, M.K. (2013) 'Identification of palmitoyl protein thioesterase 1 in human THP1 monocytes and

macrophages and characterization of unique biochemical activities for this enzyme', *Biochemistry*, 52(43), pp. 7559-7574.

Ward, B.B. (2008) 'Chapter 5 - Nitrification in marine systems', in Capone, D.G., Bronk, D.A., Mulholland, M.R. and Carpenter, E.J. (eds.) *In nitrogen in the marine environment (Second Edition)*. San Diego: Academic Press, pp. 199-261.

Wessel, D. and Flügge, U.I. (1984) 'A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids', *Analytical Biochemistry*, 138(1), pp. 141-143.

Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C. and Gough, J. (2009) 'SUPERFAMILY--sophisticated comparative genomics, data mining, visualization and phylogeny', *Nucleic acids research*, 37(Database issue), pp. D380-D386.

Wingender, J., Neu, T.R. and Flemming, H.-C. (1999) 'What are Bacterial Extracellular Polymeric Substances?', in Wingender, J., Neu, T.R. and Flemming, H.-C. (eds.) *Microbial Extracellular Polymeric Substances: Characterization, Structure and Function*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1-19.

Woith, E., Fuhrmann, G. and Melzig, M.F. (2019) 'Extracellular Vesicles-Connecting Kingdoms', *International journal of molecular sciences*, 20(22), p. 5695.

Won, S.J., Cheung See Kit, M. and Martin, B.R. (2018) 'Protein depalmitoylases', *Critical reviews in biochemistry and molecular biology*, 53(1), pp. 83-98.

Yang, C., Zhang, W., Liu, R., Zhang, C., Gong, T., Li, Q., Wang, S. and Song, C. (2013) 'Analysis of polyhydroxyalkanoate (PHA) synthase gene and PHA-producing bacteria in activated sludge that produces PHA containing 3-hydroxydodecanoate', *FEMS Microbiology Letters*, 346(1), pp. 56-64.

Yang, Y., Zang, Y., Hu, Y., Wang, X.C. and Ngo, H.H. (2020) 'Upflow anaerobic dynamic membrane bioreactor (AnDMBR) for wastewater treatment at room temperature and short HRTs: Process characteristics and practical applicability', *Chemical Engineering Journal*, 383, p. 123186.

Yao, J. and Rock, C.O. (2017) 'Exogenous fatty acid metabolism in bacteria', *Biochimie*, 141, pp. 30-39.

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.-H., Whitman, W.B., Euzéby, J., Amann, R. and Rosselló-Móra, R. (2014) 'Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences', *Nature Reviews Microbiology*, 12(9), pp. 635-645.

Yin, W., Wang, K., Xu, J., Wu, D. and Zhao, C. (2018) 'The performance and associated mechanisms of carbon transformation (PHAs, polyhydroxyalkanoates) and nitrogen removal for landfill leachate treatment in a sequencing batch biofilm reactor (SBBR)', *RSC Advances*, 8(74), pp. 42329-42336.

Zeeman, G. and Lettinga, G. (1999) 'The role of anaerobic digestion of domestic sewage in closing the water and nutrient cycle at community level', *Water Science and Technology*, 39(5), pp. 187-194.

Zenaidee, M.A., Lantz, C., Perkins, T., Jung, W., Loo, R.R.O. and Loo, J.A. (2020) 'Internal Fragments Generated by Electron Ionization Dissociation Enhance Protein Top-Down Mass Spectrometry', *Journal of the American Society for Mass Spectrometry*, 31(9), pp. 1896-1902.

Zeng, X., Xiao, X., Wang, P. and Wang, F. (2004) 'Screening and characterization of psychrotrophic, lipolytic bacteria from deep-sea sediments', *Journal of Microbiology and Biotechnology*, 14(5), pp. 952-958.

Appendices

Appendix A. Details of reactor set-up and performance.

Eight 1 litre Quickfit® anaerobic membrane bioreactors (AnMBR) with putative psychrophilic biomass were developed to operate at 4 ℃ and 15 ℃ to treat domestic wastewater (Figure 0-1). Putative psychrophilic biomass was an equal mixture (final concentration: 11.5 g/l mixed liquor suspended solids) of sediment of Lake Geneva N 46°23′04″, E 6°25′07″ (minimum and maximum annual average temperature: −11,17 °C) and soils from Svalbard N 78° and E 11, 15, and 16° (minimum and maximum annual average temperature: −16,6 °C). The sediment from Lake Geneva had the average temperature of 4.8 ℃ and were collected in August 2011 (200 m depth). By contrast, soils from Svalbard had the average temperature of 3 ℃ and were collected in September 2009. The reactors were fed with a primary domestic wastewater from an activated sludge plant (Tudhoe Mill, County Durham, UK) at two conditions of Sterile or Non-sterile. The Sterile feed was subjected to a pre-treatment with an ultraviolent lamp (irradiation dose of 110 kJ cm^{-2}) to exclude the mesophilic microbial community of the activated sludge plant. This way, we could compare the performance of the putative psychrophilic community. The membrane was hydrophobic hollow-fiber polyvinylidene difluoride (PVDF) with the following properties: pore size: 0.1 µm, fiber diameter: 1 mm, membrane area: 0.022 m^2. The hydraulic retention time (HRT), organic loading rate (OLR), up flow velocity and membrane flux were set at 60 hrs, 0.1 $kgCOD.m^3.d^{-1}$, 0.8 m h^{-1} and 0.4 L m^{2-} h^{-1}, respectively to minimise the biofouling, and the membrane backwashing (30 min relaxation per day and 30 min backwashing every 2 HRTs).

The reactors used in this study were first developed and adapted to both operating temperatures (4 ℃ and 15 ℃) and UV-treated feed in a series of batch and continuous experiments that lasted for 1073 days as described by Petropoulos *et al.* (2017), Petropoulos *et al.* (2018), Petropoulos *et al.* (2019), and Petropoulos *et al.* (2021). Reactor's performance, specific

methane production rate, and volatile fatty acid analyses are discussed in these articles. In brief, the AnMBRs in continuous operation had more than 86% COD removal which was slightly higher than the COD removal efficiency of the UASB with the same biomass and operational conditions. 6.29 and 10.25 fmol CH_4. $Cell^{-1}$. Day^{-1} was produced at 4 °C and 15 °C, respectively (Petropoulos *et al.*, 2019; Petropoulos *et al.*, 2021). In the present study, prior to sampling, the reactors were re-acclimated to the operational conditions and worked continuously at steady state for two months. We sampled from both the liquid bulk and the biofilm formed on the membranes on Day 65. The chemical oxygen demand (COD) of the feed was measured based on the standard methods of the American Public Health Association (APHA, 2006). The COD of the feed varied throughout the year (100-800 mg/l) and at the time of re-acclimation for the current study the COD was 281±13.2 mg/l. The feed (primary influent) had similar characteristics to the one Petropoulos *et al.* (2018) measured and was composed of 60% carbohydrates, 38% lipids, and less than 2% proteins. At the time of sampling, lipid content of the Sterile and Non-sterile feed, were 0.62 ± 0.07 gr/l and 0.55 ± 0.0 gr/l, respectively, measured gravimetrically based on Bligh and Dyer (1959) protocol.

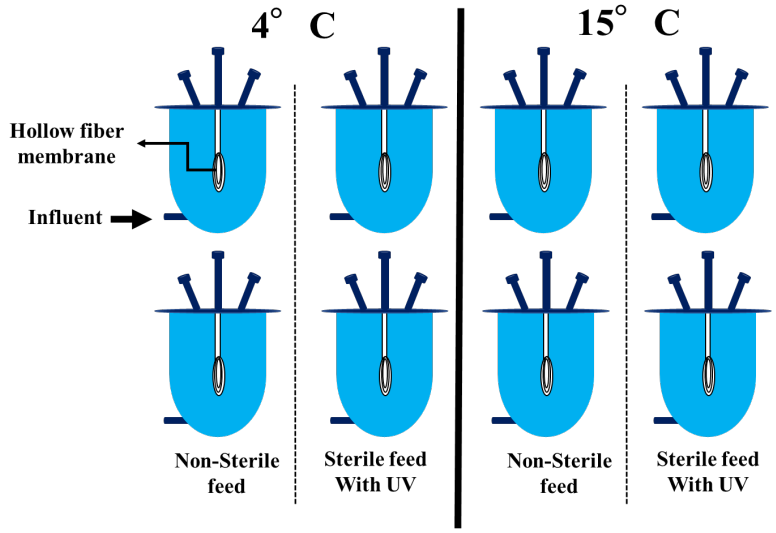


Figure 0-1. Schematic diagram of anaerobic membrane bioreactor with psychrophilic biomass working at 4 and 15 °C.

Appendix B. List of hydrolytic enzymes that was searched against metagenomics data and putative lipolytic MAGs.

Enzyme class	Enzyme name	EC number	Gene counts	
			Whole metagenome	MAGs
Carbohydrate degrading enzymes	β-galactosidase	3.2.1.23	5957	34
	β-glucosidase	3.2.1.21	5307	60
	α-galactosidase	3.2.1.22	4473	39
	β-hexosaminidase	3.2.1.52	2980	56
	non-reducing end α-L-arabinofuranosidase	3.2.1.55	2153	7
	lysozyme	3.2.1.17	1865	28
	6-phospho-β-glucosidase	3.2.1.86	1692	10
	Endo-β-xylanase	3.2.1.8	1665	39
	α-amylase	3.2.1.1	1482	42
	Cellulase	3.2.1.4	1276	43
	oligo-1,6-glucosidase	3.2.1.10	1241	25
	non-reducing end β-L-arabinofuranosidase	3.2.1.185	1128	5
	α,α-trehalase	3.2.1.28	1017	21
	α-D-xyloside xylohydrolase	3.2.1.177	1014	9
	licheninase	3.2.1.73	994	16
	xylan 1,4-β-xylosidase	3.2.1.37	862	7
	exo-α-sialidase	3.2.1.18	834	14
	α-N-acetylgalactosaminidase	3.2.1.49	805	10
	Levanase	3.2.1.80	781	3
	glucan 1,4-β-glucosidase	3.2.1.74	655	6
	neopullulanase	3.2.1.135	634	12
	α-glucosidase	3.2.1.20	597	5
	UDP-N,N'-diacetylbacillosamine 2-epimerase	3.2.1.184	585	1
	unsaturated rhamnogalacturonyl hydrolase	3.2.1.172	538	4
	6-phospho-β-galactosidase	3.2.1.85	526	0
	exo-1,4-β-D-glucosaminidase	3.2.1.165	491	14
	galacturan 1,4-α-galacturonidase	3.2.1.67	477	1
	Mannosylglycerate hydrolase	3.2.1.170	403	18
	β-glucuronidase	3.2.1.31	378	9
	glucan 1,4-α-glucosidase	3.2.1.3	375	6
	β-fructofuranosidase	3.2.1.26	374	8
	cyclomaltodextrinase	3.2.1.54	338	7
	pullulanase	3.2.1.41	326	2
	mannan endo-1,4-β-mannosidase	3.2.1.78	274	3
	Endo-β-glucosidase	3.2.1.39	246	6
	chitinase	3.2.1.14	243	11
	4-α-D- (1→4)-α-D-glucano trehalose trehalohydrolase	3.2.1.141	231	8
	sulfoquinovosidase	3.2.1.199	219	9
	cellulose 1,4-β-cellobiosidase	3.2.1.91	204	5

Enzyme class	Enzyme name	EC number	Gene counts	
			Whole metagenome	MAGs
Carbohydrate degrading enzymes	gellan tetrasaccharide unsaturated glucuronyl hydrolase	3.2.1.179	200	1
	unsaturated chondroitin disaccharide hydrolase	3.2.1.180	199	0
	xylan α-1,2-glucuronosidase	3.2.1.131	169	0
	Arabinosidase	3.2.1.99	158	0
	xyloglucan-specific endo-beta-1,4-glucanase	3.2.1.151	122	0
	keratan-sulfate endo-1,4-β-galactosidase	3.2.1.103	106	0
	UDP-N-acetylglucosamine 2-epimerase	3.2.1.183	106	0
	α,α-phosphotrehalase	3.2.1.93	94	0
	arabinogalactan endo-β-1,4-galactanase	3.2.1.89	90	0
	oligosaccharide reducing-end xylanase	3.2.1.156	69	0
	glucan 1,3-β-glucosidase	3.2.1.58	65	0
	β-porphyranase	3.2.1.178	62	2
	glucan 1,4-α-maltohexaosidase	3.2.1.98	61	0
	maltose-6'-phosphate glucosidase	3.2.1.122	58	0
	chitosanase	3.2.1.132	51	2
	xylan 1,3-β-xylosidase	3.2.1.72	44	0
	(Ara-f)3-Hyp β-L-arabinobiosidase	3.2.1.187	44	0
	κ-carrageenase	3.2.1.83	43	3
	limit dextrin α-1,6-maltotetraose-hydrolase	3.2.1.196	43	0
	endo-polygalacturonase	3.2.1.15	42	1
	glucan 1,6-α-glucosidase	3.2.1.69	40	0
	exo-poly-α-galacturonosidase	3.2.1.82	40	0
	glucan 1,4-α-maltotetraohydrolase	3.2.1.60	36	1
	isoamylase	3.2.1.68	36	2
	glucuronoarabinoxylan endo-1,4-β-xylanase	3.2.1.136	35	3
	protein O-GlcNAcase	3.2.1.169	33	4
	λ-carrageenase	3.2.1.162	31	1
	α-agarase	3.2.1.158	30	33
	mannosyl-glycoprotein endo-β-N-acetylglucosaminidase	3.2.1.96	25	0
	endo-1,3-β-xylanase	3.2.1.32	23	0
	glucan 1,6-α-isomaltosidase	3.2.1.94	22	0
	glucan 1,4-α-maltohydrolase	3.2.1.133	21	3
	2,6-β-fructan 6-levanbiohydrolase	3.2.1.64	15	0
	dextranase	3.2.1.11	11	0
	β-agarase	3.2.1.81	10	0
	endo-α-N-acetylgalactosaminidase	3.2.1.97	8	0
	hyaluronoglucosaminidase	3.2.1.35	3	0
	blood-group-substance endo-1,4-β-galactosidase	3.2.1.102	3	1
	ι-carrageenase	3.2.1.157	3	0
	β-amylase	3.2.1.2	2	3

Enzyme class	Enzyme name	EC number	Gene counts	
			Whole metagenome	MAGs
Lipolytic enzymes	Triacylglycerol lipase	3.1.1.3	903	78
	Carboxylesterase	3.1.1.1	2997	109
	Acylglycerol lipase	3.1.1.23	2150	73
	Phospholipase D	3.1.4.4	463	9
	Putative phospholipase	3.1.1.32	354	1
	Phospholipase C	3.1.4.3	126	5
	Lysophospholipase	3.1.1.5	99	3
	Lipoprotein lipase	3.1.1.34	0	0
	Phospholipase A2	3.1.1.4	0	0
	Phosphatidate phoshphohydrolase	3.1.3.4	0	0
Proteases	leucyl aminopeptidase	3.4.11.1	12094	48
	enteropeptidase	3.4.21.9	8870	0
	endopeptidase Clp	3.4.21.92	8862	70
	thrombin	3.4.21.5	7978	0
	endopeptidase La	3.4.21.53	7916	75
	repressor LexA	3.4.21.88	7490	46
	methionyl aminopeptidase	3.4.11.18	6878	54
	chymotrypsin	3.4.21.1	6037	0
	serine-type D-Ala-D-Ala carboxypeptidase	3.4.16.4	5941	146
	acrosin	3.4.21.10	5843	0
	gastricsin	3.4.23.3	5072	0
	signal peptidase II	3.4.23.36	5072	45
	signal peptidase I	3.4.21.89	4036	57
	peptidase Do	3.4.21.107	3437	58
	tripeptide aminopeptidase	3.4.11.4	2078	10
	carboxypeptidase A	3.4.17.1	1941	0
	Xaa-Pro aminopeptidase	3.4.11.9	1931	25
	membrane alanyl aminopeptidase	3.4.11.2	1912	55
	D-stereospecific aminopeptidase	3.4.11.19	1792	1
	interstitial collagenase	3.4.24.7	1756	0
	HslU—HslV peptidase	3.4.25.2	1687	12
	cytosol nonspecific dipeptidase	3.4.13.18	1666	16
	oligopeptidase A	3.4.24.70	1431	7
	Xaa-Pro dipeptidase	3.4.13.9	1369	6
	C-terminal processing peptidase	3.4.21.102	1301	38
	carboxypeptidase Taq	3.4.17.19	1295	7
	Serine-type D-Ala-D-Ala carboxypeptidase	3.4.13.22	1274	20
	prolyl aminopeptidase	3.4.11.5	1258	26
	acylaminoacyl-peptidase	3.4.19.1	1152	0
	chymotrypsin C	3.4.21.2	1068	0
	prolyl oligopeptidase	3.4.21.26	1068	15
	rhomboid protease	3.4.21.105	1057	48
	dipeptidyl-peptidase I	3.4.14.1	1034	0

Enzyme class	Enzyme name	EC number	Gene counts	
			Whole metagenome	MAGs
Proteases	glutathione γ-glutamate hydrolase	3.4.19.13	1002	0
	peptidyl-dipeptidase Dcp	3.4.15.5	874	12
	dipeptidyl-peptidase IV	3.4.14.5	854	8
	bacterial leucyl aminopeptidase	3.4.11.10	800	16
	dipeptidase E	3.4.13.21	651	9
	β-peptidyl aminopeptidase	3.4.11.25	634	6
	prolyltripeptidyl aminopeptidase	3.4.14.12	568	8
	β-aspartyl-peptidase	3.4.19.5	538	5
	oligopeptidase B	3.4.21.83	480	1
	γ-D-glutamyl-L-lysine dipeptidyl-peptidase	3.4.14.13	385	0
	glutamate carboxypeptidase	3.4.17.11	383	8
	pyroglutamyl-peptidase I	3.4.19.3	381	3
	proteasome endopeptidase complex	3.4.25.1	356	16
	coagulation factor Xa	3.4.21.6	323	0
	bleomycin hydrolase	3.4.22.40	321	5
	arginyl aminopeptidase	3.4.11.6	315	34
	cathepsin D	3.4.23.5	307	0
	HycI peptidase	3.4.23.51	307	5
	subtilisin	3.4.21.62	299	17
	cyanophycinase	3.4.15.6	260	16
	muramoyltetrapeptide carboxypeptidase	3.4.17.13	170	1
	lysostaphin	3.4.24.75	162	21
	gpr endopeptidase	3.4.24.78	153	0
	γ-D-glutamyl-meso-diaminopimelate peptidase I	3.4.19.11	150	1
	SpoIVB peptidase	3.4.21.116	144	0
	Xaa-Pro dipeptidyl-peptidase	3.4.14.11	81	0
	ficain	3.4.22.3	68	0
	gingipain R	3.4.22.37	68	19
	carboxypeptidase T	3.4.17.18	63	11
	lysyl endopeptidase	3.4.21.50	62	45
	serralysin	3.4.24.40	55	9
	glutamyl aminopeptidase	3.4.11.7	54	0
	clostripain	3.4.22.8	51	5
	pitrilysin	3.4.24.55	46	1
	chymosin	3.4.23.4	38	0
	aminopeptidase S	3.4.11.24	35	4
	plasminogen activator Pla	3.4.23.48	34	0
	gingipain K	3.4.22.47	33	7
	Zinc D-Ala-D-Ala carboxypeptidase	3.4.17.14	30	4
	trypsin	3.4.21.4	30	12
	PepB aminopeptidase	3.4.11.23	28	0
	sedolisin	3.4.21.100	27	15
	glutamyl endopeptidase	3.4.21.19	25	1

Enzyme class	Enzyme name	EC number	Gene counts	
			Whole metagenome	MAGs
Proteases	thermitase	3.4.21.66	24	2
	bacillolysin	3.4.24.28	22	3
	xanthomonalisin	3.4.21.101	21	6
	cathepsin B	3.4.22.1	18	0
	streptopain	3.4.22.10	18	17
	thermolysin	3.4.24.27	15	1
	aqualysin 1	3.4.21.111	13	0
	microbial collagenase	3.4.24.3	12	23
	pseudolysin	3.4.24.26	9	0
	α-Lytic endopeptidase	3.4.21.12	8	2
	lactocepin	3.4.21.96	8	1
	flavastacin	3.4.24.76	8	0
	atrolysin A	3.4.24.1	6	0
	IgA-specific metalloendopeptidase	3.4.24.13	6	5
	C5a peptidase	3.4.21.110	4	1
	omptin	3.4.23.49	4	0
	vibriolysin	3.4.24.25	4	0
	Pro-Pro endopeptidase	3.4.24.89	4	0
	aureolysin	3.4.24.29	3	0
	streptogrisin B	3.4.21.81	2	1
	β-lytic metalloendopeptidase	3.4.24.32	2	0
	snapalysin	3.4.24.77	2	0
	streptogrisin A	3.4.21.80	1	0
	mycolysin	3.4.24.31	1	0
Phosphatases	Alkaline phosphatase	3.1.3.1	23328	22
	Acid phosphatase	3.1.3.2	5588	11
	Phosphoserine phosphatase	3.1.3.3	5584	159
	Phosphatidate phosphatase	3.1.3.4	5442	0
	5'-nucleotidase	3.1.3.5	5653	74
	3'-nucleotidase	3.1.3.6	366	2
	3'(2'),5'-bisphosphate nucleotidase	3.1.3.7	2437	19
	3-phytase	3.1.3.8	2455	1
	Glucose-6-phosphatase	3.1.3.9	686	0
	Glucose-1-phosphatase	3.1.3.10	2362	20
	Fructose-bisphosphatase	3.1.3.11	4159	38
	Trehalose-phosphatase	3.1.3.12	473	15
	Histidinol-phosphatase	3.1.3.15	1540	29
	Protein-serine/threonine phosphatase	3.1.3.16	3308	65
	Phosphoglycolate phosphatase	3.1.3.18	10228	72
	Glycerol-1-phosphatase	3.1.3.21	187	12
	Mannitol-1-phosphatase	3.1.3.22	118	0
	Sugar-phosphatase	3.1.3.23	1137	13
	Inositol-phosphate phosphatase	3.1.3.25	2638	44

Enzyme class	Enzyme name	EC number	Gene counts	
			Whole metagenome	MAGs
Phosphatases	Phosphatidylglycerophosphatase	3.1.3.27	1223	15
	3-deoxy-manno-octulosonate-8-phosphatase	3.1.3.45	2208	21
	Protein-tyrosine-phosphatase	3.1.3.48	3234	41
	Phosphatidylinositol-3-phosphatase	3.1.3.64	48	8
	2-deoxyglucose-6-phosphatase	3.1.3.68	262	5
	Mannosyl-3-phosphoglycerate phosphatase	3.1.3.70	175	1
	2-phosphosulfolactate phosphatase	3.1.3.71	1045	10
	Adenosylcobalamin/alpha-ribazole phosphatase	3.1.3.73	228	7
	Pyridoxal phosphatase	3.1.3.74	152	3
	Acireductone synthase	3.1.3.77	90	0
	Phosphatidylinositol-4,5-bisphosphate 4-phosphatase	3.1.3.78	1	0
	Mannosylfructose-phosphate phosphatase	3.1.3.79	68	3
	D-glycero-beta-D-manno-heptose 1,7-bisphosphate 7-phosphatase	3.1.3.82	1067	19
	D-glycero-alpha-D-manno-heptose-1,7-bisphosphate 7-phosphatase	3.1.3.83	304	5
	Glucosyl-3-phosphoglycerate phosphatase	3.1.3.85	120	13
	2-hydroxy-3-keto-5-methylthiopentenyl-1-phosphate phosphatase	3.1.3.87	384	4
	5'-deoxynucleotidase	3.1.3.89	426	2
	Maltose 6'-phosphate phosphatase	3.1.3.90	33	1
	3',5'-nucleoside bisphosphate phosphatase	3.1.3.97	653	10
	Validoxylamine A 7'-phosphate phosphatase	3.1.3.101	644	0
	5-amino-6-(5-phospho-D-ribitylamino)uracil phosphatase	3.1.3.104	1093	0

Appendix C.. Genome completeness, contamination, count of ubiquitous marker genes per MAGs identified by GTDB-Tk v0.3.2.

Name	Unique Gene Count	Multiple Gene Count	Missing Gene Count	Genome completeness (%)	Contamination (%)
Bin1001	97	6	17	93.1	5.98
Bin1020	107	3	10	93.93	4.25
Bin1036	84	6	30	91.67	7.77
Bin1059	103	8	9	91.38	5.33
Bin1091	101	9	10	95.68	2.25
Bin1111	117	2	1	97.52	0.56
Bin1152	117	1	2	96.24	0.54
Bin1306	109	4	7	92.31	4.92
Bin1359	107	2	11	91.28	0.87
Bin1501	105	5	10	96.43	1.18
Bin154	111	7	2	100	2.2
Bin204	113	5	2	96.77	1.21
Bin205	101	9	10	92.08	5.05
Bin22	117	2	1	100	0.48
Bin231	115	3	2	93.64	2
Bin265	108	3	9	95.02	4.25
Bin328	95	6	19	92.24	5.56
Bin336	107	9	4	97.72	3.94
Bin367	116	3	1	97.13	2.15
Bin396	109	10	1	98.12	6.45
Bin403	109	2	9	97.04	1.61
Bin428	100	13	7	95.77	5.96
Bin481	102	6	12	91.83	0.65
Bin484	115	3	2	93.64	1.45
Bin493	117	3	0	94.47	2
Bin50	106	13	1	97.85	3.46
Bin583	114	1	5	98.9	1.1
Bin609	105	3	12	95.1	4.3
Bin617	108	10	2	91.81	5.18
Bin631	104	8	8	96.63	4.49
Bin684	106	4	10	90.34	2.03
Bin737	114	0	6	93.96	1.46
Bin744	98	8	14	91.39	2.32
Bin768	110	3	7	95.48	6.3
Bin785	99	15	6	91.43	5.12
Bin790	93	7	20	90.34	3.53
Bin803	118	1	1	94.57	0.55
Bin820	112	4	4	94.84	0.22
Bin931	110	2	8	95.76	1.15
Bin967	111	2	7	99.26	0.84

Appendix D. Catabolite repression resistance genes in putative lipolytic MAGs.

MAGs ID	Gene quantity	Sugar	Gene name
583	1	PTS system fructose-specific EIIABC component	fruA_2
	1	PTS system mannose-specific EIIAB component	manX, 2.7.1.191
803	1	PTS system fructose-specific EIIABC component	fruA
403	1	putative sugar kinase YdjH	ydjH, 2.7.1.-
396	1	putative sugar kinase YdjH	ydjH, 2.7.1.-
1152	2	putative sugar kinase YdjH	ydjH, 2.7.1.-
367	1	putative sugar kinase YdjH	ydjH, 2.7.1.-
50	1	putative sugar kinase YdjH	ydjH, 2.7.1.-
684	2	PTS system fructose-specific EIIA component	fruA
1036	2	PTS system mannose-specific EIIAB component	manX, 2.7.1.191
	3	putative sugar kinase YdjH ydjH, 2.7.1.-	ydjH, 2.7.1.-
1091	1	PTS system mannose-specific EIIAB component	manX, 2.7.1.191
1359	1	PTS system mannose-specific EIIAB component	manX, 2.7.1.191
22	1	PTS system mannose-specific EIIAB component	manX, 2.7.1.191
265	1	PTS system mannose-specific EIIAB component	manX, 2.7.1.191
967	1	PTS system mannose-specific EIIAB component	manX, 2.7.1.191
	1	PTS system fructose-specific EIIB'BC component	fruA
154	2	PTS system fructose-specific EIIABC component	fruA
	2	putative sugar kinase YdjH	ydjH, 2.7.1.-
609	1	PTS system fructose-specific EIIABC component	fruA
631	3	PTS system fructose-specific EIIABC component	fruA
820	1	PTS system mannose-specific EIIBCA component	manP
	1	PTS system fructose-specific EIIABC component	fruA
	2	putative sugar kinase YdjH	ydjH, 2.7.1.-
617	1	putative sugar kinase YdjH	ydjH, 2.7.1.-
1501	3	PTS system fructose-specific EIIABC component	fruA
481	0	-	-
484	2	PTS system fructose-specific EIIB'BC component	fruA
	1	putative sugar kinase YdjH	ydjH, 2.7.1.-

MAGs ID	Gene quantity	Sugar	Gene name
231	0	-	-
204	2	putative sugar kinase YdjH	ydjH, 2.7.1.-
1059	1	putative sugar kinase YdjH	ydjH, 2.7.1.-
1001	0	-	-
328	0	-	-
931	1	putative sugar kinase YdjH	ydjH, 2.7.1.-
336	1	putative sugar kinase YdjH	ydjH, 2.7.1.-
1020	1	PTS system fructose-specific EIIABC component	fruA
	1	PTS system mannitol-specific EIICBA component	mtlA
	1	PTS system glucose-specific EIIA component	crr, 2.7.1.199
	1	PTS system glucose-specific EIICBA component	ptsG, 2.7.1.199
744	1	PTS system fructose-specific EIIABC component	fruA
	1	PTS system glucose-specific EIIA component	crr, 2.7.1.199
	1	PTS system glucose-specific EIICBA component	ptsG, 2.7.1.199
	1	PTS system beta-glucoside-specific EIIBCA component	bglF
768	2	PTS system glucose-specific EIICBA component	ptsG 1, 2.7.1.199
	1	PTS system fructose-specific EIIABC component	fruA
1111	1	PTS system beta-glucoside-specific EIIBCA component	bglF
	1	PTS system fructose-specific EIIABC component	fruA
	1	PTS system glucose-specific EIIA component	crr, 2.7.1.199
493	1	PTS system mannitol-specific EIICB component	mtlA
	1	PTS system beta-glucoside-specific EIIBCA component	bglF
	1	PTS system fructose-specific EIIABC component	fruA
785	1	PTS system beta-glucoside-specific EIIBCA component	bglF
205	1	putative sugar kinase YdjH	ydjH, 2.7.1.-
790	0	-	-
1306	2	PTS system fructose-specific EIIABC component	fruA
	1	PTS system mannitol-specific EIICBA component	mtlA
	1	PTS system beta-glucoside-specific EIIBCA component	bglF
	1	putative sugar kinase YdjH	ydjH 1, 2.7.1.-

MAGs ID	Gene quantity	Sugar	Gene name
428	1	PTS system beta-glucoside-specific EIIBCA component	bglF
	1	PTS system mannitol-specific EIICBA component	mtlA
	2	PTS system fructose-specific EIIB'BC component	fruA
	1	putative sugar kinase YdjH	ydjH, 2.7.1.-
737	2	Ascorbate-specific PTS system EIIB component	ulaB, 2.7.1.194
	2	PTS system fructose-specific EIIABC component	fruA
	1	PTS system 2-O-alpha-mannosyl-D-glycerate-specific EIIABC component	mngA

Appendix E. Taxonomic classification of MAGs at different level by GTDB-Tk.

Bin No.	Phylum	Class	Order	Family	Genus	Species
583	Krumholzibacteriota	Krumholzibacteria	SSS58A[1]	SSS58A	Unassigned[2]	Unassigned
803	Bacteroidota	Chlorobia	Chlorobiales	Chlorobiaceae	Chlorobium	Unassigned
403	Bacteroidota	Bacteroidia	Flavobacteriales	PHOS-HE28	PHOS-HE28	Unassigned
396	Bacteroidota	Bacteroidia	Flavobacteriales	PHOS-HE28	PHOS-HE28	Unassigned
1152	Bacteroidota	Bacteroidia	Bacteroidales	WCHB1-69	UBA5266	Unassigned
367	Bacteroidota	Bacteroidia	Bacteroidales	Lentimicrobiaceae	Lentimicrobium	Unassigned
50	Bacteroidota	Bacteroidia	Bacteroidales	4484-276	Unassigned	Unassigned
684	UBA10199	UBA10199	GCA-002796325	1-14-0-20-49-13	Unassigned	Unassigned
1036	Proteobacteria	Alphaproteobacteria	Rhizobiales	Anderseniellaceae	QKVK01	Unassigned
1091	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Paracoccus	Unassigned
1359	Proteobacteria	Gammaproteobacteria	UBA6002	UBA6002	Unassigned	Unassigned
22	Proteobacteria	Gammaproteobacteria	Burkholderiales	Nitrosomonadaceae	Nitrosomonas	Unassigned
265	Proteobacteria	Gammaproteobacteria	Burkholderiales	Rhodocyclaceae	Unassigned	Unassigned
967	Proteobacteria	Gammaproteobacteria	Burkholderiales	Burkholderiaceae	Rhodoferax	Unassigned
154	Hydrogenedentota	Hydrogenedentia	Hydrogenedentiales	Unassigned	Unassigned	Unassigned
609	Omnitrophota	koll11	UBA1560	2-01-FULL-45-10	FEN-1322	Unassigned
631	Spirochaetota	UBA4802	UBA4802	UBA5368	Unassigned	Unassigned
820	RBG-13-61-14	RBG-13-61-14	RBG-13-61-14	Unassigned	Unassigned	Unassigned
617	Myxococcota	Polyangia	HGW-17	Unassigned	Unassigned	Unassigned
1501	Desulfobacterota	Syntrophorhabdia	Unassigned	Unassigned	Unassigned	Unassigned
481	Desulfobacterota	Desulfobacteria	Desulfobacterales	Desulfobacteraceae	Desulfobacter	Desulfobacter postgatei
484	Chloroflexota	Anaerolineae	Anaerolineales	envOPS12	Unassigned	Unassigned
231	Chloroflexota	Anaerolineae	Anaerolineales	envOPS12	Unassigned	Unassigned
204	Firmicutes_A	Clostridia	Christensenellales	CAG-74	DTU024	Unassigned
1059	Firmicutes_A	Clostridia	Oscillospirales	Acutalibacteraceae	UBA1447	Unassigned
1001	Cyanobacteria	Vampirovibrionia	Obscuribacterales	Obscuribacteraceae	Ga0077546	Unassigned
328	Cyanobacteria	Vampirovibrionia	Obscuribacterales	Obscuribacteraceae	Ga0077546	Unassigned

Bin No.	Phylum	Class	Order	Family	Genus	Species
931	Actinobacteriota	Thermoleophilia	Solirubrobacterales	70-9	67-14	Unassigned
336	Actinobacteriota	Acidimicrobiia	Microtrichales	Microtrichaceae	IMCC26207	Unassigned
1020	Actinobacteriota	Actinobacteria	Mycobacteriales	Mycobacteriaceae	Mycolicibacterium	Unassigned
744	Actinobacteriota	Actinobacteria	Mycobacteriales	Mycobacteriaceae	Mycolicibacterium	Unassigned
768	Actinobacteriota	Actinobacteria	Mycobacteriales	Mycobacteriaceae	Mycolicibacterium	Unassigned
1111	Actinobacteriota	Actinobacteria	Mycobacteriales	Mycobacteriaceae	Corynebacterium	Unassigned
493	Actinobacteriota	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Propionicimonas	Unassigned
785	Actinobacteriota	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Propionicimonas	Unassigned
205	Actinobacteriota	Actinobacteria	Nanopelagicales	GCA-2699445	Unassigned	Unassigned
790	Actinobacteriota	Actinobacteria	Nanopelagicales	UBA10799	UBA10799	Unassigned
1306	Actinobacteriota	Actinobacteria	Actinomycetales	Dermatophilaceae	Austwickia	Unassigned
428	Actinobacteriota	Actinobacteria	Actinomycetales	Dermatophilaceae	Austwickia	Unassigned
737	Actinobacteriota	Actinobacteria	Actinomycetales	Microbacteriaceae	Rhodoluna	Unassigned

1. *The Taxa with letters and numbers means that they are still unclassified as they are not cultured yet.*

2. Unassigned : For instance, if it is in the species level it means that the MAG was either placed outside a named genus or its average nucleotide identity (ANI) to the closest intra-genus reference genome with the alignment fraction (AF) of more than/equal to 0.65 was not within the species-specific ANI ranges

Appendix F. Details of taxonomic classification for putative lipolytic MAGs by GTDB-Tk.

User Genome	Classification	FastANI Reference[1]	FastANI Reference	FastANI Taxonomy[3]	FastANI ANI[4]	FastANI Alignment	Closest Placement	Closest Placement Taxonomy[7]	Closest Placement ANI[8]	Closest Placement Alignment	Classification Method[10]	AA Percent[11]	RED Value[12]
Bin 1001	p__Cyanobacteria; c__Vampirovibrionia; o__Obscuribacterales; f__Obscuribacteraceae; g__Ga0077546; s__						GCA_001464165.1	p__Cyanobacteria; c__Vampirovibrionia; o__Obscuribacterales; f__Obscuribacteraceae; g__Ga0077546; s__Ga0077546 sp001464165	84.39	0.67	RED	83.43	0.980
Bin 1020	p__Actinobacteriota; c__Actinobacteria; o__Mycobacteriales; f__Mycobacteriaceae; g__Mycolicibacterium; s__										Topology	89.38	0.963
Bin 1036	p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Anderseniellaceae; g__QKVK01; s__						GCF_003234965.1	p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Anderseniellaceae; g__QKVK01; s__QKVK01 sp003234965	86.92	0.67	RED	72.48	0.982
Bin 1059	p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Acutalibacteraceae; g__UBA1447; s__										RED	89.86	0.909
Bin 1091	p__Proteobacteria; c__Alphaproteobacteria; o__Rhodobacterales; f__Rhodobacteraceae; g__Paracoccus; s__										RED	89.86	0.944
Bin 1111	p__Actinobacteriota; c__Actinobacteria; o__Mycobacteriales; f__Mycobacteriaceae; g__Corynebacterium; s__										Topology	97	0.986

User Genome	Classification	FastANI Reference[1]	FastANI Reference	FastANI Taxonomy[3]	FastANI ANI[4]	FastANI Alignment	Closest Placement	Closest Placement Taxonomy[7]	Closest Placement ANI[8]	Closest Placement Alignment	Classification Method[10]	AA Percent[11]	RED Value[12]
Bin 1152	p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__WCHB1-69; g__UBA5266; s__						GCA_002411545.1	p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__WCHB1-69; g__UBA5266; s__UBA5266 sp002411545	77.31	0.2	RED	95.22	0.923
Bin 1306	p__Actinobacteriota; c__Actinobacteria; o__Actinomycetales; f__Dermatophilaceae; g__Austwickia; s__						GCF_000298175.1	p__Actinobacteriota; c__Actinobacteria; o__Actinomycetales; f__Dermatophilaceae; g__Austwickia; s__Austwickia chelonae	78.24	0.27	RED	90.97	0.898
Bin 1359	p__Proteobacteria; c__Gammaproteobacteria; o__UBA6002; f__UBA6002; g__; s__										RED	90.04	0.745
Bin 1501	p__Desulfobacterota; c__Syntrophorhabdia; o__; f__; g__; s__										RED	89.11	0.434
Bin 154	p__Hydrogenedentota; c__Hydrogenedentia; o__Hydrogenedentiales; f__; g__; s__										RED	96.15	0.715
Bin 204	p__Firmicutes_A; c__Clostridia; o__Christensenellales; f__CAG-74; g__DTU024; s__						GCA_002428405.1	p__Firmicutes_A; c__Clostridia; o__Christensenellales; f__CAG-74; g__DTU024; s__DTU024 sp002428405	77.63	0.17	Topology	95.36	0.946

User Genome	Classification	FastANI Reference[1]	FastANI Reference	FastANI Taxonomy[3]	FastANI ANI[4]	FastANI Alignment	Closest Placement	Closest Placement Taxonomy[7]	Closest Placement ANI[8]	Closest Placement Alignment	Classification Method[10]	AA Percent[11]	RED Value[12]
Bin 205	d__Bacteria; p__Actinobacteriota; c__Actinobacteria; o__Nanopelagicales; f__GCA-2699445; g__; s__										RED	89.03	0.770
Bin 22	p__Proteobacteria; c__Gammaproteobacteria; o__Burkholderiales; f__Nitrosomonadaceae; g__Nitrosomonas; s__						GCF_003201565.1	p__Proteobacteria; c__Gammaproteobacteria; o__Burkholderiales; f__Nitrosomonadaceae; g__Nitrosomonas; s__Nitrosomonas sp003201565	77.59	0.25	Topology	98.25	0.953
Bin 231	p__Chloroflexota; c__Anaerolineae; o__Anaerolineales; f__envOPS12; g__; s__										Topology	95.4	0.903
Bin 265	p__Proteobacteria; c__Gammaproteobacteria; o__Burkholderiales; f__Rhodocyclaceae; g__; s__										RED	91.47	0.916
Bin 328	p__Cyanobacteria; c__Vampirovibrionia; o__Obscuribacterales; f__Obscuribacteraceae; g__Ga0077546; s__						GCA_001464165.1	p__Cyanobacteria; c__Vampirovibrionia; o__Obscuribacterales; f__Obscuribacteraceae; g__Ga0077546; s__Ga0077546 sp001464165	86.62	0.74	RED	81.39	0.985
Bin 336	p__Actinobacteriota; c__Acidimicrobiia; o__Microtrichales; f__Microtrichaceae; g__IMCC26207; s__						GCF_001025035.1	p__Actinobacteriota; c__Acidimicrobiia; o__Microtrichales; f__Microtrichaceae; g__IMCC26207; s__IMCC26207 sp001025035	76.31	0.05	RED	93.71	0.857

User Genome	Classification	FastANI Reference[1]	FastANI Reference	FastANI Taxonomy[3]	FastANI ANI[4]	FastANI Alignment	Closest Placement	Closest Placement Taxonomy[7]	Closest Placement ANI[8]	Closest Placement Alignment	Classification Method[10]	AA Percent[11]	RED Value[12]
Bin 367	p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Lentimicrobiaceae; g__Lentimicrobium; s__						GCA_002426025.1	p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Lentimicrobiaceae; g__Lentimicrobium; s__Lentimicrobium sp002426025	77.09	0.17	Topology	97.48	0.949
Bin 396	p__Bacteroidota; c__Bacteroidia; o__Flavobacteriales; f__PHOS-HE28; g__PHOS-HE28; s__						GCA_002342985.1	p__Bacteroidota; c__Bacteroidia; o__Flavobacteriales; f__PHOS-HE28; g__PHOS-HE28; s__PHOS-HE28 sp002342985	82.57	0.55	Topology	97.22	0.943
Bin 403	p__Bacteroidota; c__Bacteroidia; o__Flavobacteriales; f__PHOS-HE28; g__PHOS-HE28; s__						GCA_002396605.1	p__Bacteroidota; c__Bacteroidia; o__Flavobacteriales; f__PHOS-HE28; g__PHOS-HE28; s__PHOS-HE28 sp002396605	79.16	0.47	Topology	90.44	0.949
Bin 428	p__Actinobacteriota; c__Actinobacteria; o__Actinomycetales; f__Dermatophilaceae; g__Austwickia; s__						GCF_000298175.1	p__Actinobacteriota; c__Actinobacteria; o__Actinomycetales; f__Dermatophilaceae; g__Austwickia; s__Austwickia chelonae	78.86	0.28	RED	91.15	0.896
Bin 481	p__Desulfobacterota; c__Desulfobacteria; o__Desulfobacterales; f__Desulfobacteraceae; g__Desulfobacter; s__Desulfobacter postgatei	GCF_000233695.2	95	p__Desulfobacterota; c__Desulfobacteria; o__Desulfobacterales; f__Desulfobacteraceae; g__Desulfobacter; s__Desulfobacter postgatei	96.18	0.81	GCF_000233695.2	p__Desulfobacterota; c__Desulfobacteria; o__Desulfobacterales; f__Desulfobacteraceae; g__Desulfobacter; s__Desulfobacter postgatei	96.18	0.81	Topology and ANI	87.52	
Bin 484	p__Chloroflexota; c__Anaerolineae; o__Anaerolineales; f__envOPS12; g__; s__										Topology	95.26	0.903

User Genome	Classification	FastANI Reference[1]	FastANI Reference	FastANI Taxonomy[3]	FastANI ANI[4]	FastANI Alignment	Closest Placement	Closest Placement Taxonomy[7]	Closest Placement ANI[8]	Closest Placement Alignment	Classification Method[10]	AA Percent[11]	RED Value[12]
Bin 493	p__Actinobacteriota; c__Actinobacteria; o__Propionibacteriales; f__Propionibacteriaceae; g__Propionicimonas; s__						GCA_002841335.1	p__Actinobacteriota; c__Actinobacteria; o__Propionibacteriales; f__Propionibacteriaceae; g__Propionicimonas; s__Propionicimonas sp002841335	86.39	0.84	Topology	97.02	0.987
Bin 50	p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__4484-276; g__; s__										RED	96.15	0.800
Bin 583	p__Krumholzibacteriota; c__Krumholzibacteria; o__SSS58A; f__SSS58A; g__; s__										RED	94.05	0.856
Bin 609	p__Omnitrophota; c__koll11; o__UBA1560; f__2-01-FULL-45-10; g__FEN-1322; s__						GCA_003140915.1	p__Omnitrophota; c__koll11; o__UBA1560; f__2-01-FULL-45-10; g__FEN-1322; s__FEN-1322 sp003140915	76.57	0.21	RED	87.16	0.910
Bin 617	p__Myxococcota; c__Polyangia; o__HGW-17; f__; g__; s__										RED	95.85	0.588
Bin 631	p__Spirochaetota; c__UBA4802; o__UBA4802; f__UBA5368; g__; s__						GCA_002407865.1	p__Spirochaetota; c__UBA4802; o__UBA4802; f__UBA5368; g__UBA5368; s__UBA5368 sp002407865	76.69	0.11	RED	90.38	0.811

User Genome	Classification	FastANI Reference[1]	FastANI Reference	FastANI Taxonomy[3]	FastANI ANI[4]	FastANI Alignment	Closest Placement	Closest Placement Taxonomy[7]	Closest Placement ANI[8]	Closest Placement Alignment	Classification Method[10]	AA Percent[11]	RED Value[12]
Bin 684	p__UBA10199; c__UBA10199; o__GCA-002796325; f__1-14-0-20-49-13; g__; s__										RED	87.84	0.783
Bin 737	p__Actinobacteriota; c__Actinobacteria; o__Actinomycetales; f__Microbacteriaceae; g__Rhodoluna; s__						GCF_000699505.1	p__Actinobacteriota; c__Actinobacteria; o__Actinomycetales; f__Microbacteriaceae; g__Rhodoluna; s__Rhodoluna lacicola	79.25	0.41	Topology	92.86	0.952
Bin 744	p__Actinobacteriota; c__Actinobacteria; o__Mycobacteriales; f__Mycobacteriaceae; g__Mycolicibacterium; s__						GCA_001510415.1	p__Actinobacteriota; c__Actinobacteria; o__Mycobacteriales; f__Mycobacteriaceae; g__Mycolicibacterium; s__Mycolicibacterium sp001510415	80.57	0.59	Topology	85.91	0.967
Bin 768	p__Actinobacteriota; c__Actinobacteria; o__Mycobacteriales; f__Mycobacteriaceae; g__Mycolicibacterium; s__						GCA_001510415.1	p__Actinobacteriota; c__Actinobacteria; o__Mycobacteriales; f__Mycobacteriaceae; g__Mycolicibacterium; s__Mycolicibacterium sp001510415	77.91	0.33	Topology	91.31	0.956
Bin 785	p__Actinobacteriota; c__Actinobacteria; o__Propionibacteriales; f__Propionibacteriaceae; g__Propionicimonas; s__										Topology	91.69	0.983
Bin 790	p__Actinobacteriota; c__Actinobacteria; o__Nanopelagicales; f__UBA10799; g__UBA10799; s__						GCA_003452655.1	p__Actinobacteriota; c__Actinobacteria; o__Nanopelagicales; f__UBA10799; g__UBA10799; s__UBA10799 sp003452655	78.04	0.1	RED	80.97	0.861

User Genome	Classification	FastANI Reference[1]	FastANI Reference	FastANI Taxonomy[3]	FastANI ANI[4]	FastANI Alignment	Closest Placement	Closest Placement Taxonomy[7]	Closest Placement ANI[8]	Closest Placement Alignment	Classification Method[10]	AA Percent[11]	RED Value[12]
Bin 803	p__Bacteroidota; c__Chlorobia; o__Chlorobiales; f__Chlorobiaceae; g__Chlorobium; s__										Topology	97.02	0.944
Bin 820	p__RBG-13-61-14; c__RBG-13-61-14; o__RBG-13-61-14; f__; g__; s__						GCA_001797815.1	p__RBG-13-61-14; c__RBG-13-61-14; o__RBG-13-61-14; f__RBG-13-61-14; g__RBG-13-61-14; s__RBG-13-61-14 sp001797815	76.29	0.16	RED	95.58	0.639
Bin 931	p__Actinobacteriota; c__Thermoleophilia; o__Solirubrobacterales; f__70-9; g__67-14; s__						GCA_001897355.1	p__Actinobacteriota; c__Thermoleophilia; o__Solirubrobacterales; f__70-9; g__67-14; s__67-14 sp001897355	82.74	0.7	RED	89.98	0.968
Bin 967	p__Proteobacteria; c__Gammaproteobacteria; o__Burkholderiales; f__Burkholderiaceae; g__Rhodoferax; s__										Topology	93.08	0.981

1. Indicates the accession number of the closest reference genome as determine by ANI. This genome is used along with the placement of the genome in the reference tree to determine the species assignment on the genome. ANI values are only calculated when a query genome is placed within a defined genus and are calculated for all reference genomes in the genus 2. indicates the species-specific ANI circumscription radius of the reference genomes used to determine if a query genome should be classified to the same species as the reference 3. Indicates the GTDB taxonomy of the closest reference genome 4. Indicates the ANI between the query and the closest reference genome 5. Indicates the AF between the query and the closest reference genome 6. Indicates the accession number of the reference genome when a genome is placed on a terminal branch. This genome is used along with the ANI information to determine the species assignment on the genome 7. Indicates the GTDB taxonomy of the reference genome 8. Indicates the ANI between the query and the reference genome 9. Indicates the AF between the query and the reference genome 10. Indicates the rule used to classify the genome. This field will be one of: i) ANI/Placement, indicating a species assignment was made based on both the calculate ANI and placement of the genome in the reference tree; ii) taxonomic classification fully defined by topology, indicating that the classification could be determine based solely on the genome's position in the reference tree; or iii) taxonomic novelty determined using RED, indicating that the relative evolutionary divergence (RED) and placement of the genome in the reference tree were used to determine the classification 11. Indicates the percentage of the MSA spanned by the genome (i.e. percentage of columns with an amino acid) 12. Indicates, when required, the relative evolutionary divergence (RED) for a query genome. RED is not calculated when a query genome can be classified based on ANI.

Appendix G. Grouping the putative lipolytic MAGs based on the role of the lipase on the genome.

MAG ID	Lowest classified level	Phyla	Gram stain	fadL	Denitrification/PHA genes
Bin1001.gff	Family- Obscuribacteraceae	Cyanobacteria	Gram negative	None	Only Denitrification
Bin1020.gff	Genus-Mycolicibacterium	Actinobacteriota	Gram positive	None	Only PHA
Bin1036.gff	Family-Anderseniellaceae	Proteobacteria	Gram negative	None	Both
Bin1059.gff	Family-Acutalibacteraceae	Firmicutes_A	Gram positive	None	None
Bin1091.gff	Genus-Paracoccus	Proteobacteria	Gram negative	None	Both
Bin1111.gff	Genus-Corynebacterium	Actinobacteriota	Gram positive	None	None
Bin1152.gff	Order-Bacteroidales	Bacteroidota	Gram negative	None	None
Bin1306.gff	Genus-Austwickia	Actinobacteriota	Gram positive	None	Only PHA
Bin1359.gff	Class-Gammaproteobacteria	Proteobacteria	Gram negative	None	None
Bin1501.gff	Class-Syntrophorhabdia	Desulfobacterota	Gram negative	Yes	None
Bin154.gff	Order-Hydrogenedentiales	Hydrogenedentota	Not known	None	Only Denitrification
Bin204.gff	Order-Christensenellales	Firmicutes_A	Gram positive	None	Only Denitrification
Bin205.gff	Order-Nanopelagicales	Actinobacteriota	Gram positive	None	Only PHA
Bin22.gff	Genus-Nitrosomonas	Proteobacteria	Gram negative	None	Only Denitrification
Bin231.gff	Order-Anaerolineales	Chloroflexota	Mostly gram negative	None	None
Bin265.gff	Family-Rhodocyclaceae	Proteobacteria	Gram negative	None	Only PHA
Bin328.gff	Family-Obscuribacteraceae	Cyanobacteria	Gram negative	None	None
Bin336.gff	Family-Microtrichaceae	Actinobacteriota	Gram positive	None	Only PHA
Bin367.gff	Genus-Lentimicrobium	Bacteroidota	Gram negative	None	Only Denitrification
Bin396.gff	Order-Flavobacteriales	Bacteroidota	Gram negative	None	Only Denitrification
Bin403.gff	Order-Flavobacteriales	Bacteroidota	Gram negative	None	Only Denitrification
Bin428.gff	Genus-Austwickia	Actinobacteriota	Gram positive	None	Both
Bin481.gff	Species-Desulfobacter postgatei	Desulfobacterota	Gram negative	None	None
Bin484.gff	Order-Anaerolineales	Chloroflexota	Mostly gram negative	None	Only Denitrification
Bin493.gff	Genus-Propionicimonas	Actinobacteriota	Gram positive	None	Only Denitrification
Bin50.gff	Order-Bacteroidales	Bacteroidota	Gram negative	None	Both
Bin583.gff	Class-Krumholzibacteria	Krumholzibacteriota	Gram negative	None	None

MAG ID	Lowest classified level	Phyla	Gram stain	fadL	Denitrification/PHA genes
Bin609.gff	Phylum-Omnitrophota	Omnitrophota	Not known	None	None
Bin617.gff	Class-Polyangia	Myxococcota	Gram negative	None	Only PHA
Bin631.gff	Phylum-Spirochaetota	Spirochaetota	Weak Gram negative in some	None	Only Denitrification
Bin684.gff	Unassigned	Unassigned	Not known	None	None
Bin737.gff	Genus-Rhodoluna	Actinobacteriota	Gram positive	None	None
Bin744.gff	Genus-Mycolicibacterium	Actinobacteriota	Gram positive	None	Only PHA
Bin768.gff	Genus-Mycolicibacterium	Actinobacteriota	Gram positive	None	Both
Bin785.gff	Genus-Propionicimonas	Actinobacteriota	Gram positive	None	Only Denitrification
Bin790.gff	Order-Nanopelagicales	Actinobacteriota	Gram positive	None	None
Bin803.gff	Genus-Chlorobium	Bacteroidota	Gram negative	None	None
Bin820.gff	Unassigned	Unassigned	Not known	None	None
Bin931.gff	Order-Solirubrobacterales	Actinobacteriota	Gram positive	None	None
Bin967.gff	Genus-Rhodoferax	Proteobacteria	Gram negative	Yes	Both

Appendix H. MAGs linked to the taxa, reactor conditions and lipases, the status of the conditions in each MAG are based on the ANOVA in Appendix H.

MAGs ID	Cat.[1]	Phylum	Lowest classified level		Treatment	Phase	Temp. (°C)	Lip. Quant[2]	Class	Length (aa)
583	1	Krumholzibacteriota	Class	Krumholzibacteria	Ster[3] ~ Nster[4]	Liq[5] ~ Bio[6]	4 > 15	1	Lipase 2	274
803	3	Bacteroidota	Genus	Chlorobium	Nster >> Ster	Liq > Bio	4 >> 15	1	Lipase 1	287
403	1	Bacteroidota	Order	Flavobacteriales	Ster > Nster	Liq ~ Bio	4 ~ 15	1	Lipase 3	362
396	6	Bacteroidota	Order	Flavobacteriales	Ster > Nster	Liq ~ Bio	4 ~ 15	1	Lipase 3	363
1152	1	Bacteroidota	Order	Bacteroidales	Ster ~ Nster	Bio > Liq	4 > 15	1	Lipase 2	321
367	2	Bacteroidota	Genus	Lentimicrobium	Ster >> Nster	Liq ~ Bio	4 ~ 15	1	Lipase 2	305
50	2	Bacteroidota	Order	Bacteroidales	Ster ~ Nster	Liq ~ Bio	4 > 15	1	Lipase 1	265
684	5	Unassigned	-	-	Ster ~ Nster	Liq ~ Bio	4 ~ 15	1	Lipase	247
1036	4	Proteobacteria	Family	Anderseniellaceae	Ster ~ Nster	Liq > Bio	4 ~ 15	1	Putative	391
1091	2	Proteobacteria	Genus	Paracoccus	Ster ~ Nster	Liq ~ Bio	4 ~ 15	1	Lipase 3	294
1359	3	Proteobacteria	Class	Gammaproteobacteria	Ster ~ Nster	Liq ~ Bio	4 ~ 15	1	Est A	215
22	1	Proteobacteria	Genus	Nitrosomonas	Ster >> Nster	Liq > Bio	4 ~ 15	1	Lipase 3	320
265	2	Proteobacteria	Family	Rhodocyclaceae	Nster > Ster	Bio > Liq	4 ~ 15	1	Lipase 1	325
967	1	Proteobacteria	Genus	Rhodoferax	Ster > Nster	Liq ~ Bio	4 > 15	1	Lipase	306
154	2	Hydrogenedentota	Order	Hydrogenedentiales	Ster ~ Nster	Liq > Bio	15 >> 4	1	Lipase 2	306
609	2	Omnitrophota	Phylum	Omnitrophota	Nster > Ster	Bio >> Liq	15 >> 4	1	Lipase 1	220
631	2	Spirochaetota	Phylum	Spirochaetota	Ster > Nster	Liq ~ Bio	15 > 4	1	Lactonizing	297
820	3	Unassigned	-	-	Ster ~ Nster	Liq ~ Bio	15 > 4	1	Lipase 2	308
617	4	Myxococcota	Class	Polyangia	Ster ~ Nster	Bio > Liq	15 > 4	4	Lipase	423
										442
									Est A	289
									Lipase 2	419
1501	1	Desulfobacterota	Class	Syntrophorhabdia	Ster ~ Nster	Liq ~ Bio	4 ~ 15	1	Lactonizing	253
481		Desulfobacterota	Species	Desulfobacter postgatei	Ster > Nster	Bio >> Liq	15 >> 4	1	Lipase 3	312
484	3	Chloroflexota	Order	Anaerolineales	Ster > Nster	Bio > Liq	15 > 4	3	Lipase 1	274
										246
									Est A	618
231	3	Chloroflexota	Order	Anaerolineales	Ster >> Nster	Bio > Liq	15 >> 4	1	Lipase 3	246

MAGs ID	Cat.[1]	Phylum	Lowest classified level		Treatment	Phase	Temp. (°C)	Lip. Quant[2]	Class	Length (aa)
204	1	Firmicutes_A	Order	Christensenellales	Ster ~ Nster	Liq > Bio	4 > 15	1	Lipase 1	339
1059	4	Firmicutes_A	Family	Acutalibacteraceae	Nster > Ster	Liq ~ Bio	4 > 15	3	Lipase	211
										561
										404
1001	4	Cyanobacteria	Family	Obscuribacteraceae	Ster ~ Nster	Bio >> Liq	15 > 4	3	Lipase 2	293
										313
										333
328	4	Cyanobacteria	Family	Obscuribacteraceae	Nster >> Ster	Bio >> Liq	15 >> 4	1	Lipase 2	317
931	1	Actinobacteriota	Order	Solirubrobacterales	Ster ~ Nster	Liq ~ Bio	4 ~ 15	2	Lipase 3	264
									Putative	414
336	2	Actinobacteriota	Family	Microtrichaceae	Ster ~ Nster	Liq ~ Bio	4 ~ 15	2	Lipase	319
									Putative	180
1020	5	Actinobacteriota	Genus	Mycolicibacterium	Ster ~ Nster	Liq ~ Bio	4 ~ 15	6	Lipase 2	397
										362
									Triacylglycerol	562
										570
									Putative	445
										412
744	5	Actinobacteriota	Genus	Mycolicibacterium	Ster ~ Nster	Liq ~ Bio	4 ~ 15	2	Lipase	348
									Putative	446
768	6	Actinobacteriota	Genus	Mycolicibacterium	Ster ~ Nster	Liq ~ Bio	4 ~ 15	7	Lipase	352
										353
									Lipase 2	291
										253
									Putative	445
									Triacylglycerol	477
										537

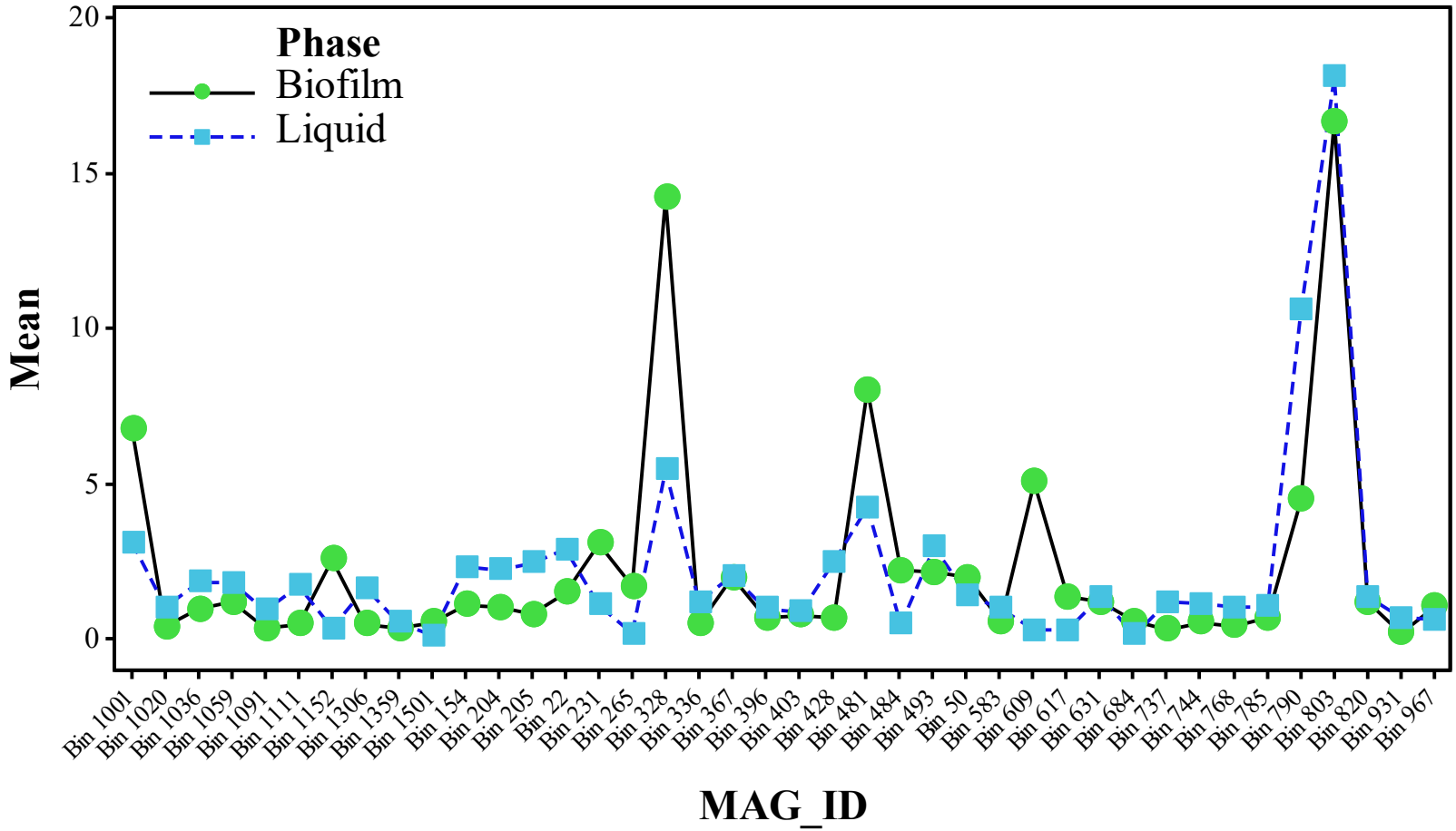
MAGs ID	Cat.[1]	Phylum	Lowest classified level		Treatment	Phase	Temp. (°C)	Lip. Quant[2]	Class	Length (aa)
1111	1	Actinobacteriota	Genus	Corynebacterium	Ster ~ Nster	Liq > Bio	4 ~15	7	Lipase	339
									Lipase 2	250
										273
										298
									Putative	450
										456
										471
493	3	Actinobacteriota	Genus	Propionicimonas	Ster ~ Nster	Liq > Bio	4 ~ 15	3	Lipase 2	258
									Lipase 3	720
									Putative	565
785	4	Actinobacteriota	Genus	Propionicimonas	Ster > Nster	Liq ~ Bio	4 ~ 15	1	Putative	569
205	4	Actinobacteriota	Order	Nanopelagicales	Ster > Nster	Liq > Bio	4 ~ 15	1	Lipase 1	293
790	5	Actinobacteriota	Order	Nanopelagicales	Nster > Ster	Liq >> Bio	4 ~ 15	4	Lipase 1	361
									Lipase 3	308
									Putative	420
										443
1306	5	Actinobacteriota	Genus	Austwickia	Ster > Nster	Liq > Bio	4 ~ 15	3	Lipase 1	358
									Triacylglycerol	306
									Putative	368
428	6	Actinobacteriota	Genus	Austwickia	Ster >> Nster	Liq > Bio	4 ~ 15	3	Lipase	819
									Lipase 1	339
									Putative	369
737	3	Actinobacteriota	Genus	Rhodoluna	Nster > Ster	Liq > Bio	4 ~ 15	1	Lipase 3	311

1- Category 2- Lipase quantity 3- Sterile 4- Non-sterile 5- Liquid 6- Biofilm

Appendix I. P-values (two-way ANOVA): Abundance of reads per putative lipolytic MAGs that mapped to different reactor conditions including phase, treatment, and temperature (α=0.05); highlighted cells in yellow had P-value ≤ 0.05. P-value zero means that the value is very close to zero and hence is significant.

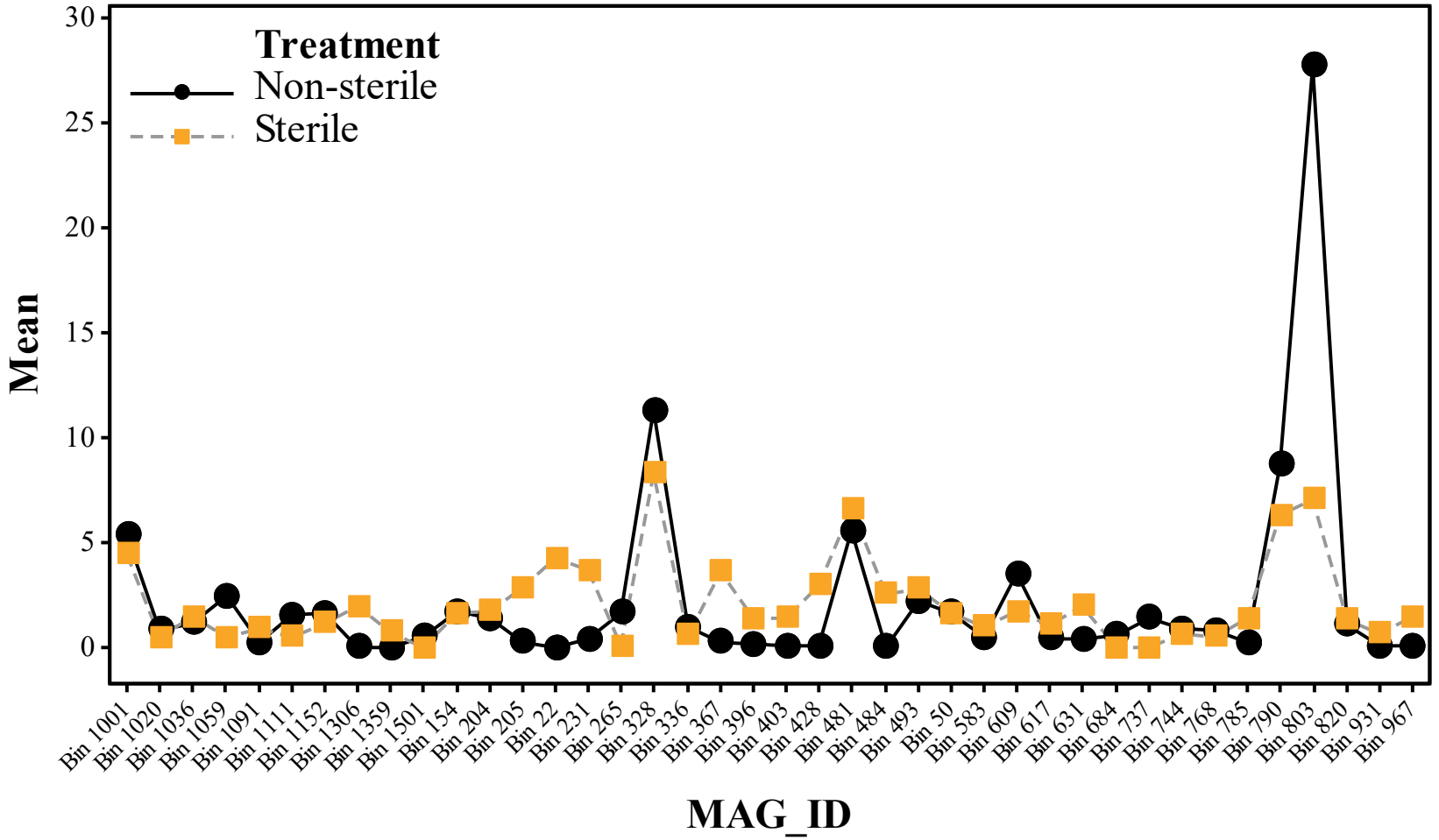
MAG_ID	P-value			MAG_ID	P-value		
	Phase	Treatment	Temperature		Phase	Treatment	Temperature
Bin 1001	0.037	0.604	0.134	Bin 403	0.881	0.329	0.925
Bin 1020	0.653	0.865	0.886	Bin 428	0.255	0.045	0.597
Bin 1036	0.548	0.797	0.768	Bin 481	0.032	0.451	0.084
Bin 1059	0.654	0.224	0.142	Bin 484	0.343	0.082	0.16
Bin 1091	0.672	0.587	0.842	Bin 493	0.563	0.627	0.734
Bin 1111	0.409	0.543	0.972	Bin 50	0.807	0.968	0.211
Bin 1152	0.216	0.807	0.101	Bin 583	0.742	0.674	0.364
Bin 1306	0.449	0.184	0.674	Bin 609	0.006	0.267	0.001
Bin 1359	0.846	0.559	0.957	Bin 617	0.575	0.625	0.316
Bin 1501	0.861	0.78	0.816	Bin 631	0.852	0.26	0.179
Bin 154	0.418	0.996	0.03	Bin 684	0.87	0.753	0.664
Bin 204	0.421	0.746	0.07	Bin 737	0.547	0.387	0.899
Bin 205	0.286	0.08	0.987	Bin 744	0.665	0.926	0.763
Bin 22	0.375	0.004	0.595	Bin 768	0.671	0.912	0.864
Bin 231	0.277	0.025	0.021	Bin 785	0.762	0.406	0.929
Bin 265	0.412	0.327	0.894	Bin 790	0.000	0.129	0.798
Bin 328	0	0.059	0.045	Bin 803	0.352	0.000	0.000
Bin 336	0.621	0.86	0.929	Bin 820	0.85	0.824	0.101
Bin 367	0.902	0.022	0.641	Bin 931	0.72	0.649	0.897
Bin 396	0.787	0.391	0.868				

Appendix J. Two-way ANOVA interaction plot (Minitab 18) for the abundance of reads per MAGs mapped to different phases (Biofilm and bulk Liquid) in the reactors.



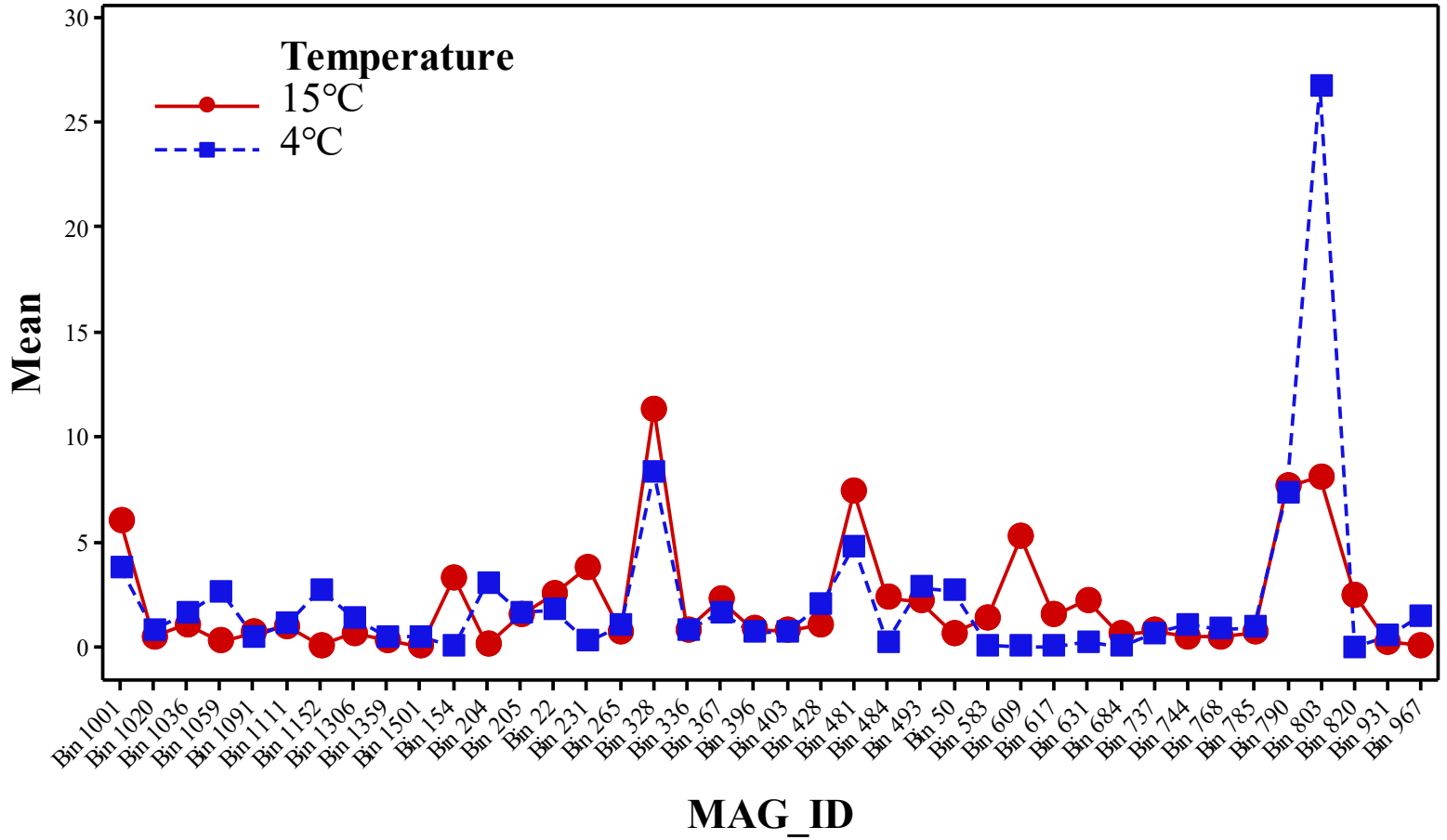
Two-way ANOVA interaction plot (Minitab 18) for the abundance of reads per MAGs mapped to different phases (Biofilm and bulk Liquid) in the reactors.

Appendix K. Two-way ANOVA interaction plot (Minitab 18) for the abundance of reads per MAGs mapped to different treatment (Sterile and Non-sterile) in the reactors.



Two-way ANOVA interaction plot (Minitab 18) for the abundance of reads per MAGs mapped to different treatment (Sterile and Non-sterile) in the reactors.

Appendix L. Two-way ANOVA interaction plot (Minitab 18) for the abundance of reads per MAGs mapped to different temperature (4°C and 15°C) in the reactors.

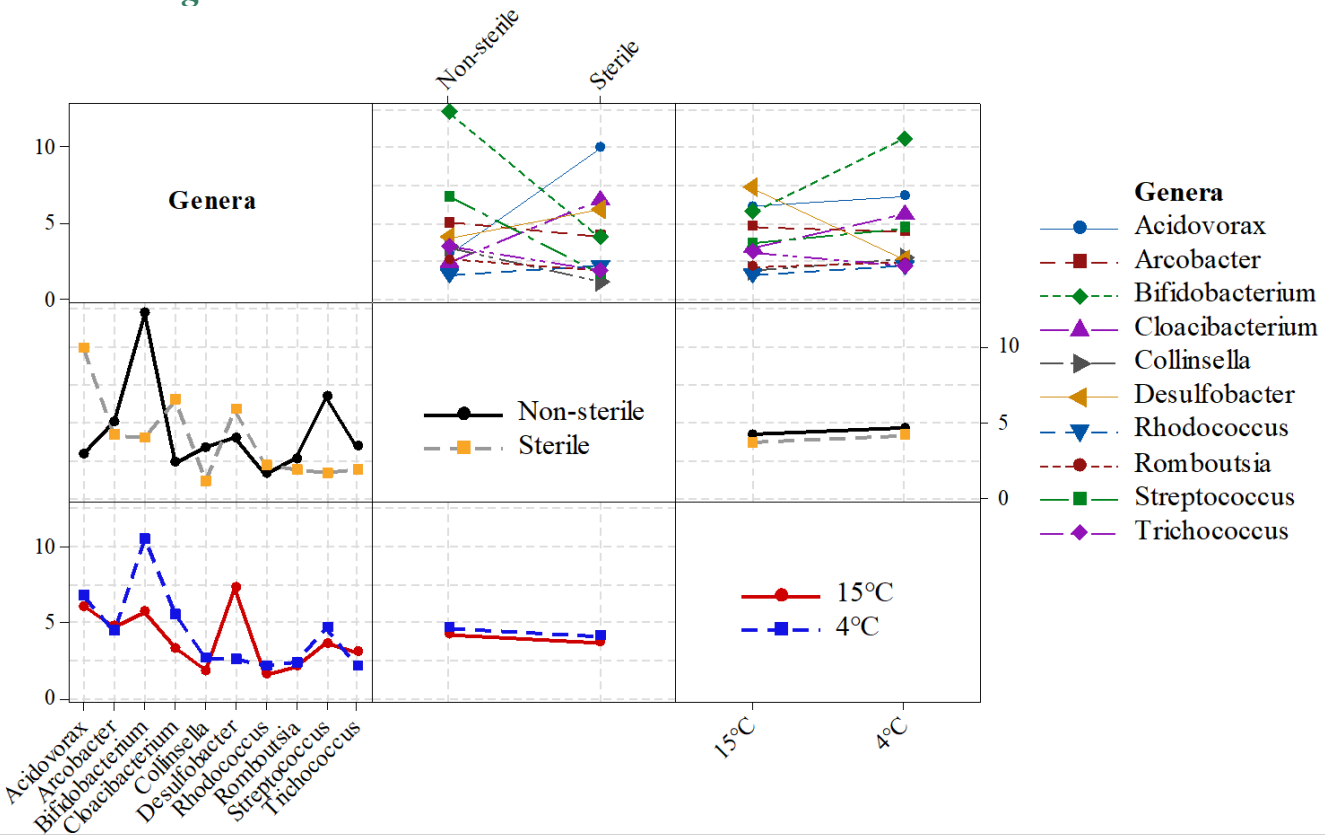


Two-way ANOVA interaction plot (Minitab 18) for the abundance of reads per MAGs mapped to different temperature (4°C and 15°C) in the reactors.

Appendix M. Significant putative lipolytic MAGs (MAGs with the highest mapped reads, but not statistically) at 4℃ and 15℃ and the average length of their lipases.

	MAG_ID	Length (aa)
MAGs more significant at 4 ℃	Bin 1059	392
	Bin 204	339
	Bin 1152	321
	Bin 803	287
	Bin 967	306
	Bin 50	265
MAGs more significant at 15 ℃	Bin 484	379
	Bin 328	317
	Bin 481	312
	Bin 820	308
	Bin 154	306
	Bin 631	297
	Bin 1001	293
	Bin 231	246
	Bin 609	220
	Bin 617	393

Appendix N. Interaction plot (ANOVA, Minitab 18): Effect of temperature (4°C and 15°C) and treatment (sterile and non-sterile) on relative abundance of microbes at genus level.



The value on Y-axis is a relative abundance of genera identified by GOTTCHA2 at different reactor temperatures and treatments.

Interaction plot (ANOVA, Minitab 18): Effect of temperature (4°C and 15°C shown as blue and red lines) and treatment (sterile and non-sterile, shown as yellow and black lines) on relative abundance of microbes at genus level.

Appendix O. Protein extraction and its downstream processes

VSS measurement

Microfiber Whatman filter papers were first dried in an oven at 105 ℃ for 15 min then in a furnace at 550 ℃ for 5 min. they were later cooled down in a desiccator, labelled by a soft pencil and weighed to constant value. 10 ml of bulk liquid and 1ml of scarped biofilm (the volume was estimated by a 1 ml microcentrifuge tube prior to being transferred to a Whatman filter paper) from AnMBRs were filtered and first dried for 1 hr at 105 ℃ then at 550 ℃ for 5 min. After the ignition, filter papers were cooled down in a desiccator and weighed to the constant value. The initial weight of the empty filter papers was subtracted from the weight obtained after the ignition and reported as g/l.

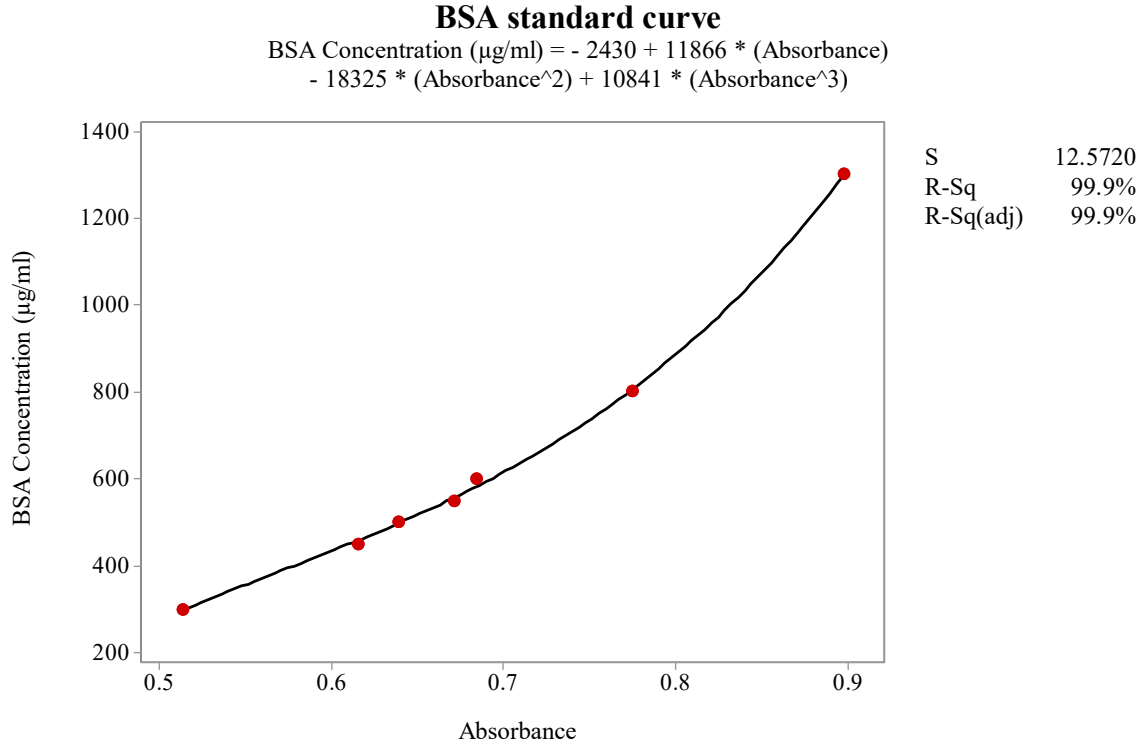
Protein extraction

From each reactor, 10 ml of bulk liquid and 1 ml of biofilm were collected and transferred to individual 50 ml conical centrifuge tubes. 9 ml autoclaved distilled water was added to biofilm-containing tubes to retain the same volume. 5 gr cation exchange resin (DOWEX, 50X8, 20-50 mesh, Na^+ form, strong acidic, Sigma Aldrich) pre-washed for 1 h in sample buffer (2 mM Na_3PO_4, 4 mM NaH_2PO_4, 9 mM NaCl and 1 mM KCl at pH=7) along with 10 µl Triton X-100 (final concentration of 0.1% v/v) was added to each tube. The quantity of resin is usually determined based on the gr VSS of samples. Gessesse *et al.* (2003) and Frølund *et al.* (1996) recommended 70 gr resin/ gr VSS for wastewater samples. However, the VSS for biofilm samples was high and it was not possible to add resin on such basis and work at 50 ml final volume (the maximum accessible capacity for a high-speed centrifuge was for 50 ml tubes). Therefore, 5 gr was the maximum quantity that could be added to all samples. Samples were shaken for 1.5 h at 400 rpm and 4℃ and then centrifuged twice at the same temperature (20 min at 15,000g and 10 min at 10,000g). The supernatant was collected for protein quantification and precipitation.

Protein quantification

Pierce™ Modified Lowry Protein Assay Kit was used for measuring the concentration of proteins in the supernatant and plotting the standard curve of Bovine serum albumin (BSA). The BSA with concentration of 2 mg/ml was diluted into various ranges of 1, 5, 25, 125, 250, 500, 750, 1000 and 1500 µg/ml according to the kit instructions. 0.2 ml of each dilution was mixed with 1 ml of modified Lowry reagent, vortexed and incubated for 10 min at room

temperature. Finally, 0.1 ml of phenol reagent (already diluted with distilled water to yield 1 N solution) was added to each sample, vortexed and incubated at room temperature for another 30 min. The absorbances were read at 750 nm and plotted against concentrations of diluted samples for further calculations.



BSA standard curve, cubic regression model, Minitab 18.

Protein precipitation

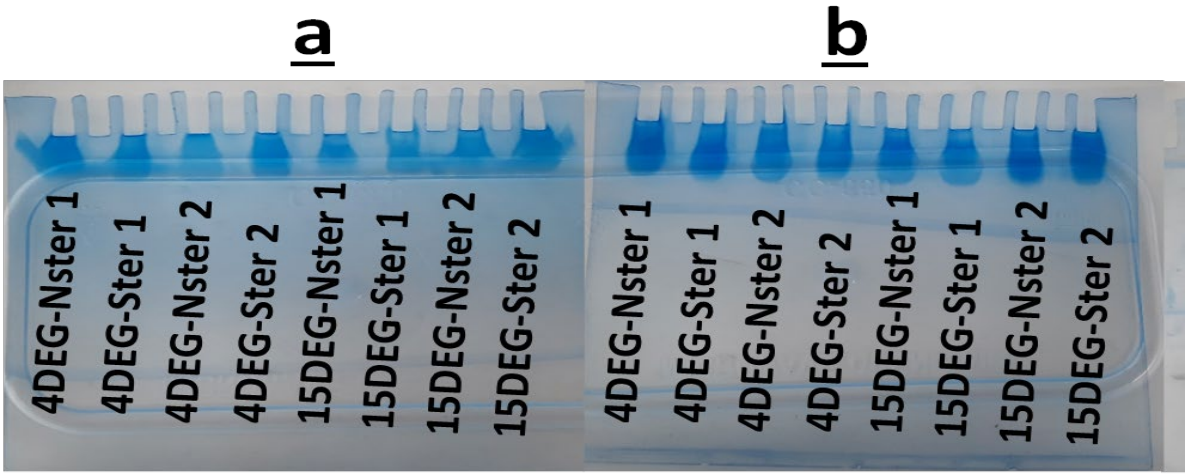
1 part of the supernatant was mixed with 4 parts of ice-cold methanol and was vortexed. A mixture of 1 part of ice-cold chloroform and 3 parts of cold distilled water were then added respectively and vortexed too. The mixture was then centrifuged for 1 min at 15500g and 4 ºC to form three phases (proteins form a circular flake in the interface of water and chloroform). The top aqueous layer containing salts and hydrophilic contaminants was carefully removed by pipette. 4 part of methanol was added again and after being vortexed, the mixture was centrifuged for 5 min at 15500g and 4 ºC. After removing the supernatant, the pellets were air-dried and stored at -80 ℃ for further analysis.

1D SDS-PAGE

100 µl of BME and 900 µl of Laemmli buffer were mixed and 10 µl of the mixture was added to each tube containing protein pellets (defrosted in room temperature). The tubes were sonicated for 20 min at cool temperature, then heated at 60 ℃ for 5 min and centrifuged for 10

min at 4 °C. 10 µl of supernatant was injected into wells (4–15% Mini-PROTEAN® TGX™ Precast Protein Gels, 15-well, 15 µl) and run for 5 min at 120 V (Bio-Rad Mini-PROTEAN®). The gel was removed from the tank, was immersed in distilled water, microwaved for 1 min, and shaken at 360 rpm for 1 min (PMS-1000i Microplate Shaker, Grant Instruments™). The water was removed, and the washing/microwaving/shaking procedure was repeated for three times. After removing the water, the gel was stained by 60 ml Bio-Safe Coomassie Brilliant Blue G-250 (microwaved for 1 min and shaken for 5 min at 360 rpm) and destained overnight in distilled water at 360 rpm and room temperature. Destained gel was stored at 4 °C in 20 mM NaCl solution before in-gel digestion.

Gel visualisation



Visualisation of SDS-PAGE a) Biofilm phase of AnMBRs b) Liquid phase of AnMBRs. 4DEG and 15 DEG are referring to reactors working at 4 °C and 15 °C, Nster and Ster are referring to Sterile and Non-sterile conditions, and 1 and 2 are referring to replicates of each reactor.

Appendix P. In-gel digestion and mass spectrometry

In-gel digestion

Each 1D SDS-PAGE band was excised with a clean scalpel, diced into 1x1x1 mm cubes, and transferred to a clean microcentrifuge tube. Gel pieces were destained by mixture of 50mM ammonium bicarbonate and acetonitrile (50%). The destained buffer was removed and exchanged until the gel pieces were clear. As a digest control, a molecular weight marker band was also excised. Proteins were reduced with 10 mM dithiothreitol for 30 min at 60˚C to break disulphide bridges. This was followed by alkylation with 50 mM iodoacetamide for 30 min at room temperature in the dark to prevent disulphide reformation. Gel pieces were washed in 50mM ammonium bicarbonate and then dehydrated with 3 washes of 100 µL of acetonitrile. Residual moisture was removed from gel pieces in a vacuum drier. Proteins were digested by the addition of trypsin added at a ratio of 30:1 (protein: trypsin), buffered with 50 mM ammonium bicarbonate and incubated for 16 hours at 37 ˚C. The digest was stopped by the addition of 10% Trifluoroacetic acid (TFA) to a final concentration of 0.5%, shaken for 30 mins, 750 rpm. The liquid containing hydrophilic peptides was transferred to a fresh microcentrifuge tube. 80% acetonitrile with 2% TFA was then added to the gel pieces and shaken for 30 min at 750 rpm. This dehydrates the gel pieces and removes hydrophobic peptides from the gel. The solution containing hydrophobic peptides was pooled with the hydrophilic peptide mix. The peptide solution was dried in a centrifugal evaporator, peptides were dissolved in 3% acetonitrile, and 0.1% TFA. The resulting peptide solutions were desalted using home packed C18 stage tips (Rappsilber *et al.*, 2007). The sample was dissolved in 50 µL of 3% acetonitrile, 0.1% TFA giving the final concentration of ~1µg/µL.

Nano LC-MS/MS

About 1 µg of a protein digest was loaded onto a UltiMate 3000 RSLC nano HPLC and peptides separated with a 97 min nonlinear gradient (3-40%, 0.1% formic acid). Samples were first loaded onto a 300 µm x 5mm C18 PepMap C18 trap cartridge in 0.1% formic acid at 25 µl/min and passed on to an in-house made 75 µm x 15cm C18 column (ReproSil-Pur Basic-C18-HD, 3 µm, Dr. Maisch GmbH) at 400nl/min. The eluent was directed to an Ab-Sciex TripleTOF 6600 mass spectrometer through the AB-Sciex Nano-Spray 3 source, fitted with a New Objective FS360-20-10 emitter. For data-dependent data acquisition (DDA), MS1 data was acquired within a range of 400-1250m/z (250 ms accumulation time), followed by MS2 of Top 30 precursors with charge states between 2 and 5 (total cycle time 1.8s). Product ion spectra

(50 ms accumulation time) were acquired within a range of 100-1500m/z, using rolling collision energy for precursors which exceed 150 cps. Precursor ions were excluded for 15s after one occurrence. The acquired DDA data was searched against the metagenomics sequence database.

Appendix Q. Two-way ANOVA (Minitab 18) of VSS data from the AnMBRs at different conditions, α=0.05.

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Treatment	1	1432.5	2.50%	1432.5	1432.5	7.11	0.013
Temperature	1	3106.7	5.43%	3106.7	3106.7	15.43	0.001
Phase	1	37462.7	65.46%	37462.7	37462.7	186.01	0.000
Treatment*Temperature	1	575.5	1.01%	575.5	575.5	2.86	0.104
Treatment*Phase	1	2943.4	5.14%	2943.4	2943.4	14.61	0.001
Temperature*Phase	1	5052.6	8.83%	5052.6	5052.6	25.09	0.000
Treatment*Temperature *Phase	1	1825.6	3.19%	1825.6	1825.6	9.06	0.006
Error	24	4833.7	8.45%	4833.7	201.4		
Total	31	57232.6	100.00%				

Appendix R. Two-way ANOVA (Minitab 18) of protein concentration data from the AnMBRs at different conditions, α=0.05.

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Temperature	1	45476	3.64%	45476	45476	0.41	0.541
Phase	1	39	0.00%	39	39	0.00	0.986
Treatment	1	82226	6.59%	82226	82226	0.74	0.415
Temperature*Phase	1	14823	1.19%	14823	14823	0.13	0.725
Temperature*Treatment	1	71690	5.74%	71690	71690	0.64	0.446
Phase*Treatment	1	114075	9.14%	114075	114075	1.02	0.341
Temperature*Phase *Treatment	1	28308	2.27%	28308	28308	0.25	0.628
Error	8	891491	71.43%	891491	111436		
Total	15	1248128	100.00%				

Appendix S. List of identified proteins at FDR 1% and 5% by PEAKS two-round search.

Protein name	Gene name	FDR 1%	FDR 5%
Outer membrane porin protein 32	omp32	73	81
Vitamin B12 transporter BtuB	btuB	14	15
TonB-dependent receptor SusC	susC	9	14
Major outer membrane protein P. IA	porA	9	10
Succinate dehydrogenase flavoprotein subunit	sdhA	2	9
Outer membrane protein W	ompW	7	8
Putative outer membrane protein	Putative Omp	8	8
Elongation factor Tu	tufA	2	8
Outer membrane porin protein	Porin	7	7
47 kDa outer membrane protein	omp 47KDa	2	4
Citrate synthase	gltA	4	4
Long-chain fatty acid transport protein	fadL	3	4
Outer membrane protein P1	ompP1	4	4
Elongation factor G	fusA	3	4
ATP synthase subunit b	atpF	3	3
DNA-directed RNA polymerase subunit beta	rpoB	1	3
Glycerol kinase	glpK	3	3
Outer membrane protein 40	omp40	3	3
Phosphoenolpyruvate carboxykinase [GTP]	pckG	3	3
Porin D	Porin D	2	3
Porin Omp2b	Porin Omp2b	3	3
Succinate--CoA ligase [ADP-forming] subunit beta	SUCLA2	3	3
2-oxoglutarate carboxylase large subunit	cfiA	1	2
30S ribosomal protein S1	rpsA	1	2
3-methylmercaptopropionyl-CoA dehydrogenase	dmdC	2	2
50S ribosomal protein L1	rplA	0	2
50S ribosomal protein L5	rplE	1	2
Acetyl-coenzyme A synthetase	acs	2	2
ATP synthase subunit alpha	atpA	0	2
ATP synthase subunit beta	atpF	1	2
ATP synthase subunit beta 1	atpD	1	2
Biopolymer transport protein ExbB	exbB	2	2
DNA-binding protein HU-beta	hupB	2	2
Flagellin	fliC	2	2
Fumarate reductase flavoprotein subunit	frdA	1	2
Ketol-acid reductoisomerase (NADP (+))	ilvC	2	2
Major outer membrane prolipoprotein Lpp	lpp	1	2
Major outer membrane protein P. IB	porB	2	2
Malate dehydrogenase	MDH	2	2
Outer membrane protein	omp	2	2
Outer membrane protein A	ompA	2	2
Outer membrane protein IIIA	ropA	2	2

Protein name	Gene name	FDR 1%	FDR 5%
Outer membrane protein Omp38	omp38	2	2
Particulate methane monooxygenase alpha subunit	pmoB1	2	2
Peroxiredoxin	Peroxiredoxin	1	2
Phosphate-binding protein PstS	PstS	2	2
Porin	Porin	0	2
Protein oar	oar	2	2
Succinate--CoA ligase [ADP-forming] subunit alpha	sucD	2	2
Transcription termination/antitermination protein NusA	nusA	0	2
60 kDa chaperonin	groL1	0	2
V-type ATP synthase subunit C	atpC	2	2
30S ribosomal protein S16	rpsP	1	1
30S ribosomal protein S3	rpsC	1	1
30S ribosomal protein S5	rpsE	1	1
30S ribosomal protein S7	rpsG	1	1
3-isopropylmalate dehydratase large subunit	IIL1	1	1
50S ribosomal protein L13	rplM	1	1
50S ribosomal protein L28	rpmB	1	1
5-methyltetrahydrofolate: corrinoid/iron-sulfur protein co-methyltransferase	acsE	1	1
Aconitate hydratase B	acnB	1	1
Adenylylsulfate reductase subunit alpha	aprA	1	1
Aerobic glycerol-3-phosphate dehydrogenase	GlpD	1	1
ATP synthase subunit c	atpC	1	1
ATP-dependent RecD-like DNA helicase	recD2	0	1
Biotin transporter BioY	bioY	0	1
Calcium dodecin	Calcium dodecin	1	1
Carbon monoxide dehydrogenase/acetyl-CoA synthase subunit alpha	CODH/acs	1	1
Cation/acetate symporter ActP	actP	1	1
Chaperone protein DnaK	dnaK	0	1
Corrinoid/iron-sulfur protein large subunit	acsC	1	1
Cytochrome c-552	cyt-c552	0	1
DNA-binding protein HRm	HRm	1	1
Electron transfer flavoprotein subunit alpha	etfA	1	1
Electron transfer flavoprotein subunit beta	etfB	1	1
Enolase	eno	0	1
Ethanolamine ammonia-lyase heavy chain	eutB	1	1
Fatty acid oxidation complex subunit alpha	fadB	1	1
Fimbrial protein	fimA	1	1
GDP-6-deoxy-D-mannose reductase	rmd	0	1
Glutamyl-tRNA reductase	hemA	0	1
Glyceraldehyde-3-phosphate dehydrogenase 1	GAPDH	1	1
GTP-binding protein TypA/BipA	TypA/BipA	1	1

Protein name	Gene name	FDR 1%	FDR 5%
Hydrogenase-1 large chain	hyaB	0	1
Inositol 2-dehydrogenase/D-chiro-inositol 3-dehydrogenase	iolG	1	1
Isocitrate dehydrogenase [NADP]	IDH1	1	1
Isocitrate lyase	icl	1	1
Macrolide export protein MacA	macA	0	1
Maltoporin	lamB	1	1
Methylmalonyl-CoA mutase	mcm	1	1
Multidrug efflux pump subunit AcrA	acrA	0	1
NAD(P)H-quinone oxidoreductase subunit I chloroplastic	ndhI	1	1
NADP-dependent malic enzyme	maeB	1	1
Nitric oxide reductase subunit C	norC	0	1
Nitrogen regulatory protein	glnB	1	1
Nucleoside diphosphate kinase	ndk	0	1
Oligopeptide-binding protein AppA	appA	0	1
Outer membrane protein 41	omp41	1	1
Outer membrane protein C	ompC	1	1
Outer membrane protein P6	ompP6	1	1
Outer membrane protein PagN	pagN	1	1
Outer membrane protein X	ompX	1	1
5,10-methylenetetrahydromethanopterin reductase	mer	1	1
Putative adenylyl-sulfate kinase	cysC	0	1
Putative glutamine ABC transporter permease protein GlnM	GlnM	0	1
Putative phospholipase A1	pldA	0	1
Pyruvate dehydrogenase E1 component	PDHA1	1	1
Ribonuclease HII	rnhB	0	1
Ribulokinase	araB	0	1
RNA polymerase sigma factor RpoD	rpoD	1	1
S-layer protein SlpA	slpA	1	1
Superoxide dismutase [Fe]	SODB	1	1
Thioredoxin	Thioredoxin	1	1
Transcription-repair-coupling factor	mfd	0	1
Trigger factor	tig	1	1
Tubulin-like protein CetZ	cetZ	1	1
V-type ATP synthase alpha chain	atpA	1	1

Appendix T. List of all genera associated with three or less expressed proteins.

Class	Genus	Number of expressed proteins
Acidobacteria	Candidatus Solibacter	3
Actinobacteria	Aurantimicrobium	3
	Ilumatobacter	3
	Nocardiopsis	3
	Tessaracoccus	3
Alphaproteobacteria	Agrobacterium	3
	Defluviicoccus	3
	Georhizobium	2
	Caulobacter	2
	Croceicoccus	2
	Paracoccus	2
	Rhodopseudomonas	2
	Roseomonas	2
	Shinella	2
	Stella	2
	Tabrizicola	2
Bacteroidetes	Bacteroidales bacterium CF	2
	Cloacibacterium	2
	Lacinutrix	2
	Alistipes	2
	Dysgonomonas	2
	Elizabethkingia	2
	Filimonas	2
	Flavobacteriaceae bacterium UJ101	2
	Flavobacterium	2
	Labilibaculum	2
	Lutibacter	2
	Parabacteroides	2
	Petrimonas	2
	Prevotella	2
	Rhodothermaceae bacterium RA	2
	Salinivirga	2
	Sphingobacterium	1
Betaproteobacteria	Comamonas	3
	Ephemeroptericola	3
	beta proteobacterium CB	2
	Delftia	1
	Polynucleobacter	1
	Ramlibacter	1
	Rhodoferax	1
	Serpentinomonas	1
	Sulfurimicrobium	1
	Verminephrobacter	1

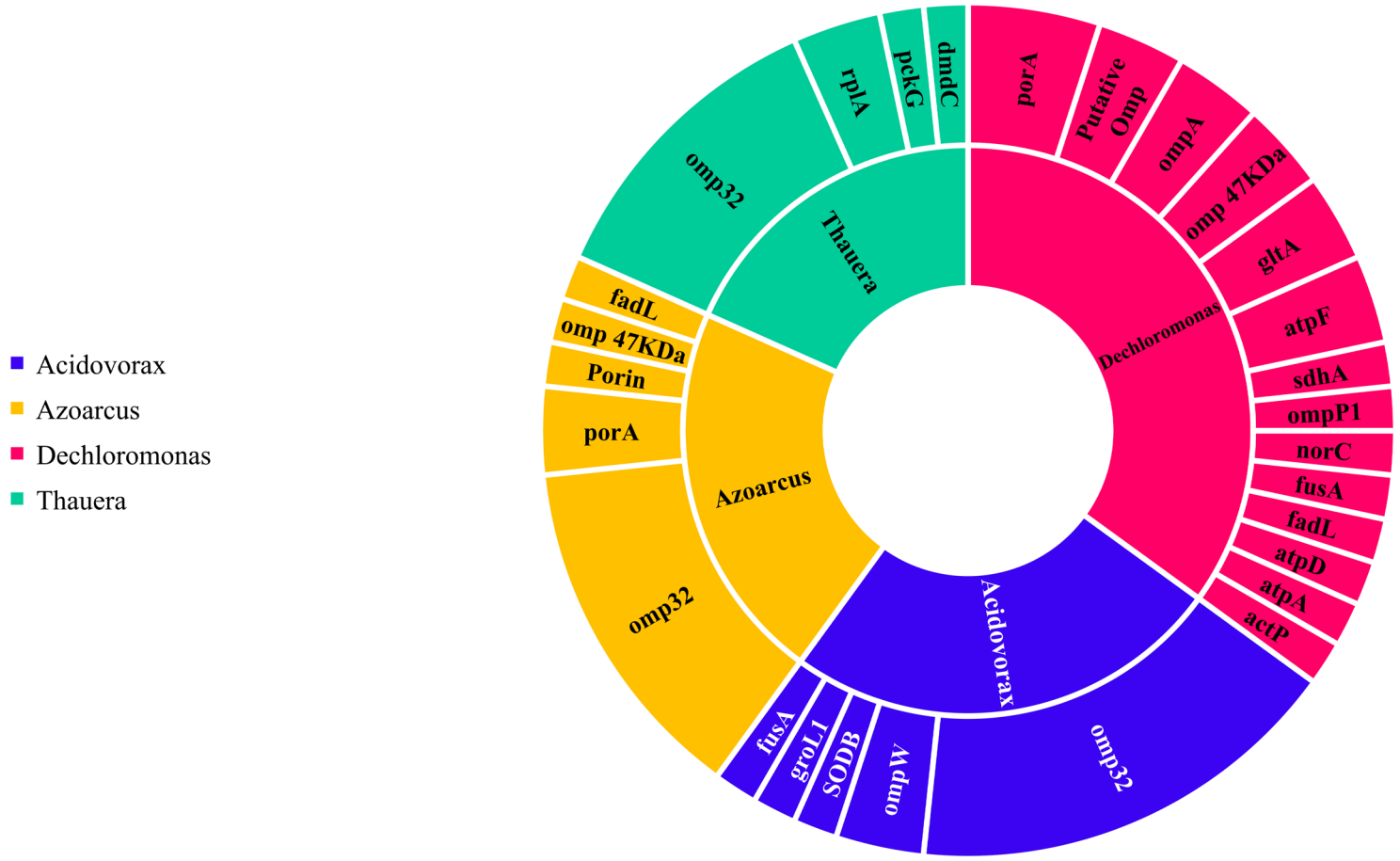
Class	Genus	Number of expressed proteins
	Achromobacter	1
	Cupriavidus	1
	Ferriphaselus	1
	Iodobacter	1
	Methylibium	1
	Methyloversatilis	1
	Nitrosomonas	1
	Pigmentiphaga	1
	Sulfuricella	1
	Sulfuriferula	1
	Undibacterium	1
Chlamydiae	Neochlamydia	1
Chlorobi	Ignavibacterium	1
	Prosthecochloris	1
Chloroflexi	Pelolinea	1
Deltaproteobacteria	Desulfosarcina	1
	Desulfuromonas	1
	Desulfobulbus	1
	Desulfococcus	1
	Anaeromyxobacter	1
	Desulfobacterium	1
	Desulfomonile	1
	Geobacter	1
	Haliangium	1
	Sorangium	1
Epsilonproteobacteria	Sulfuricurvum	1
	Sulfurimonas	1
	Pseudoarcobacter	1
	Sulfurospirillum	1
	Sulfurovum	1
Firmicutes - Bacilli	Thermobacillus	1
Firmicutes - Clostridia	Caldanaerobacter	1
	Caloramator	1
	Caproiciproducens	1
	Moorella	1
	Syntrophomonas	1
	Thermincola	1
Fusobacteria	Ilyobacter	1
Gammaproteobacteria - Enterobacteria	Escherichia	1
	Shigella	1
	Shimwellia	1
Gammaproteobacteria - Others	Acinetobacter	1
	Aeromonas	1

Class	Genus	Number of expressed proteins
	Azotobacter	1
	Methylomonas	1
	Pseudomonas	1
	Aquicella	1
	Dokdonella	1
	Dyella	1
	Entomomonas	1
	Methylocaldum	1
	Methylomicrobium	1
	Microbulbifer	1
	Oblitimonas	1
	Permianibacter	1
	Saccharophagus	1
	Tatlockia	1
	Thermomonas	1
	Thioflavicoccus	1
	Xanthomonas	1
Lentisphaerae	Victivallales bacterium CCUG 44730	1
Saccharibacteria	Candidatus Saccharibacteria oral taxon TM7x	1
Spirochaetes	Salinispira	1
	Treponema	1
	Turneriella	1
Synergistetes	Cloacibacillus	1
Unclassified Bacteria	Candidatus Campbellbacteria bacterium GW2011_OD1_34_28	1

Appendix U. List of associated expressed proteins to Paucimonas.

Gene	Number	Description	KO number
atpD	1	ATP synthase subunit beta	K02112
cfiA	1	2-oxoglutarate carboxylase large subunit	K01960
cysC	1	putative adenylyl-sulfate kinase	K00955
etfB	1	Electron transfer flavoprotein subunit beta	K03521
fadB	1	Fatty acid oxidation complex subunit alpha	K01825
GAPDH	1	Glyceraldehyde-3-phosphate dehydrogenase 1	K00134
GlpD	1	Aerobic glycerol-3-phosphate dehydrogenase	K00111
HRm	1	DNA-binding protein HRm	K03530
maeB	1	NADP-dependent malic enzyme	K00029
ompP6	1	Outer membrane protein P6	K03640
porin D	1	Porin D	K18093
PstS	1	Phosphate-binding protein PstS	K02040
rplE	1	50S ribosomal protein L5	K02931
rplM	1	50S ribosomal protein L13	K02871
rpmB	1	50S ribosomal protein L28	K02902
rpsA	2	30S ribosomal protein S1	K02945
rpsC	1	30S ribosomal protein S3	K02982
rpsE	1	30S ribosomal protein S5	K02988
rpsP	1	30S ribosomal protein S16	K02959
sucD	1	Succinate--CoA ligase [ADP-forming] subunit alpha	K01902
SUCLA2	1	Succinate--CoA ligase [ADP-forming] subunit beta	K01903
tig	1	Trigger factor	K03545
tufA	2	Elongation factor Tu	K02358
TypA/BipA	1	GTP-binding protein TypA/BipA	K06207

Appendix V. Related expressed genes for top-ranked genera identified by proteomics.



Related expressed genes for top-ranked genera identified by proteomics, actP=Cation/acetate symporter, atpA=ATP synthase subunit alpha, atpD=ATP synthase subunit beta 1, atpF= ATP synthase subunit b, dmdC=3-methylmercaptopropionyl-CoA dehydrogenase, fadL= Long-chain fatty acid transporter, fusA=Elongation factor G, gltA=Citrate synthase, groL1= 60 kDa chaperonin, norC=Nitric oxide reductase subunit C, omp 47KDa= 47 kDa outer membrane protein, omp32=Outer membrane porin protein 32, ompA= Outer membrane protein A, ompP1= Outer membrane protein P1, omp W=Outer membrane protein W, pckG=Phosphoenolpyruvate carboxykinase [GTP], porA=Major outer membrane protein P.IA, Porin=Outer membrane porin protein, Putative Omp=Putative outer membrane protein, rplA=50S ribosomal protein L1, sdhA=Succinate dehydrogenase flavoprotein subunit, SODB=Superoxide dismutase [Fe].

Appendix W. Lipolytic patterns in PROSITE.

Identifier	PROSITE accession	Pattern*
LIPASE_SER**	PS00120	[LIV]-{KG}-[LIVFY]-[LIVMST]-G-[HYWV]-S-{YAG}-G-[GSTAC]
LIPASE_GDSL_SER	PS01098	[LIVMFYAG](4)-G-D-S-[LIVM]-x(1,2)-[TAG]-G
LIPASE_GDXG_HIS***	PS01173	[LIVMF](2)-x-[LIVMF]-H-G(2)-[SAG]-[FYW]-x(3)-[STDN]- x(1,2)-[STYA]-[HAGFT]
LIPASE_GDXG_SER	PS01174	[LIVMF](2)-x-[LIVMF]-H-G(2)-[SAG]-[FYW]-x(3)-[STDN]- x(1,2)-[STYA]-[HAGFT]

* Each letter in the patterns stands for an amino acid (aa). For example, in [LIV], L= Leucine, I= Isoleucine, V= Valine. Each aa is separated by a hyphen. The aa inside the square brackets are the permitted one in the position and the aa inside the curly brackets should not exist there.

** Serine is the putative active site

*** Histidine is the putative active site

Appendix X. List of protein families in putative lipase sequences obtained from putative lipolytic MAGs in Chapter 3 that were scanned by different tools.

Tools	Family membership	Number of sequences
InterProScan	None predicted	41
	Lipase, secreted	16
	Streptomyces scabies esterase-like	6
	Epoxide hydrolase-like	4
	GPI inositol-deacylase PGAP1-like	4
	Palmitoyl protein thioesterase	4
	Lipase EstA/Esterase EstB	2
	Lecithin: diacylglycerol acyltransferase	1
PfamScan	alpha/beta hydrolase fold	34
	Sec_lip	15
	GDSL/Acyl family	9
	Palmitoyl protein thioesterase	4
	PGAP1-like protein	4
	No information	3
	Carboxylesterase family	2
	Lipase (class 2)	2
	Serine aminopeptidase, S33	2
	Helix-turn-helix	1
	Lecithin:cholesterol acyltransferase	1
	Putative serine esterase (DUF676)	1
PANTHER grafting	alpha/beta-Hydrolases	18
	BLR7622 Protein	11
	SLL1969 Protein	8
	Family Not Named	8
	Lipase 5	8
	Monoacylglycerol Lipase	7
	Lipase 2	6
	Lecithin-Cholesterol Acyltransferase-Related	3
	Fasting Induced Lipase	2
	SI:DKEY-122A22.2 (serine protease)	1
	GDSL Esterase/Lipase 3	1
	Hydrolase (serine protease)	1
	Uncharacterized	1
	Thioesterase	1
	No hits	1
	Arylacetamide Deacetylase	1
SUPERFAMILY (Superfamily)	alpha/beta-Hydrolases	69
	SGNH hydrolase	9
	SCOP hierarchy in SUPERFAMILY	32
	Bac_Lip	15
	Carboxylesterase	10
	Esterase	5

Tools	Family membership	Number of sequences
SUPERFAMILY (Family)	Carboxylesterase/thioesterase 1	3
	Acylamino-acid-releasing enzyme, C-terminal domain	2
	Carbon-carbon bond hydrolase	2
	DPP6 catalytic domain-like	2
	Epoxide hydrolase	2
	Acetyl xylan esterase-like	1
	Acetylhydrolase	1
	Biotin biosynthesis protein BioH	1
	Haloperoxidase	1
	Pancreatic lipase, N-terminal domain	1
CDD SPARCLE	Abhydrolase/LIP, Sec_lip	16
	AeS: Acetyl esterase/lipase/AeS	15
	Abhydrolase/EstA	14
	MhpC: imeloyl-ACP methyl ester carboxylesterase/MhpC	12
	SGNH_hydrolase/SEST_like	8
	Abhydrolase_1/Abhydrolase_1	3
	PRK14875/acetoin dehydrogenase E2 subunit dihydrolipoyllysine-residue acetyltransferase	3
	Abhydrolase/Abhydrolase_3	3
	SGNH_hydrolase/fatty_acylteransferase_like	1
	No hit	1
	Abhydrolase/Lipase 2	1
	Abhydrolase_1 /Abhydrolase_1 & MhpC/MhpC	1
ScanProsite	Lipases, serine active site (PS001120)	7
	Lipolytic enzymes "G-D-X-G" family, putative histidine active site (PS01173)	1
	Lipolytic enzymes "G-D-S-L" family, serine active site (PS01098)	1
	No lipolytic motif hit	69

Appendix Y. Detailed placement of lipase sequences in protein families by different tools.

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		InterPro	Pfam	Panther	Super family / Family	CDD
583	Krumholzibacteriota	Unassigned	Lipase 2	274	MILIHGGGFKEEDKSG	Yes	PS01173	NoPre	Ab_3	BLR7622	Carboxy, Evalue= 0.032	Aes
803	Bacteroidota	Chlorobium	Lipase 1	287	LYATGHSMGG	No	PS00120?	NoPre	Ab_1	a/b	SCOP, Evalue=0.030	MhpC
403	Bacteroidota	PHOS-HE28	Lipase 3	362	LFVVGSSYGG	No	PS00120?	Epoxide	Ab_1	SI:DKEY	SCOP, Evalue= 0.023	MhpC
396	Bacteroidota	PHOS-HE28	Lipase 3	363	VAVVGNSYGG	No	PS00120?	NoPre	Ab_1	Mono	Epoxide hydrolaseEvalue= 0.041	MhpC

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
1152	Bacteroidota	UBA5266	Lipase 2	321	IVLCGSSAGG	No	PS00120?	NoPre	Carboxy	BLR7622	Carboxy, Evalue= 0.047	Aes
367	Bacteroidota	Lentimicrobium	Lipase 2	305	LVYIHGGGWLGGSKEQI **OR** IVISGESAGG	No	PS01173 OR PS00120?	NoPre	Ab_3	BLR7622	Carboxy, Evalue= 0.054	Aes
50	Bacteroidota	Unassigned	Lipase 1	265	CIMVGHSMGG	No	PS00120?	NoPre	Ab_1	Mono	SCOP, Evalue=0.0012	MhpC
684	Unassigned	Unassigned	Lipase	247	VVIIGHSKGG	Yes	PS00120	Palmitoyl	Palmitoyl	SLL1969	Bac_Lip, Evalue= 0.020	EstA

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
1036	Proteobacteria	QKVK01	Putative	391	YGLWGYSQGG	No	PS00120?	Lip_Sec	Sec_lip	Not Named	DPP6, Evalue=0.061	Sec_lip
1091	Proteobacteria	Paracoccus	Lipase 3	294	AIVVGHSLGG	No	PS00120?	NoPre	Ab_1	a/b	SCOP, Evalue=0.048	MhpC
1359	Proteobacteria	Unassigned	Est A	215	IHFVGHSLGG	Yes	PS00120	NoPre	Serine aminopeptidase	SLL1969	Bac_Lip, Evalue=0.025	Ab_1
22	Proteobacteria	Nitrosomonas	Lipase 3	320	LHIVGHSYGG	No	PS00120?	Epoxide	Abhydrolase_6	a/b	Carbon-carbon bond hydrolase, Evalue=0.032	Ab_1

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
265	Proteobacteria	Rhodocyclaceae	Lipase 1	325	VVFFGDSLSDTG	Yes	PS01098	NoPre	GDSL/Acyl	GDSL Est/Lip	SCOP, Evalue=0.069	SGNH/acylteransferase
967	Proteobacteria	Rhodoferax	Lipase	306	LVLVGHSMGG	Yes	PS00120	Palmitoyl	Palmitoyl	SLL1969	Bac_Lip, Evalue=0.022	EstA
154	Hydrogenedentota	Unassigned	Lipase 2	306	IAVLGNSAGG	No	PS00120?	NoPre	Ab_3	BLR7622	Carboxy, Evalue= 0.007	Aes
609	Omnitrophota	FEN-1322	Lipase 1	220	ISIFGWSLGG	No	PS00120?	NoPre	Ab_1	Serine protease	SCOP, Evalue= 0.021	MhpC

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
631	Spirochaetota	Unassigned	Lactonizing	297	ANIIGHSHGT **OR** AFLLGDSSPDSL	No	PS00120 OR PS01098?!	Palmitoyl	Palmitoyl	Leci	Bac_Lip, Evalue=0.00014	EstA
820	Unassigned	Unassigned	Lipase 2	308	LVWIHGGSWEQFSKEAN	No	PS01173?	NoPre	Ab_3	BLR7622	Carboxy, Evalue= 0.058	Aes
617	Myxococcota	Unassigned	Lipase	423	LLVGGDSREVTVT	No	PS01174?	NoPre	Ab_1	Leci	Bac_Lip, Evalue=0.0034	EstA
617	Myxococcota	Unassigned	Lipase 2	419	LLQIHGGGWVIGDKREQ **OR** LAVTGESAGG	No	PS01173 OR PS00120?	NoPre	Ab_3	BLR7622	SCOP, Evalue=0.069	Aes

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
617	Myxococcota	Unassigned	Est A	289	VHLVTHSLGG	No	PS00120?!	Lip EstA/B	Lipase (class 2)	SLL1969	Bac_Lip, Evalue=0.014	EstA
617	Myxococcota	Unassigned	Lipase	442	FNLVAHSQGG	No	PS00120?!	NoPre	Palmitoyl	Leci	Bac_Lip, Evalue=0.0020	EstA
1501	Desulfobacterota	Unassigned	Lactonizing	253	ISVIAHSMGG	No	PS00120?!	GPI	PGAP1	SLL1969	Bac_Lip, Evalue=0.017	EstA
481	Desulfobacterota	Desulfobacter	Lipase 3	312	FHLAGCSMGG	No	PS00120?	NoPre	Ab_1	Mono	Biotin biosynthesis protein BioH, Evalue=0.056	PRK14875/ acetyltransferase

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
484	Chloroflexota	Unassigned	Lipase 1	274	FILMGHSMGG	No	PS00120?	NoPre	Ab_1	Mono	SCOP, Evalue=9.53e-05	MhpC
484	Chloroflexota	Unassigned	Lipase 1	246	AALAGHSMGG	No	PS00120?	NoPre	Abhydrolase_6	Mono	SCOP, Evalue=0.039	MhpC
484	Chloroflexota	Unassigned	Est A	618	VDLLVHSMGG	No	PS00120?!	Leci	Ab_1	SLL1969	Bac_Lip, Evalue=0.0029	EstA
231	Chloroflexota	Unassigned	Lipase 3	246	AALAGHSMGG	No	PS00120?	NoPre	Abhydrolase_6	a/b	SCOP, Evalue= 7.24e-05	MhpC

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
204	Firmicutes_A	DTU024	Lipase 1	339	WHLMGHSMGG	No	PS00120?	NoPre	Serine aminopeptidase	a/b	Carbon-carbon bond hydrolase, Evalue=0.026	MhpC
1059	Firmicutes_A	UBA1447	Lipase	561	No	No		NoPre	Helix-turn-helix	a/b	SCOP, Evalue=0.0012	EstA
1059	Firmicutes_A	UBA1447	Lipase	211	No	No		NoPre	No info	a/b	SCOP, Evalue=0.040	No hit
1059	Firmicutes_A	UBA1447	Lipase	404	INFVCHSFGG	No	PS00120?	NoPre	Putative serine esterase	a/b	SCOP, Evalue=0.071	EstA

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
1001	Cyanobacteria	Ga0077546	Lipase 2	333	IIYIHGGSFCLGDKVSS **OR** IFLLGHSAGA	No	PS01173 OR PS00120?	NoPre	Ab_3	BLR7622	SCOP, Evalue=0.067	Aes
1001	Cyanobacteria	Ga0077546	Lipase 2	313	IVFIHGGAWLQGDKSE	No	PS01173?	NoPre	Ab_3	BLR7622	Carboxy, Evalue= 0.034	Aes
1001	Cyanobacteria	Ga0077546	Lipase 2	293	VLCIHGGGWSAGHKKDM **OR** IGAMGSSAGG	No	PS01173 OR PS00120?	NoPre	Ab_3	BLR7622	Carboxy, Evalue= 0.072	Aes
328	Cyanobacteria	Ga0077546	Lipase 2	317	IGVWGVSAGG	No	PS00120?	NoPre	Ab_3	BLR7622	Carboxy, Evalue= 0.074	Aes

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
931	Actinobacteriota	67-14	Lipase 3	264	AHIVGNSLGG	No	PS00120?	NoPre	Abhydrolase_6	a/b	Epoxide hydrolase, Evalue=0.026	PRK14875/ acetyltransferase
931	Actinobacteriota	67-14	Putative	414	YLIAGHSQGG	No	PS00120?!	Lip_Sec	Sec_lip	Lipase 5	SCOP, Evalue=0.063	Sec_lip
336	Actinobacteriota	IMCC26207	Lipase	319	VDLVGHSQGG	Yes	PS00120	Lip EstA/B	Lipase (class 2)	FastLip	Bac_Lip, Evalue=0.0099	Lipase 2

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
336	Actinobacteriota	IMCC26207	Putative	180	VGIIGYSQGG	No	PS00120?!	Lip_Sec	Sec_lip	Lipase 5	Pancreatic lipase, Evalue=0.084	Sec_lip
1020	Actinobacteriota	Mycolicibacterium	Lipase 2	362	IGVGGDSAGGGLA	No	PS01174?	NoPre	Ab_3	a/b	Carboxy, Evalue= 0.022	Ab_3
1020	Actinobacteriota	Mycolicibacterium	Triacyl	562	VSVLGDSAGGNIG	No	PS01174?	NoPre	Ab_3	a/b	SCOP, Evalue=0.029	Aes
1020	Actinobacteriota	Mycolicibacterium	Triacyl	570	VSVLGDSAGGNLG	No	PS01174?	NoPre	Ab_3	a/b	SCOP, Evalue=0.072	Aes

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
1020	Actinobacteriota	Mycolicibacterium	Lipase 2	397	IAISGGSAGG	No	PS00120?	Lip_Sec	Ab_3	BLR7622	SCOP, Evalue=0.081	Aes
1020	Actinobacteriota	Mycolicibacterium	Putative	412	VAFWGYSQGG	No	PS00120?!	Lip_Sec	Sec_lip	Lipase 5	SCOP, Evalue=0.067	Sec_lip
1020	Actinobacteriota	Mycolicibacterium	Putative	445	VGLWGYSGGG	No	PS00120?!	Lip_Sec	Sec_lip	Not Named	Acylamino-acid-releasing enzyme, Evalue=0.066	Sec_lip
744	Actinobacteriota	Mycolicibacterium	Lipase	348	VDLVGHSMGG	Yes	PS00120	GPI	PGAP1	SLL1969	Bac_Lip, Evalue=0.0059	EstA

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		InterPro	Pfam	Panther	Super family / Family	CDD
744	Actinobacteriota	Mycolicibacterium	Putative	446	No	No		Lip_Sec	Sec_lip	Lipase 5	DPP6, Evalue=0.069	Sec_lip
768	Actinobacteriota	Mycolicibacterium	Triacyl	477	VSVLGDSAGGGLA	No	PS01174?	NoPre	Ab_3	a/b	SCOP, Evalue=0.095	Aes
768	Actinobacteriota	Mycolicibacterium	Triacyl	537	VSVIGDSAGGGLA	No	PS01174?	NoPre	Ab_3	a/b	SCOP, Evalue=0.066	Aes
768	Actinobacteriota	Mycolicibacterium	Lipase 2	291	YVALGDSAAAGPL **OR** YVALGDSAAAG	No	PS01174 **OR** PS01098?!	Strep_est	GDSL/Acyl	Lipase 2	Esterase, Evalue=0.0023	SGNH/SEST

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
768	Actinobacteriota	Mycolicibacterium	Lipase 2	253	YVALGSSMAAG	No	By alignment PS01098?	Strep_est	GDSL/Acyl	Lipase 2	Esterase, Evalue=0.015	SGNH/SEST
768	Actinobacteriota	Mycolicibacterium	Lipase	353	VDLVGHSMGG	Yes	PS00120	GPI	PGAP1	SLL1969	Bac_Lip, Evalue=0.0072	EstA
768	Actinobacteriota	Mycolicibacterium	Lipase	352	VDLVGHSNGG	Yes	PS00120	GPI	PGAP1	UnChr	Bac_Lip, Evalue=0.0071	EstA

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
768	Actinobacteriota	Mycolicibacterium	Putative	445	IGLWGYSGGG	No	PS00120?!	Lip_Sec	Sec_lip	Not Named	Acylamino-acid-releasing enzyme, Evalue=0.078	Sec_lip
1111	Actinobacteriota	Corynebacterium	Lipase 2	273	YVALGSSMAA	No	PS00120?!	Strep_est	GDSL/Acyl	Lipase 2	Esterase, Evalue=0.076	SGNH/SEST
1111	Actinobacteriota	Corynebacterium	Lipase 2	298	VVVFGDSLTAN	No	PS01098?!	NoPre	GDSL/Acyl	Thioesterase	Acetylhydrolase, Evalue=0.068	SGNH/SEST
1111	Actinobacteriota	Corynebacterium	Lipase 2	250	?MTFGDSFSANPN	No	PS01147?!	NoPre	GDSL/Acyl	No hits	SCOP, Evalue=0.059	SGNH/SEST

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
1111	Actinobacteriota	Corynebacterium	Lipase	339	VDIVAHSQGG	No	PS00120?!	Palmitoyl	Palmitoyl	FastLip	Bac_Lip, Evalue=0.0060	EstA
1111	Actinobacteriota	Corynebacterium	Putative	456	VAFYGYSQGG	No	PS00120?!	Lip_Sec	Sec_lip	Lipase 5	SCOP, Evalue=0.071	Sec_lip
1111	Actinobacteriota	Corynebacterium	Putative	450	No	No		Lip_Sec	Sec_lip	Not Named	SCOP, Evalue=0.072	Sec_lip
1111	Actinobacteriota	Corynebacterium	Putative	471	IGLLGYSGGA **OR** IAPPGKSDGNV	No	PS00120?!	Lip_Sec	Sec_lip	Not Named	Acetyl xylan esterase-like, Evalue=0.047	Sec_lip

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
493	Actinobacteriota	Propionicimonas	Lipase 3	720	YYLVGYSLGG	No	PS00120?	Epoxide	Ab_1	Mono	SCOP, Evalue=0.051	Ab_1 & MhpC
493	Actinobacteriota	Propionicimonas	Lipase 2	258	VLVSGDSAGAAVA	No	PS01174?	NoPre	Ab_3	Aryl	SCOP, Evalue=0.097	Ab_3
493	Actinobacteriota	Propionicimonas	Putative	565	SVLVGESGGR **OR** GSTAGDSLPLVV	No	PS00120 OR PS01098?!	Lip_Sec	Sec_lip	Lipase 5	Carboxy/thioesterase 1, Evalue=0.063	Sec_lip

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
785	Actinobacteriota	Propionicimonas	Putative	569	TVIWGHSQGG	No	PS00120?!	Lip_Sec	Sec_lip	Lipase 5	Haloperoxidase, Evalue=0.081	Sec_lip
205	Actinobacteriota	Unassigned	Lipase 1	293	VHLFGNSMGG	No	PS00120?	NoPre	Ab_1	a/b	SCOP, Evalue=0.071	PRK14875/acetyltransferase
790	Actinobacteriota	UBA10799	Lipase 1	361	VDLFGNSMGG	No	PS00120?	NoPre	Ab_1	a/b	SCOP, Evalue=0.043	MhpC
790	Actinobacteriota	UBA10799	Lipase 3	308	VHVFGNSLGG	No	PS00120?	NoPre	Ab_1	a/b	SCOP, Evalue=0.081	Ab_1

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
790	Actinobacteriota	UBA10799	Putative	420	YVVMGHSQGG	No	PS00120?!	Lip_Sec	Sec_lip	Not Named	Carboxy, Evalue=0.071	Sec_lip
790	Actinobacteriota	UBA10800	Putative	443	LGVYGKSQGG	No	PS00120?!	Lip_Sec	Sec_lip	Lipase 5	Bac_Lip, Evalue=0.058	Sec_lip
1306	Actinobacteriota	Austwickia	Triacyl	306	VIVGGDSAGGQIA	No	PS01174?	NoPre	Ab_3	a/b	SCOP, Evalue=0.018	Ab_3
1306	Actinobacteriota	Austwickia	Lipase 1	358	YVALGDSYSAGIG	No	PS01147?!	Strep_est	GDSL/Acyl	Lipase 2	SCOP, Evalue=0.072	SGNH/SEST

MAGs ID	Phyla	Genera	Prokka	Length (aa)	Putative lipolytic motif	Prosite		IntherPro	Pfam	Panther	Super family / Family	CDD
1306	Actinobacteriota	Austwickia	Putative	368	VALWGYSEGG	No	PS00120?!	Lip_Sec	Sec_lip	Not Named	Carboxy/ thioesterase 1, Evalue=0.062	Sec_lip
428	Actinobacteriota	Austwickia	Lipase	819	IIALGDSYGAREF	No	PS01174?!	Strep_est	GDSL/ Acyl	Lipase 2	Esterase, Evalue= 0.074	SGNH/S EST
428	Actinobacteriota	Austwickia	Lipase 1	339	YVALGDSFSAGIG	No	PS01174?!	Strep_est	GDSL/ Acyl	Lipase 2	Esterase, Evalue= 0.067	SGNH/S EST
428	Actinobacteriota	Austwickia	Putative	369	LALWGYSEGG	No	PS00120?!	Lip_Sec	Sec_lip	Not Named	Carboxy/ thioesterase 1, Evalue=0.066	Sec_lip
737	Actinobacteriota	Rhodoluna	Lipase 3	311	PHLLGHSFGS	No	PS00120?	Epoxide	Abhydrolase_6	Mono	SCOP, Evalue= 0.080	MhpC

Appendix Z. List of taxa and their accession number with lipase genes from Family I.2 obtained from (Kovacic et al., 2018).

Taxa	Accession Number	Conserved lipase box
Burkholderia glumae	Q05489	VNLI**GHSQGG**
Burkholderia cenocepacia	Q1BM22	VNLV**GHSQGG**
Burkholderia multivorans	Q45VN4	VNLV**GHSQGG**
Burkholderia thailandensis	Q2T7L1	VNLV**GHSQGG**
Pseudomonas KWI-56	P25275	VNLV**GHSQGG**
Burkholderia cepacia	P22088	VNLV**GHSQGG**
Pseudomonas luteola	O68551	VNLV**GHSQGG**

Appendix AA. Results of BLASTp and alignment with ClustalOmega for two putative lipases without common lipolytic motifs.

BLASTp hits Accession Number	E-value	Conserved lipolytic motif in the hit	Reference MAG/ Length (aa)	Motif in the Reference MAG
P9WK89	1.1×10^{-30}	IGLW**GYSGGG**	Bin 1111 (450)	VGLF**GIAGGG**
P9WK88	1.5×10^{-30}	IGLW**GYSGGG**		
P9WK89	1.1×10^{-39}	IGLW**GYSGGG**	Bin 744 (446)	IGLW**GWLTGG**
P9WK88	1.5×10^{-39}	IGLW**GYSGGG**		