PhD-FSTC-2013-21
The Faculty of Sciences, Technology and Communication

DISSERTATION

Presented on 19/09/2013 in Luxembourg
to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN BIOLOGIE

by

Hugo ROUME
Born on 14 March 1985 in Paris (France)

MOLECULAR ECO-SYSTEMS BIOLOGY OF LIPID ACCUMULATING MICROBIAL COMMUNITIES IN BIOLOGICAL WASTEWATER TREATMENT PLANTS

Dissertation defense committee

Dr. Paul Wilmes, dissertation supervisor
Luxembourg Centre of Biomedicine (LCSB), Université du Luxembourg

Dr. Karsten Hiller
Luxembourg Centre of Biomedicine (LCSB), Université du Luxembourg

Prof. Dr. Thomas Sauter, Chairman
Life Science Research Unit (LSRU), Université du Luxembourg

Prof. Dr. Gerard Muyzer
Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Neetherlands

Prof. Dr. Thomas Curtis, Vice Chairman
Civil Engineering and Geoscience, Newcastle University, England

Dr. Falck Warnecke
Institute of Microbiology, Friedrich Schiller University of Jena, Germany

Abstract

Biological wastewater treatment is based on the use of microorganisms capable of intense metabolic activity that results in the removal of a large proportion of organic and inorganic contaminants. Given copious amounts of energy-dense organic molecules such as lipids accumulated by the microbial biomass, chemical energy may be directly harnessed from this for biofuel production. Here, lipid accumulating organism (LAO)-enriched microbial communities were studied using a molecular eco-systems biology approach. This involved the development of necessary methodologies including a new comprehensive biomolecular extraction method, yielding high-quality DNA, RNA, proteins and metabolites, as well as bioinformatic approaches for integrating and analysing the derived high-throughput genomic, transcriptomic, proteomic and metabolomic data. At the inception of the project, a full-scale wastewater treatment plant (WWTP) system with a strong presence of LAOs especially during winter months, i.e. the Schifflange WWTP (Esch-sur-Alzette, Luxembourg), was identified and selected for further study. 16S rRNA amplicon sequencing highlighted the presence of ubiquitous lipid accumulating bacteria closely related to *Candidatus* Microthrix parvicella which increase in abundance from autumn to winter over other highly abundant community members belonging to *Alkanindiges* spp. and *Acinetobacter* spp. In order to elucidate compositional, genetic and functional differences between autumn and winter LAO communities, a comparative integrative omic analysis was carried out on rationally selected autumn and winter LAO community samples. The results from metabolomic/lipidomic analyses between intra- and extracellular compartments support previous models of uptake of unprocessed long chain fatty acids (LCFAs) from the wastewater environment and their storage as triacyglycerols within LAOs. Furthermore, a tailored computational framework for the integration of multi-omic datasets into reconstructed community-wide metabolic networks and models was developed. The resulting networks provide overviews of functional capacity of the sampled LAO communities by incorporating gene copy numbers, transcript levels and protein frequency across the two studied environmental conditions. The identification of genes overexpressed, strongly associated with a specific season and/or possessing a high clustering coefficient suggests the existence of keystone genes, analogous to keystone species in species interaction networks. Examples of such keystone genes in the context of the LAO communities include genes coding for proteins involved in nitrogen and glycerolipid metabolism. The existence of such keystone genes opens up exciting possibilities for prediction and control strategies of microbial communities at the dawn of the field of Eco-Systems Biology.

Acknowledgments

My first thanks goes to my supervisor, Dr. Paul Wilmes. I count myself fortunate to have had the opportunity to work on this stimulating and important project with him. The first two years of close supervision were particularly important for finalising this thesis. I would like to further address all my gratefulness to my thesis committee, Dr. Paul Wilmes, Dr. Karsten Hiller and Prof. Thomas Sauter, who gave me indispensable encouragements and support for the hardest times encountered during my PhD. Thanks also goes to Emilie for her guidance and support of my project. Her honest, thorough and thoughtful approach to carrying out science and ability to see the big picture was essential for me. To Venkata, Patrick, René and Nicolàs all my thanks for their indispensable bioinformatic treatment of this "big data". Extracting meaning from meta-omics datasets is still an outstanding challenge, but their involvement in this project largely contributed to the advancement of our understanding of complex microbial communities. I appreciated working with Anna on the writing of the technical paper. Through her contributions, she largely advanced the propagation of the method. Thanks to all of these people for their encouragement, enthusiasm and efforts in explaining things clearly and simply to me. I have learnt a lot from them, in particular from their intellectual curiosity, endless quest for scientific observations and attention to details. I appreciate the countless number of meetings that I had with them, which were always inspiring and refreshing. I might have been lost without these people.

I also thank Dr. Michael Hoopmann, Dr. Mark Sartain and Dr. Robert Moritz from the Proteomics group at the Institute for Systems Biology (ISB, Seattle, USA) for the extraordinary amount of time that they invested in getting me up-to-speed and for

providing me with very detailed protocols. I hope that we will continue to work together on future projects.

Thanks to Nathan Hicks, Dr. John Gillece, Dr. Jim Schupp, Prof. Paul Keim, from the Pathogen Genomics Division at the Translational Genomic Research Institute (TGen, Flagstaff, Arizona, USA), for the great sequencing effort that they contributed to the analyses described in this thesis.

Thanks to Cécile Walczak, Sébastien Planchon, Emmanuelle Coco, Dr. Christian Penny, Dr. Jenny Renaut and Dr. Cedric Guignard from the Centre de Recherche Public Gabriel Lippmann (Belvaux, Luxembourg), and Laura Lebrun from the Luxembourg Centre Systems for Biomedicine (LCSB, Luxembourg) for the assistance and support they provided me during my thesis project.

I would also like to thank Thekla Cordes and the entire Metabolomics group at the LCSB for their dedicated effort to the metabolite identifications of my diverse environmental samples. Thanks also to Maria Pacheco and Prof. Thomas Sauter from the Life Science Research Unit (LSRU) of the University of Luxembourg for the time that they dedicated for the manual curation of my microbial community-wide metabolic model. Mr. Bissen and Mr. Di Pentima from the Syndicat Intercommunal à Vocation Ecologique (SIVEC) are thanked for their permission to collect samples and gain access to the technical platform of the Schifflange wastewater treatment plant.

The present work was supported by an Aide à la Formation Recherche (AFR) grant (PHD-MARP-04) funded by the Luxembourg National Research Fund (FNR).

Finally, I am very grateful to all of the Eco-Systems Biology group at the LCSB, for giving me the opportunity to meet each of them. They progressively joined Paul and

myself during the last year of my PhD and I really enjoyed spending fun and interesting times with them. Thanks to them. I have realised the importance of working with pleasant team members.

Table of Contents

List of Figures

List of Tables

List of Supplementary Tables in Appendix

Supplementary Table I Operational taxonomic units (OTUs) found within the LAO-enriched microbial communities across the four different sampling dates.

Supplementary Table II Raw polar metabolite intensity data for the LAO-enriched microbial communities. The data were filtered so that metabolites had to be detected in more than 70 % of technical replicates. D1 to D4 denote individual sampling dates, I1-1 to I4-4 denote individual islets (biological replicates) and technical replicates, thereof.

Supplementary Table III Bacterial gene abundances identified in the LAO-enriched microbial communities in each of the biological and technical replicates obtained on D4 (23 February 2011).

Supplementary Table IV Abundances of KOs in the metagenomic (DNA) dataset in each of the biological and technical replicates obtain on D4 (23 February 2011).

Supplementary Table V Raw polar and non-polar metabolite intensity data for the intra- and extra-cellular fractions of the representative LAO-enriched microbial community samples obtained on D4 (4 biological replicates analysed). The data are filtered so that metabolites have to be detected in more than 70 % of replicates.

Supplementary Table VI Raw polar and non-polar metabolite intensity data for an LAO-enriched microbial community, river water filtrate and human faeces. The data are filtered so that metabolites have to be detected in more than 70 % of replicates. R1 to R3 denote technical replicates for each of the microbial communities analysed.

Supplementary Table VII Raw non-polar metabolite intensity data for the LAO-enriched microbial communities for the intracellular (INT) and extracellular (EXT) compartment in autumn (04 October 2010) and in winter (25 January 2011) sampling date. The data are filtered so that metabolites have to be detected in more than 70 % of technical replicates. 1 to 3 denote individual technical replicates.

Supplementary Table VIII Abundances of triacylglycerols (TAGs) in the intra- and extracellular compartments in autumn (04 October 2010) and in winter (25 January 2011).

Supplementary Table IX Concentrations of the main long chain fatty acids (LCFAs) in the intracellular and extracellular compartments in autumn (04 October 2010) and in winter (25 January 2011).

Supplementary Table X Abundances of gene copy and transcript numbers corresponding to identified KOs in the metagenomic (DNA) and metatranscriptomic (RNA) datasets in autumn (04 October 2010) and in winter (25 January 2011) respectively, considering the combined metagenomic and metatranscriptomic dataset.

Supplementary Table XI Rate of core reactions (%) in all KEGG pathways using the COBRA model and its refinement using FASTCORE.

Supplementary Table XII KO gene number in combined metagenomic (DNA) and metatranscriptomic (RNA) in autumn (04 October 2010) and winter (25 January 2011).

Supplementary Table XIII Frequency of KOs at the protein level in the autumn (04 October 2010) and in winter (25 January 2011).

Supplementary Table XIV Raw polar metabolite intensity data for the LAO-enriched microbial communities for the intracellular (INT) and extracellular (EXT) compartments in autumn (04 October 2010) and in winter (25 January 2011). The data are filtered so that metabolites have to be detected in more than 70 % of technical replicates. 1 to 3 denote individual technical replicates.

Supplementary Table XV Abundance of pre-selected genes related to lipid metabolism inferred from the metagenomic datasets also identified in the *Microthrix parvicella* BIO17-1 genome.

List of Abbreviations

1D-PAGE	one-dimensional polyacrylamide gel electrophoresis
AMOVA	analysis of molecular variance
AOA	ammonia-oxidising archaea
AOB	ammonia-oxidising bacteria
BOD	biological oxygen demand
bp	base pair
BR	biological replicates
CCC	concordance correlation coefficient
cDNA	complementary DNA
CoA	coenzyme A
COD	chemical oxygen demand
Da	Dalton
DAG	diacylglycerols
DO	dissolved oxygen
EBPR	enhanced biological phosphorus removal
EI	electron impact
emPCR	emulsion polymerase chain reaction
EPS	extracellular polymeric substances
ESI	electrospray impact
eV	electron volt
FAME	fatty acid methyl ester
FFA	free fatty acid
FISH	fluorescence *in situ* hybridisation
FU	fluorescent units
GAO	glycogen accumulating organisms
Gb	gigabyte or gigabases
GC	gas chromatography
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KO	orthologous gene
KOA	KO abundances
L	Ladder
LAO	lipid accumulating organism
LC	liquid chromatography
LCFA	long chain fatty acid
m/z	mass to charge ratio
MALDI	matrix assisted laser desorption/ionization
MAR-FISH	microautoradiography fluorescent in situ hybridisation
MBR	membrane bioreactor
MG-RAST	metagenome rapid annotation using subsystem technology
miRNA	micro ribonucleic acid
mRNA	messenger ribonucleic acid
MS	mass spectrometry

NA	Norgen All-in-One Purification kit
NCBI	National Center for Biotechnology Information
NGA	normalised gene abundance
NGS	next generation sequencing
NMR	nuclear magnetic resonance
NOB	nitrite-oxidising bacteria
Norm	Normalised
nt	Nucleotide
ORF	open reading frames
OTU	operational taxonomic unit
PAH	polycyclic aromatic hydrocarbon
PAO	polyphosphate accumulating organism
PBS	phosphate buffered saline
PC	principal component
PCA	principal component analysis
PCoA	principal coordinate analysis
PCR	polymerase chain reaction
PFA	Paraformaldehyde
PHA	Polyhydroxyalkanoate
PHB	poly-3-hydroxybutyrate
pM	pico molar
ppm	Parts per million
PTS	phosphotransferase system
QA	Qiagen allPrep DNA/RNA/Protein Mini kit
qPCR	quantitative polymerase chain reaction
rRNA	ribosomal ribonucleic acid
RDP	ribosomal database project
RIN	RNA integrity number
RM	reference method
rpm	rotation per minute
RGE	relative gene expression
rRNA	ribosomal ribonucleic acid
RT	retention time
SAF	season association factors
SCFA	short chain fatty acids
SDS	sodium dodecyl sulphate
SIVEC	Syndicat Intercommunal à Vocation Ecologique
SOR	seasonal overrepresentation factor
st. dev.	standard deviation
TAG	Triacylglycerol
TE	Tris/ethylenediaminetetraacetic acid
TIC	total ion chromatograms
ToF	time-of-flight
TR	TRI Reagent

tRNA	transfer ribonucleic acid
TSS	total suspended solids
v/v	volume per volume
VFA	volatile fatty acid
w/v	weight per volume
WE	wax ester
WWTP	wastewater treatment plant
x	solution time concentrated
X	Magnification

CHAPTER 1: INTRODUCTION

1.1. Biological wastewater treatment: the fate of fat

1.1.1. Lipids: molecules of immediate bioenergy interest

1.1.1.1. Lipids

The term lipid (from Greek "lipos", meaning "fat, grease"), defined as an "organic substance of the fat group" by Gabriel Bertrand in 1923 (Gidez, 1984), constitutes a broad group of naturally occurring compounds (e.g. oils, fatty acids, glycerides) and other key biological constituents, such as cholesterols, some vitamins and hormones.

Lipids are broadly defined as hydrophobic, non-polar, or amphilic molecules. They are ubiquitous in nature and are synthetised or degraded using various biosynthetic pathways. Lipids are major structural components of cellular membranes, important signalling molecules and efficient compounds for energy storage. A recent review classified lipids into eight different families comprising fatty acids, glycerolipids, glycerophospholipids, sphingolipids, saccharolipids, polyketides, sterol lipids and prenol lipids (Fahy *et al.*, 2007). Glycerophospholipids, the main components of biological membranes, consist of hydrophobic tails and a hydrophilic head (amphiphilic molecule). They allow the self-assembly of lipid bilayer structures such as vesicles and liposomes in aqueous environments and guarantee cell membrane integrity. Lipids provide energy buffers when energy intake is not equal to energy expenditure. The complete oxidation of a gram of fatty acids provides approximately 9 kcal, compared with 4 kcal for the breakdown of a gram of carbohydrates or proteins (Alvarez & Steinbüchel, 2002). By being anhydrous they exhibit a minor grade of oxidation and possess a higher calorific value than protein or carbohydrates (Manilla-Pérez *et al.*, 2010). Glycerolipids, such as triacyglycerides (TAGs), a

glycerol derivative and three long chain fatty acids (LCFAs), are commonly found in most eukaryotes and in some specific bacterial groups. Each TAG has a glycerol backbone to which is esterified three fatty acids differing in chain length and number of unsaturations. TAGs represent particular convenient storage compound for carbon and energy, because of their relative compactness. At ambient temperature, most of the saturated TAGs occur as solids (fats), while unsaturated TAGs exist as liquids (oils). The chain length of the main fatty acids in naturally occurring TAGs varies between 16 and 20 carbon atoms (i.e. LCFAs, **Table 1.1**).

Table 1.1 Most commonly occurring LCFAs consituting naturally occurring TAGs.

Common name	Chemical structure	Lipid numbers*
Palmitic acid	$CH_3(CH_2)_{14}COOH$	16:0
Palmitoleic acid	$CH_3(CH_2)_5CH=CH(CH_2)_7COOH$	16:1
Stearic acid	$CH_3(CH_2)_{16}COOH$	18:0
Oleic acid	$CH_3(CH_2)_7CH=CH(CH_2)_7COOH$	18:1
Linoleic acid	$CH_3(CH_2)_4CH=CHCH_2CH=CH(CH_2)_7COOH$	18:2
α-Linolenic acid	$CH_3CH_2CH=CHCH_2CH=CHCH_2CH=CH(CH_2)_7COOH$	18:3
Arachidic acid	$CH_3(CH_2)_{18}COOH$	20:0

* The lipid number has the form c:d, where c is the number of carbon atoms in the fatty acid and d is the number of double bonds in the fatty acid moiety.

1.1.1.2. Lipids: a feedstock for biodiesel production

Decreasing reserves of fossil fuels, as well as progressing climate change caused by the net increase in atmospheric CO_2 levels due to their combustion, are major drivers for initiatives to search for alternative source of energy (Singh & Singh, 2010). Fuels

derived from renewable resources, e.g. biofuel, will form an important part of the energy mix in years to come (Subramaniam *et al.*, 2010).

Chemically, biodiesel is a mixture of mono-alkyl esters, commonly referred to as fatty acid methyl esters (FAMEs; **Figure 1.1**) and can be produced from vegetable oils or animal fat. The use of plant oils as fuels was first demonstrated at the International Exhibition in 1900 in Paris, when the inventor of the diesel engine, Rudolph Diesel, first tested peanut oil in a compression ignition engine (Parawira, 2010). In 1912, at the Great Britain Technological Institute, Diesel proclaimed that "Usage of plant oil as fuel for cars may seem insignificant today. But, such oils may become in the course of time as important as the petroleum and coal tar products of the present time" (Feofilova *et al.*, 2010). The name biodiesel was introduced in 1992 by the US National Soy Diesel Development Board, which pioneered the commercialisation of biodiesel. Since then, biodiesel has attracted worldwide growing interest as a renewable, biodegradable and non-toxic fuel (Singh *et al.*, 2010). Biodiesel provides a similar energy density to fossil fuels, is compatible with the existing fuel infrastructure and can be used in most diesel-based engines in pure form or as a mixture with standard petroleum diesel (Shi *et al.*, 2011). Additionally, use of biodiesel as fuel would reduce sulphur and carbon monoxide emissions from vehicles by 30 and 10 %, respectively (Subramaniam *et al.*, 2010).

Biodiesel production is based on the transesterification and/or esterification of various lipid sources, in the presence of an alcohol (methanol or ethanol), a homogenous base, acid or an enzyme catalyst (Siddiquee *et al.*, 2011). Lipid feedstocks for biodiesel production are, mainly, composed of TAGs and/or free fatty acids (FFAs). TAGs react with the alcohol to produce FAMEs, in addition to glycerol using the broad

range of catalysts (**Figure 1.1, A**). Acid catalysts are being preferentially used to esterify FFAs to produce biodiesel with water as a by-product (**Figure 1.1, B**). The feedstock for biodiesel production is mainly comprised of TAGs with long-chain carboxylic acids, from C_{12} to C_{20}, which can either be saturated or unsaturated (Azócar *et al.*, 2010). The esterification of LCFAs results in biodiesel with a higher cetane number and reduced NO_x emissions (Subramaniam *et al.*, 2010).

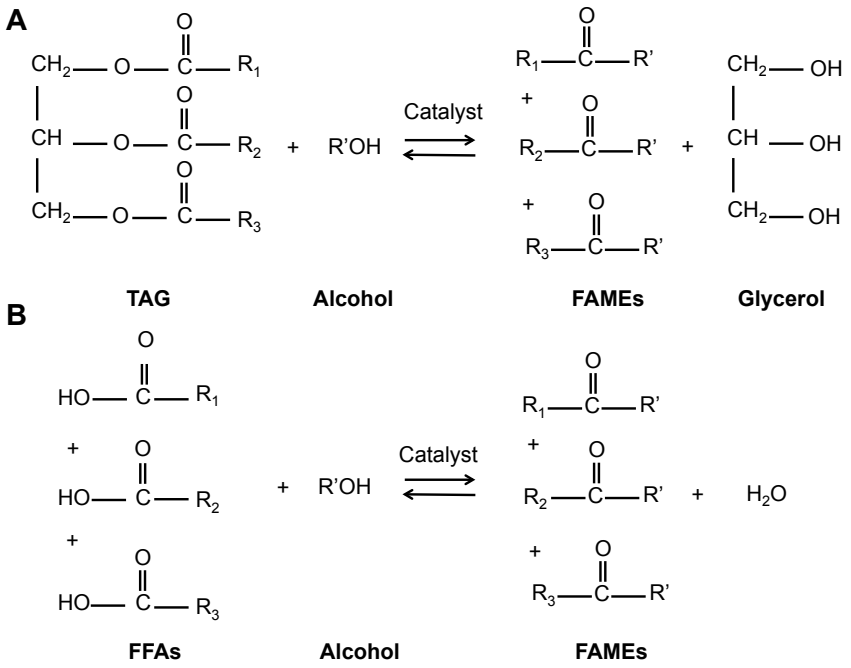

Figure 1.1 Transesterification of TAGs (**A**) and FFAs (**B**) with an alcohol yielding biodiesel. R1, R2, R3 and R' represent alkyl groups. Reproduced from Mondala *et al.*, 2009.

In conventional industrial biodiesel production processes, a mixture of methanol, vegetable oil (molar ratio 1:6) and NaOH as catalyst is used (Rashid *et al.*, 2008). As the solubility of these compounds is low, stirring (200 rpm) and high temperature (60 °C) are typically used to guarantee acceptable kinetic reactions (Azòcar *et al.,* 2010). After the transesterification reaction, the glycerol (bottom layer) is separated

from the FAMEs (upper layer) by phase separation, usually, using water. The excess methanol or water (in case of esterification) in the FAME fraction is then recovered by distillation and the glycerol may be further refined for future use (Azócar *et al.*, 2010). Therefore, the excess of methanol or water can be fully recovered and recycled back to the start of the process.

Lipases, which are enzymes that hydrolyse lipids, may replace alkaline or acid catalysts. They should theoretically be more compatible with variations in the quality of the raw material due to their specificity and should be reusable. Furthermore, the use of lipases should allow the production of biodiesel in fewer process steps using less energy and lead to improved product separation and a higher quality of glycerol in the case of using TAG feedstocks (Fjerbaek *et al.*, 2009). However, commercially available lipases currently have a low reaction rate, are expensive and lose their activity within short period of time, if methanol is added in more than equimolar amounts (Fjerbaek *et al.*, 2009). Consequently, there exists currently a concrete need for the discovery of novel efficient lipases for biodiesel production.

Over the last 10 years, biodiesel production has increased 92-fold in the USA and the U.S Energy Information Administration (EIA) has estimated a production of 969.4 million gallons (3.7 billion of liters) in 2012. However, more than 95 % of the world's biodiesel is currently produced from edible vegetable oils such as soybean, sunflower and palm oil (Azócar *et al.*, 2010) and 70 to 85 % of the overall biodiesel production cost is associated with the primary raw materials (Siddiquee *et al.*, 2011). Moreover, the raw materials for biodiesel production compete for the land available to agriculture, thereby, reducing food security. Alternative process feedstocks like algae, waste cooking oil, wastewater sludge and non-edible plant oils such as those from

Jatropha curcas, *Ricinus communis*, *Azadirachta indica*, *Millettia pinnata* are, therefore, increasingly being considered for biodiesel production (Siddiquee *et al.*, 2011).

Microbial biomass, potentially, has important advantages over plant-based biodiesel feedstocks, because of its rapid generation time related to short microbial cell doubling times, less demand on space and ease of industrial scale-up (Subramaniam *et al.*, 2010). Additionally, certain microorganisms belonging to several different families such as microalgae, bacteria and fungi, possess the ability to accumulate a large fraction of their dry mass in the form of lipids. Microbes possessing a lipid content of more than 20 % of their biomass are classified as 'oleaginous'. Oleaginous bacteria represent promising raw materials for biodiesel production. Especially since they exhibit particularly rapid cell growth rates and include certain species of *Mycobacterium* spp., *Streptomyces* spp., *Rhodococcus* spp. and *Norcodia* spp. (Alvarez & Steinbüchel, 2002).

1.1.2. Biological wastewater treatment

A conventional wastewater treatment plant (WWTP) consists of a combination of physical, chemical and biological stages to remove organic matter and nutrients from wastewater. Discharges of excess amounts of organic and inorganic nutrients such as carbohydrates, fats, nitrogen and phosphorus from populated areas directly into receiving water bodies may lead to severe perturbations of these aquatic ecosystems. The organic compounds stimulate the growth of heterotrophic organisms, leading to a reduction in dissolved oxygen and entailing nitrogen and phosphorus removal. The need to reduce both nitrogen and phosphorus inputs to freshwater and costal marine

ecosystems has been widely recognised, in order to protect drinking-water supplies and reduce eutrophication, including the algal blooms and aquatic "dead zones" (Conley *et al.*, 2009). From these considerations, the objectives of wastewater treatment are (i) the reduction of organic bond energy, to a level where it will no longer sustain heterotrophic growth and, thereby, avoid deoxygenation (i.e. removal of oxygen) effects, (ii) oxidise ammonia to nitrate to reduce its toxicity and deoxygenation effects, and (iii) reduce eutrophic substances such as phosphate (Mara & Horan, 2003). In biological wastewater treatment, two basic categories of microorganisms are of specific interest: heterotrophic organisms and lithoautotrophic nitrifying organisms (Paul & Liu, 2012). Heterotrophic organisms utilise organic compounds within wastewater as electron donors and either oxygen or nitrate as terminal electron acceptor, depending on their cellular respiration. The advantages of biological wastewater treatment processes include lower energy and chemical consumption as well as lower waste production when compared to chemical processes. In Luxembourg, 109 biological wastewater treatment plants (WWTPs), handle approximately 95 % of the total wastewater (Administration de la Gestion de l'Eau, 2013).

1.1.2.1. Constituents of wastewater

The characteristics of wastewater vary depending on their source, primarily on domestic, agricultural and/or industrial sectors supplying the WWTP. Wastewater contains a variety of organic and inorganic compounds, either in solution or as particulate matter, which is of anthropogenic, i.e. household waste liquid, and/or natural origin. According to Metcalf and Eddy (2003), total suspended solids (TSS) in municipal wastewater are around 720 mg/l of which approximately 70 % is organic.

1.1.2.1.1. Organic constituents of wastewater

The degradation of organic matter in wastewater is measured by the biological oxygen demand (BOD) and the chemical oxygen demand (COD), both expressed in milligrams per litre. The BOD is defined as the measure of oxygen necessary to decompose the organic matter present in wastewater through the action of bacteria under aerobic condition at 20 °C over 5 days (BOD_5, Metcalf & Eddy, 2003). The COD is the most commonly used parameter for determining the organic content of wastewater. It indicates the amount of oxygen consumed to chemically oxidise organic compounds to inorganic end products, such as ammonia (Metcalf & Eddy, 2003). In domestic wastewater, the organic matter fraction is comprised of total protein, carbohydrate and lipids which correspond, approximately, to 28, 18 and 31 % of COD, respectively (Raunkjær *et al.*, 1994). Additionally, municipal wastewater typically contains a large number of different synthetic organic molecules such as pesticides, herbicides, pharmaceuticals and endocrine disrupting chemicals (Kim *et al.*, 2007).

The lipid fraction present in wastewater can be comprised of oils, greases, waxes, and fats (Chipasa & Mędrzycka, 2006). The main sources for the lipids in raw domestic wastewater are kitchen wastes (14-36 %) and human excreta (4-23 %; Quéméneur & Marty, 1994). The major part of lipids in wastewater are present as TAGs and a minor part as free LCFAs (Dueholm *et al.*, 2001). The fatty acid compositions of TAG and LCFAs are very similar. Sources of animal- and vegetable-based fats and oils, the lipid composition of wastewater is dominated by four long chain fatty acids: palmitic acid (C16:0, **Table 1.1**), stearic acid (C18:0), oleic acid (C18:1) and linoleic acid (C18:2), which make up approximately 80 % of the LCFAs in wastewater (Quéméneur & Marty, 1994).

1.1.2.1.2. Reclamation of chemical energy contained within wastewater organic molecules

Organic molecules, especially lipids, within wastewater incorporate large amounts of chemical energy. Unfortunately, this energy is currently not effectively recoverable from wastewater because current processes promote the oxidation of organic matter to carbon dioxide. Only a minor fraction of the energy is recoverable after assimilation of the organic matter into the activated sludge biomass followed by anaerobic digestion and the burning of the resulting biogas for electricity generation (Metcalf & Eddy, 2003). Alternative energy recovery approaches involving the generation of electricity from wastewater oxidation using microbial fuel cell technology have yet to match anaerobic digestion in terms of energy efficiency (Logan & Rabaey, 2012).

However, Mondala *et al.* (2009) were able to produce biodiesel (FAMEs) from primary and secondary wastewater sludge using a standard acid-catalyzed *in situ* transesterification reaction. They obtained a maximum yield of 14.5 % FAMEs per primary sludge dry weight (Mondala *et al.*, 2009). Based on a 10 % FAMEs yield, they estimate a production cost of $ 3.23 per gallon (0.88 € per litre) of neat biodiesel from their process which is comparatively lower than the costs of petroleum diesel and alternative biodiesels (Mondala *et al.*, 2009). Furthermore, based on a similarly efficient transesterification process to Mondala *et al.* (2009), Dufreche *et al.* (2007) calculated that the integration of lipid extraction and transesterification processes in 50 % of all existing municipal WWTPs in the United States of America could produce approximately 1.8 billion gallons (6.8 billion litres) of biodiesel which is roughly equivalent to 0.5 % of the yearly national petroleum diesel demand. However, the yield and cost-effectiveness of wastewater biodiesel may still be dramatically improved using lipid rich biomass as a feedstock and a lipase-catalysed

transesterification approach. Furthermore, the above calculations do not take into account energy production from the residual non-lipid biomass (Rittmann, 2008). Overall, large-scale biodiesel production from lipid-rich biomass in wastewater would have several advantages including carbon neutrality, high energy density, handling convenience and low cost. Consequently, wastewater biodiesel would circumvent numerous limitations of other biofuels (Siddiquee *et al.,* 2011; Kwon *et al*., 2012).

1.1.2.1.3. Inorganic constituents of wastewater

Inorganic constituents play a major role in biological wastewater treatment processes and include compounds such as nitrogen, phosphorus and trace amounts of sulphur, zinc, copper and iron (Metcalf & Eddy, 2003). The nitrogen present in wastewater is partially derived from organic compounds, mainly proteins and urea (or is hydrolysed from ammonia) compounds. The main source of nitrogen is fertilizer run-off from agricultural lands which is diffuse and, hence, difficult to control (Beman *et al.*, 2005). During wastewater treatment, nitrogen-containing compounds can undergo several chemical/biological reactions through nitrification/denitrification processes which result in the transformation of ammonium (NH_4^+) into dinitrogen gas (N_2). Phosphorus is present in organic form, orthophosphate (PO_4^{3-}) or polyphosphate. Most of the phosphorus results from localised anthropogenic sources, i.e. faecal material, industrial and commercial sources, and synthetic detergents (Seviour *et al.*, 2003).

1.1.2.2. Wastewater treatment

Traditional wastewater treatment involves three stages, defined as preliminary, primary and secondary treatments (Sonune & Ghate, 2004). The preliminary

treatment consists of the physical separation of suspended solids from the wastewater flow using coarse screens and grit removal (Sonune & Ghate, 2004). The objective of primary treatment is the removal of settleable organic and inorganic solids by sedimentation as well as grease and oil removal by skimming. The secondary treatment involves the removal or reduction of organic matter in the wastewater using aerobic biological treatment processes. Microorganisms metabolising the organic matter convert non-settleable solids, i.e. biomass to settleable solids, allowing the suspended organic nutrients to be removed by sedimentation (Tchobanoglous & Burton, 2003).

1.1.2.2.1. The activated sludge process

The activated sludge wastewater treatment process relies on microorganisms forming aggregates (flocs) which drive the reduction in nutrient levels. In the activated sludge process, the main fates of degradable organic compounds are either to be assimilated into the microbial biomass or to be oxidised and released as carbon dioxide. The activated sludge process, arguably, represents the most widely used biotechnological process in the world (Seviour, 2010). It, along with its derivatives, is typically used as the secondary treatment stage in WWTPs and has the advantage of producing a high quality effluent for a reasonable operating and maintenance cost.

One hundred years ago, E. Arden and W.T. Lockett first described the conventional activated sludge process in England (Kraume *et al.*, 2005). After small-scale trials, Arden and Lockett observed that the aeration of wastewater led to the formation of flocs. They also discovered that organic contaminants were most efficiently removed when the flocs were recycled within the system. The critical step of the Ardern-Lockett setup involves recycling most of the sludge from the end of the treatment

process. The resulting sludge from the settling process was, thus, referred to as "activated" (Kraume *et al.*, 2005).

1.1.2.2.2. Process configuration

Since its development, the activated sludge process has undergone and continues to experience many changes in its operational features and designs to improve its efficiency. In the past fifty years, many plants have been designed and built to further allow nitrogen and phosphate removal. The first indication of phosphate removal by an activated sludge-type system was described by Srinath *et al.*, (1959), after an accidental failure at aeration of an activated sludge plant. Barnard and co-workers (Barnard, 1975) recommended then the separation and recycling of the anoxic and aerobic zones to improve plant performance. They further designed an evolving series of plant configurations, culminating in a new activated sludge process configuration for enhanced biological phosphorus removal (EBPR; Seviour *et al.*, 2003). Nowadays, most activated sludge wastewater treatment systems, with alternating anaerobic and aerobic phases, are designed and operate globally for EBPR **(Figure 1.2**; Seviour *et al.*, 2003). The main feature of this process is the accumulation of phosphorus in the sludge, achieved by the enrichment of polyphosphate accumulating organisms (PAOs) through the alternating anaerobic/aerobic conditions and followed by the removal of phosphorus rich biomass by gravity filtration or sedimentation (Blackall *et al.*, 2002).

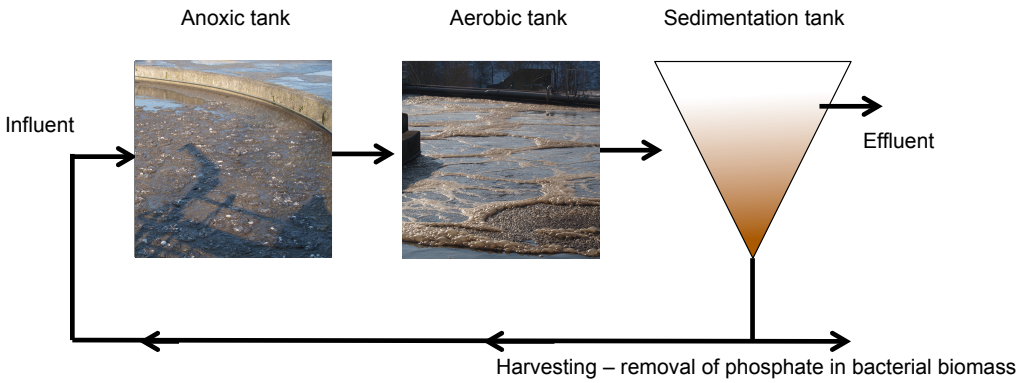

Figure 1.2 Typical layout of an activated sludge plant operated for enhanced biological phosphorus removal. Modified from Blackall *et al.*, 2002.

For particularly high nitrogen concentrated wastewater such as industrial wastewater, leachate or effluent from anaerobic digesters, several specific nitrogen removal processes and technologies have been discovered and developed (Paredes *et al.,* 2007).

1.1.2.2.3. The monitoring of wastewater treatment processes

Physical and chemical factors are essential to allow engineers to monitor and assess the performance of the activated sludge process. Oxidation-reduction potential is the measurement of the ability of the system to either accept electrons (reduction) or donate electrons (oxidation) and provides valuable information about the biological activity of the biomass. It is used as a control parameter and indicator of process efficiency (Schuyler *et al.*, 2011). Aeration has two purposes in activated sludge systems: oxygen supply and mixing. When oxygen limits the growth, filamentous microorganisms may predominate, causing formation of porous flocs, deteriorated settling and a poor activated sludge quality. Sufficient amount of oxygen have to be supplied to the system to allow microbial metabolism of organic matter and the conversion of ammonium to nitrate (i.e. nitrification). However, excessive aeration

negatively affects the activated sludge treatment and, therefore, needs to be properly controlled (Brdjanovic *et al.*, 1998).

The pH of the environment is a key factor for regulating the growth and selection of microorganisms (Kim *et al.*, 2001). As most of the bacteria cannot tolerate pH levels below 4 or above 9, an optimal pH in activated sludge is general between 6.5 and 7.5 (Libhaber & Orozco-Jaramillo, 2012). To retain the system at this optimum pH range, bases are added to the process (Libhaber & Orozco-Jaramillo, 2012).

Wastewater temperature is another critical factor for maintaining efficient wastewater treatment as the optimal growth and activity of a particular organismal groups are typically related to a specific temperature ranges. Temperature also has a strong effect on gas transfer rates and sludge settling characteristics (Rittmann & McCarty, 2001). The solubility of oxygen in liquids increases as the temperature decreases, suggesting that less aeration is needed to supply the required amount of oxygen in case of cold temperature (Rittmann & McCarty, 2001). Seasonal variations of biomass density and settling ability have been reported in some wastewater treatment systems, although this is not majorly dependent on the presence of specific microorganisms which cause bulking events themselves the results of overgrowth by filamentous bacteria (Jones & Schuler, 2010).

In order to monitor process performance, the on-line monitoring sensors are used to measure the characteristics of influents and effluents. These measurements are typically limited to orthophosphate (PO_4), ammonium (NH_4), nitrate (NO^-_3) and dissolved oxygen (DO) concentrations allowing the operator to ensure phosphate and ammonium removal as well as the control of anaerobic conditions (Metcalf & Eddy, 2003). The control of nitrate concentrations allows anoxic conditions to be optimally

maintained. The DO is, generally, one of the major parameters which is monitored. It is commonly suggested that a minimum of 1.5 – 2 mg/l of DO (Metcalf & Eddy, 2003) and optimum temperatures between 25 °C and 35 °C (Rossetti *et al.*, 2005) should be maintained in order to suppress the growth of filamentous organisms, which can severely disrupt the performance of activated sludge systems.

1.1.2.3. The microbial ecology of the activated sludge process

The activated sludge process is a microbiological process which relies on the selection and activity of microorganisms which are able to hydrolyse and then either assimilate or respire diverse organic substrates and to recycle elements such as nitrogen and phosphorus (Daims *et al.,* 2006b). Activated sludge comprises different specialised organisms which occupy distinct niches defined by their substrate usage. For example in relation to organic biotransformations, specialised glycogen accumulating organisms (GAOs) accumulate intracellular glycogen stores under anaerobic conditions and respire these under aerobic conditions (Kong *et al.,* 2006). The fate of organic molecules within biological wastewater treatment is of immediate scientific interest due to their pivotal importance as pollutants and as rate-limiting nutrients for microbial processes related to nitrogen and phosphorus removal. Furthermore, they are of bioenergy interest as discussed previously (Section 1.1.2.1.2). However, the fate of organic matter within biological wastewater treatment biomass is rather under-researched compared to phosphorus and nitrogen transformations (Morgenroth *et al.*, 2002). In the activated sludge process, microbial ecology and environmental biotechnology are inherently tied to one another (Rittmann, 2006).

Organisms found in activated sludge are typically divided into filamentous and non-filamentous bacteria. Filamentous bacteria are further subdivided into Gram-negative,

such as α-proteobacteria, β-proteobacteria or γ-proteobacteria and Gram-positive bacteria such as *Actinobacteria*, *Chloroflexi* or *Planctomycetes* (Seviour & Nielsen, 2010). Non-filamentous bacteria include Gram-negative organisms such as *Bacteroidetes* or Gram-positive organisms such as *Firmicutes* (Yu & Zhang, 2012).

Investigating changes in microbial population dynamics in response to operational conditions and plant performance should be the most meaningful, if not the simplest way, to monitor activated sludge process performance (Seviour & Nielsen, 2010). Additionally, wastewater treatment systems, by being a chemically and physically well-defined system, provide a fertile testing ground to address a range of fundamental ecological questions (Daims *et al.*, 2006b). In this context, the application of modern molecular tools to activated sludge biomass has yielded important insights into general microbial diversity and physiology. An important example in this context is the discovery of the anaerobic ammonium oxidation (anammox) process (Mulder *et al.*, 1995). In the following part, two key aspects of activated sludge microbiology in relation to the results from the present work are discussed: nitrogen removal and filamentous bacterial growth associated with sludge bulking.

1.1.2.3.1. Nitrogen removal

Microbially mediated nitrogen removal from wastewater comprises three main processes: nitrification, denitrification and anaerobic ammonium oxidation (anammox) involving three groups of chemo-autolithotrophic bacteria (Schmidt *et al.,* 2003). In nitrification, aerobic ammonia-oxidising bacteria (AOB) and archaea (AOA) oxidise ammonium (NH_4^+) to nitrite (NO_2^-) via hydroxylamine (NH_2OH; Junier *et al.*, 2010). The key enzyme involved in this process is ammonia monooxygenase which

catalyses the first essential oxidation step of ammonia to hydroxylamine (Martens-Habbena *et al.,* 2009). In the second step of nitrification, nitrite-oxidising bacteria (NOB) oxidise nitrite to nitrate (NO_3^-; Sorokin *et al.*, 2012). In the anoxic denitrification phase, nitrite and nitrate are reduced to gaseous dinitrogen with a variety of electron donors from organic substances present or added to wastewater. In the anammox process, ammonia is oxidised to nitrogen by anaerobic AOB with nitrite as an electron acceptor (Zhu *et al.*, 2008). For nitrification, the main nitrifiers originally detected in wastewater plants comprised of *Nitrosomonas* spp. as major AOB organism and *Nitrobacter* species as main NOB. However with the development of cultivation independent methods, the total diversity of nitrifiers has largely extended to additional key members of AOB such as *Nitrosococcus mobilis* (Juretschko *et al.*, 1998) and NOB as *Nitrospira* spp. (Daims *et al.*, 2001). The NOB community members, within the *Nitrospira* genus, revealed ecological differentiation for different nitrite concentrations (Maixner *et al.*, 2006). The detection of a unique monooxygenase gene (*amoA*) on an archaea-associated scaffold from the samples of the Sargasso Sea (Venter *et al.*, 2004) allowed the detection of AOA occurring in activated sludge systems and these have been found to play a major role in ammonia removal (Park *et al.*, 2006).

Bacterial members capable of catalysing anaerobic ammonium oxidation, identified as a novel member of the phylum *Planctomycetes* (Strous *et al.*, 1999), occur in large amounts in activated sludge systems (Schmid *et al.*, 2000). Conventional wastewater treatment systems using the traditional nitrification-denitrification setup require abundant energy to create aerobic conditions for bacterial nitrification and to also use organic carbon (i.e. methanol) to help remove nitrate by bacterial denitrification. Because the aeration is required only for the initial partial oxidation of ammonia to

nitrite, the anammox process requires less energy. These advantages have been exploited for nitrogen removal (Windey *et al.*, 2005). The anammox process is currently applied as a cost-effective and environmentally friendly system for removal of nitrogen from wastewater (Kartal *et al.*, 2013).

1.1.2.3.2. Filamentous bacterial growth associated with activated sludge bulking and foaming problems

In activated sludge systems, selective pressures favour floc-forming organisms. Flocs are complex heterogeneous structures composed of aggregates of bacteria and other organisms embedded in a polymeric matrix, often referred to as extracellular polymeric substances (EPS; Bura *et al.*, 1998). Particulate organic and inorganic matter may also be absorbed onto such flocs. Most bacteria exist as microcolonies or microflocs bound to the floc structure, although some occur as filamentous bacteria (Nielsen *et al.*, 2003). The floc matrix comprises high hydrolytic enzymatic activity which enhances the degradation of carbohydrates and lipids before assimilation by bacterial cells (Geesey *et al.*, 2004). Importantly, floc formation is essential to guarantee good settling proprieties of activated sludge, which is crucial for the separation of sludge from treated wastewater. It is this property that actually selects for floc-forming organisms in the activated sludge process.

Bulking and foaming refers to poor settling sludges which themselves contain excess amounts of filamentous bacteria with a hydrophobic cell surface and which tend to float to the surface of activated sludge tanks. Importantly, dominance of such filamentous bacteria will lead to operational difficulties, such as accumulation of significant amounts of biomass as foam that does not play an active role in the treatment process and, thereby, leads to a deterioration of the final effluent quality

(Martins *et al.*, 2004). Molecular phylogenetic surveys have enabled the identification of several filamentous bacteria in activated sludge systems. These include, for example, *Thiothrix* spp. (Wagner *et al.*, 1994), *Norcodia* spp. (Soddell & Seviour, 1994) and *Microthrix* spp. (Erhart *et al.*, 1997).

The advent of molecular tools, in combination with fluorescence *in situ* hybridisation (FISH) allowed the quantification of previously unknown filamentous bacteria (*Chloroflexi* spp.; Björnsson *et al.*, 2002). However, until recently the presence and abundance of filamentous organisms did not reveal many correlations with a specific plant designs or process parameters (Mielczarek *et al.*, 2012).

To investigate further the physiological characteristics, which will determine the ecology and competitive behaviour of filamentous bacteria, i.e. their ecophysiology, efforts have been made to focus on addressing the *in situ* hydrolysis, assimilation and degradation of organic compounds by different filamentous species. The development of the microautoradiography and fluorescence *in situ* hybridization (MAR-FISH) method, enabled detailing of the ecophysiological characterisation of *Thiothrix* spp. (Nielsen *et al.*, 2000), *Candidatus* Microthrix parvicella (Nielsen *et al.,* 2002), *Meganema peroderoedes* (Kragelund *et al.*, 2005) and *Candidatus* Epiflobacter spp. (Xia *et al.*, 2008). Despite the existence of a clear relationship between filamentous organisms and bulking and foaming, the ecological factors responsible for controlling individual population densities and the reasons for certain filamentous bacteria to reach excessive abundances remains unclear (Nielsen *et al.*, 2009).

1.1.3. The microbiology of lipids accumulating bacteria in activated sludge

Due to their hydrophobic nature, lipids within wastewater are typically absorbed onto particles and promote cell adhesion within flocs. However within anaerobic/aerobic sludge cycling, specialised filamentous lipid accumulating organisms (LAOs), e.g. '*Candidatus* Microthrix parvicella' (henceforth, referred to as *Microthrix parvicella*), export highly efficient extracellular lipases and take up and store resulting LCFAs as neutral lipids during the anaerobic treatment phases (Nielsen *et al.*, 2002). These characteristics make such organisms very interesting for potential lipid-based energy-recovery strategies from wastewater.

1.1.3.1. *Candidatus* Microthrix parvicella

Microthrix parvicella is a Gram-positive filamentous *Actinobacterium* (Slijkhuis, 1983) which causes severe problems with foaming and bulking in activated sludge treatment plants with nutrient removal (Nielsen *et al.*, 2002). Consequently, to devise control strategies aimed at limiting the proliferation of *Microthrix parvicella*, substantial efforts have been directed towards gaining more information about the identity, physiology and ecology of this filamentous organism in activated sludge.

1.1.3.1.1. Characteristics and isolation

Microthrix parvicella was first isolated in pure culture and characterised by microscopic examination of bulking activated sludge samples, in 1973 (Van Veen, 1973). Isolation was performed on I-medium, with glucose as carbon source. Van Veen carefully described the filamentous organism (see **Table 1.2** for the main characteristics of *Microthrix parvicella*). However, since the early work of Van Veen

several difficulties concerning the isolation and the preservation of isolates have limited extensive physiological investigations (Blackall *et al.*, 1995).

Table 1.2 Main physiological characteristics of *Microthrix parvicella* in pure culture.

	Characteristics
Morphology	Non-branched filaments
Cell length	< 200 µm
Cell diameter	0.5 µm
Mobility	Immobile
Cell wall structure	Gram-positive
Respiration	Facultative anaerobe
Physical growth	Mesophile, Neutrophile

Microthix parvicella is described as a slow growing bacterium with unclear growth and nutritional requirements. The first tentative to characterise its carbon substrates was reported by Slijkhuis in 1983. In contrast to the first isolation realized by Van Veen, the isolates obtained by Slijkhuis did not grow on simple substrates such as glucose, but required oleic acid as a carbon and energy source (Dutch isolate; Slijkhuis, 1983). Later, additional strains of *Microthrix parvicella* were cultured using other complex culture media such as R2A (Seviour *et al.*, 1994; Seviour *et al.*, 1984), RN1 (Rossetti *et al.*, 2005), modified NTM medium (strain DAN1-3; Blackall *et al.*, 1995), MSV based medium (strains BIO17; Levantesi *et al.*, 2006), or GS medium (strains EU; Levantesi *et al.*, 2006). However, none of the later obtained *Microthrix parvicella* strains actively grow on medium using the Slijkuis formulation. These results suggest that considerable physiological differences exist among these strains

and the original "Dutch isolate". The difference in substrate utilisation may be explained with respect to the different culture environments. In pure culture, isolates can grow in a non-competitive environment and can use a range of simple carbon sources, in contrast to activated sludge plants where substrates are clearly more complex (Rossetti *et al.*, 2005) and much more specialised carbon utilisation profiles exist (Nielsen & Vidal, 2010). **Table 1.3** details the different media used thus far for the isolation and cultivation of *Microthrix parvicella*.

Table 1.3 Growth media and methods used for the isolation of *Microthrix parvicella*.

Strains	Medium	Carbon source	Nitrogen source	Sulphur source	Growth factor	Temparture (°C) / pH	Ref.
"Dutch"	I	Glucose	Ammonium sulphate	Sulphate	Vitamin B12 + thiamine	17-20 / 7.0	(Van Veen, 1973)
Slijkhuis	H	Sludge hydrolysate	/	/	Vitamin complex	25 / 8.0	(Slijkhuis *et al.,* 1983)
RN1	R2A	Protease peptone, starch, glucose, sodium pyruvate	Organic	Organic sulfur/ sulfate	Yeast extract	20-22 / 7.2	(Seviour *et al.*, 1994)
DAN1-3	NTM	Succinate + peptone	Ammonium + sulphate	Organic sulphur	Vitamin B12 + thiamine	20-22/ 8.0	(Blackall *et al.*, 1995)
BIO17	MSV	Sodium lactate + peptone	Ammonium sulphate	Thiosulphate	Vitamin mixture	20 / 7.0	(Levantesi *et al.*, 2006)

1.1.3.1.2. Taxonomy

Microthrix parvicella (strain DAN1-3) was first reported as a novel deeply branching member of the *Actinobacteria* based on its 16S rRNA gene (Blackall *et al.*, 1995). A similar phylogenetic position was found for the "Italian isolate RN1" (Rossetti *et al.*, 1997) as well as for the strains BIO17 and EU (Levantesi *et al.*, 2006). However, *Microthrix parvicella* is not a taxonomically valid name (McIlroy *et al.*, 2013). Despite many attempts during the last forty years, conflicting data have been obtained concerning the phenotypic traits and the physiology of *Microthrix parvicella*, as show in **Tables 1.2** and **1.3**, justifying the 'Candidatus' status for all *Microthrix* spp. isolates (Blackall *et al.*, 1996). Another species, "*Candidatus* Microthrix calida" (strains TNO 1-2), has been isolated from an industrial activated sludge sample (Levantesi *et al.*, 2006). 16S rRNA gene sequence analysis revealed maximal 96.7 % sequence similarity between both species (Levantesi *et al.*, 2006). Current evidence based on few studies suggests that "*Candidatus* Microthrix calida" is not common and probably plays no role in bulking incidents (Nielsen *et al.*, 2009). **Figure 1.3** show recent published phylogenic tree of publicly available 16S rRNA gene sequence of *Microthrix parvicella*.

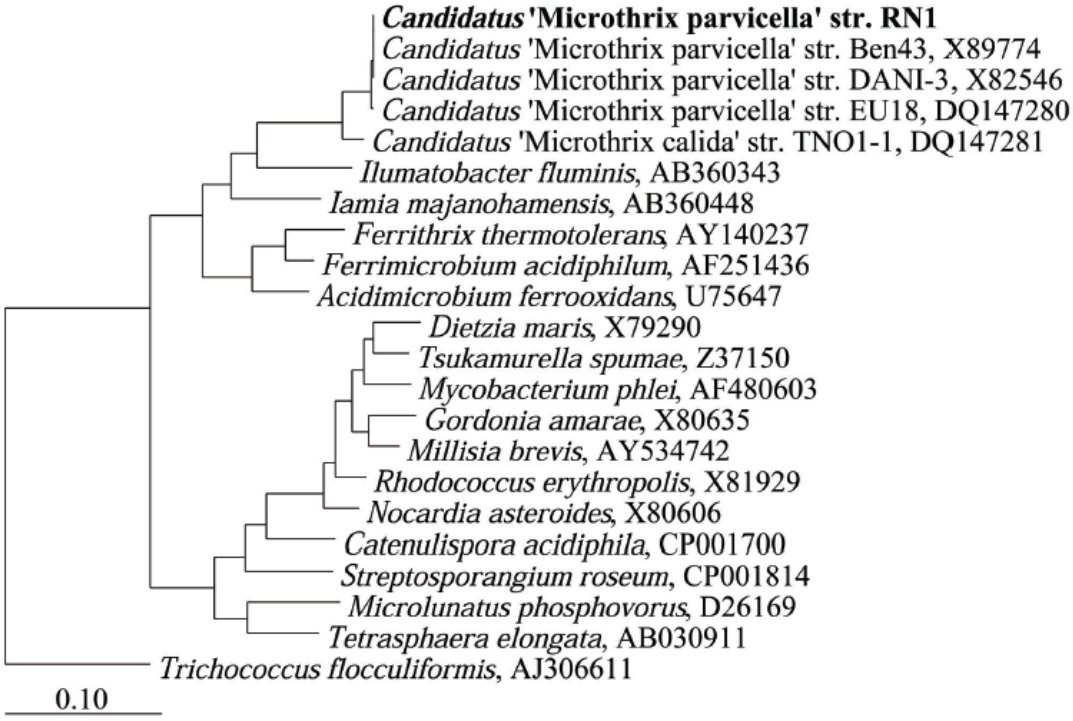

Figure 1.3 Phylogenetic tree showing publicly available 16S rRNA gene sequences of *Microthrix* spp. isolates. Reproduced from McIlroy *et al.*, 2013.

1.1.3.1.3. Basic metabolism

The pre-requisites for efficient control measures against *Microthrix* spp. have motivated *in situ* studies aimed at the detection and quantification of *Microthrix parvicella* and its close relatives. However, inconsistencies, controversy and sometimes contradictory data regarding important aspects of *Microthrix parvicella*'s metabolism, mainly regarding organic substrate catabolism, have been reported between pure culture studies and culture-independent techniques (McIlroy *et al.*, 2013; Noutsopoulos *et al.*, 2012). It is well established that in pure cultures, *Microthrix parvicella* is a chemoorganotroph with a slow growth rate, microaerophile with low temperature tolerance. Pure culture studies have revealed an optimal growth rate at 22 °C, but very little growth in the temperature range of 25-30 °C. However,

considerable growth has been observed at 7 °C in both cultures and full-scale plants (**Tables 1.2&1.3**; Rossetti *et al.*, 2005). This adaptability of *Microthrix parvicella* provides a significant competitive advantage during cold temperatures. Significant seasonal variations in the abundance of *Microthrix parvicella* have been reported from different WWTPs locations, e.g. in Italy (Miana *et al.*, 2002), Czech Republic (Wanner *et al.*, 1998), South Africa (Lacko *et al.*, 1999), The Netherlands (Slijkhuis & Deinema, 1988), China (Xie *et al.*, 2007) and Denmark (Kristensen *et al.*, 1994). The preferential growth of *Microthrix parvicella* at low temperatures has been suggested to be related to the reduction of lipid solubility in wastewater thereby preventing their uptake by non-specialised bacteria (Hug *et al.*, 2006). Physiological characterisations of different strains (Tandoi *et al.,* 1998) as well as recent genome sequences (Muller *et al.,* 2012) suggest that *Microthrix parvicella* can reduce nitrate to nitrite but appears to lack nitrite reductase for further reducing the produced nitrite.

1.1.3.2. Lipid metabolism in *Microthrix parvicella*

Early pure culture studies on the physiological characteristics of *Microthrix parvicella* indicated exclusive utilisation of LCFAs as carbon and energy sources (Slijkhuis *et al.*, 1984). These characteristics have been confirmed under *in situ* conditions using MAR-FISH, demonstrating that LCFAs such as oleic and palmitic acids are taken up by *Microthrix parvicella* under anaerobic, anoxic and aerobic conditions (Andreasen & Nielsen, 1998). Additionally, microautoradiography has shown that oleic acid is stored mainly as neutral lipids under anaerobic conditions, whereas conversion to membrane phospholipids occurs almost exclusively under aerobic conditions, thereby indicating growth (Hesselsoe *et al.*, 2005; Nielsen *et al.*, 2002). Extraction and identification of the isotopically labelled lipids by gas chromatography (GC) showed

that the resulting storage compounds were probably TAGs (Rossetti *et al.*, 2005). The combination of FISH with hydrophobic fluorescent microsphere adhesion cell surface assays demonstrated that *Microthrix parvicella* filaments are more hydrophobic than most other activated sludge bacteria (Nielsen *et al.*, 2010). This cell surface hydrophobicity can attract lipids, LCFAs and other non-polar substrates thereby providing a competitive advantage in substrate uptake (Nielsen *et al.*, 2002). Additional investigations using enzyme-linked fluorescence assays have indicated that *Microthrix parvicella* possesses lipase activity on its cell surface (Nielsen *et al.*, 2002). The presence of these extracellular enzymes enhances its ability to take up lysis products and LCFAs derived from TAGs bound to the cell surface (Nielsen *et al.*, 2002). It has been hypothesised that these two characteristics provide *Microthrix parvicella* with the ability to out-compete other bacteria (e.g. floc-formers) in the uptake and storage of lipids in activated sludge systems operating with alternating anaerobic-aerobic zones (Nielsen *et al.*, 2002). This hypothesis has been confirmed by measuring the uptake and storage capacity of oleic acid (C18:1) by *Microthrix parvicella* under anoxic conditions compared to typical floc-forming microorganisms (Noutsopoulos *et al.*, 2012). Additionally, further recent investigations on EBPR systems showed that *Microthrix parvicella* is very stable in its substrate uptake profile (Kindaichi *et al.*, 2013). **Figure 1.4** provides a summary of LCFA utilisation by LAOs, such as *Microthrix parvicella*, under anaerobic and aerobic conditions in EBPR systems.

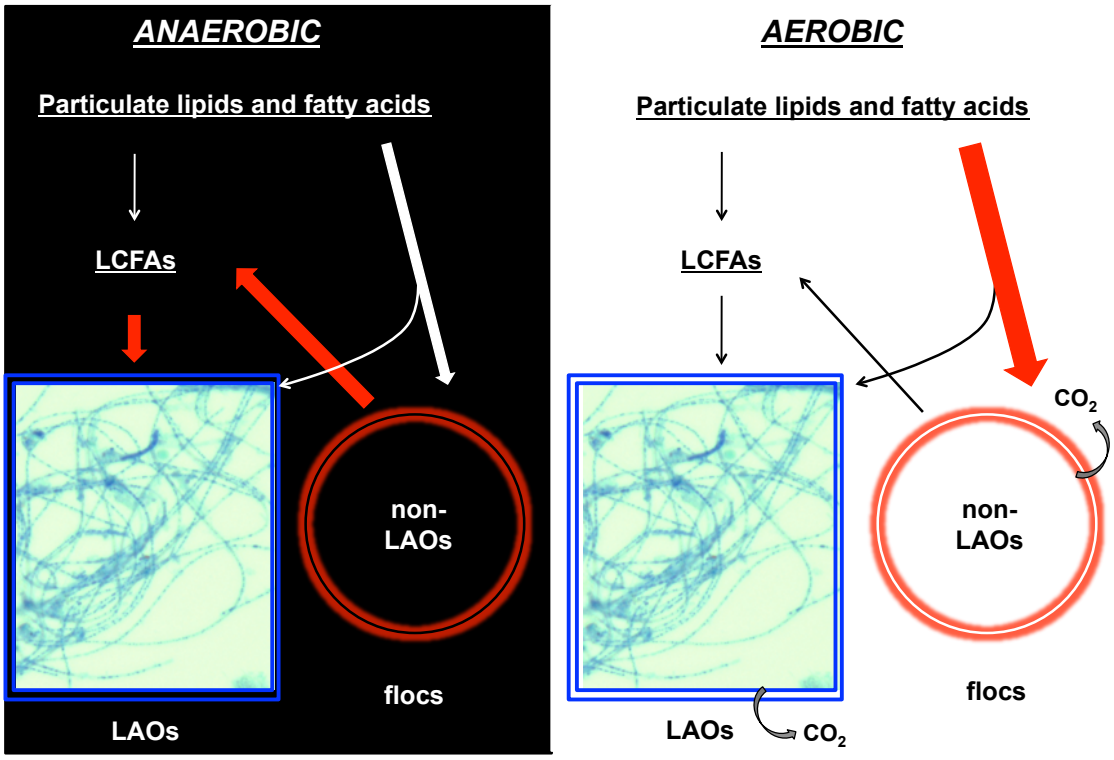

Figure 1.4 Model of LCFA utilisation by LAOs in alternating anaerobic–aerobic activated sludge plants. LCFAs are preferentially accumulated by LAOs during anaerobic conditions whereas during aerobic conditions floc-forming organisms are able to assimilate these better. Modified from Rossetti *et al.*, 2005.

A first draft genome sequence of *Microthrix parvicella* (strain BIO17-1, Muller *et al.*, 2012) confirmed its ability to process and accumulate excessive amounts of fatty acids. The genome encodes multiple copies of two key enzymes involved in fatty acid metabolism: 28 homologues of long-chain fatty acid-acyl coenzyme A (LCFA acyl-CoA) ligase and 17 homologues of enoyl-CoA hydratase. Additionally, in a second draft genome sequence of *Microthrix parvicella* (strain RN1, McIlroy *et al.*, 2013) at least 10 putative copies of the wax ester synthase/acyl coenzyme A:diacylglycerol acyltransferase (WS/DGAT) were identified. This enzyme catalyses the final step of TAG biosynthesis in Gram-negative bacteria, e.g. in the genus *Acinetobacter* (strain ADP1; Stöveken *et al.*, 2005). However, a candidate gene coding for the prior

reaction, i.e. phosphatidic acid phosphatase (PAP), within the TAG synthesis has yet to be identified (McIlroy *et al.*, 2013).

LAOs appear to employ a similar strategy to PAOs and their competitors, i.e. GAO (Seviour *et al.*, 2000) in taking up and storing organic matter under anaerobic conditions and, subsequently, using the storage compounds for growth with nitrate or oxygen as electron acceptors (Nielsen *et al.*, 2002). Recent draft genome sequences of *Microthrix parvicella* confirm an active accumulation of polyphosphate (Muller *et al.,* 2012). Investigations on EBPR systems have shown that *Microthrix parvicella* is able to utilise glycerol only when oleic acid is present as a co-substrate (Kindaichi *et al.*, 2013). As *Microthrix parvicella* produces extracellular lipases that can degrade TAGs into LCFAs and glycerol, it seems plausible that they can consume both types of substrates; a putative transporter gene for glycerol has been annotated in the RN1 draft genome, but the exact transport mechanism for LCFAs in Gram-positive bacteria is still unclear (McIlroy *et al.*, 2013).

1.1.4. Lipid biochemistry

Many organisms synthesise neutral lipids as an integral part of their metabolism and as energy storage compounds (Alvarez & Steinbüchel, 2002). Particularly, TAGs are convenient storage compounds for carbon and energy, because they are less oxidised and posess a higher calorific value than carbohydrates or proteins (Section 1.1.1.1; Alvarez & Steinbüchel, 2002) . Consequently, lipid accumulation is an advantageous strategy for survival in natural habitats with fluctuating nutrient availabilities and the capacity for lipid storage most likely provided a strong selective advantage during evolution. Occurrence of TAG storage is widespread in many eukaryotes (Zweytick *et*

al., 2000). In contrast, most prokaryotes synthetise polymers such as poly-3-hydroxybutyrate (PHB) or other polyhydroxyalkanoates (PHAs; Wältermann & Steinbüchel, 2005). The accumulation of TAG or wax esters (WE) in intracellular lipid-bodies is a virtue of only a few prokaryotes (Wältermann & Steinbüchel, 2005).

1.1.4.1. LCFA uptake and fatty acid β-oxidation

The genetic and biochemical details of LCFA transport were first elucidated in the Gram-negative bacterium *Eschericha coli* (Fujita *et al.*, 2007). The hallmark of the uptake is the coupled transport and activation of exogenous LCFAs (Fujita *et al.*, 2007). The transport/acyl-activation mechanism of LCFAs across the cell envelope of *E. coli* requires both the outer membrane transport protein FadL (encoded by the *fad*L gene) and the membrane-associated acyl-CoA synthase, FadD (encoded by the *fadD* gene). The product of transport, long chain acyl-CoA, has a regulatory role and controls on the DNA binding activity of the transcription factor FadR (encoded by the *fadR* gene). This controls the expression of nine genes primarily involved in fatty acid β-oxidation (**Figure 1.5**, Fujita *et al.*, 2007). Fatty acid β-oxidation generates acetyl-CoA, which enters the citric acid cycle or TAG biosynthesis, and NADH as well as $FADH_2$ which are used by the electron transport chain (Röttig & Steinbüchel, 2013).

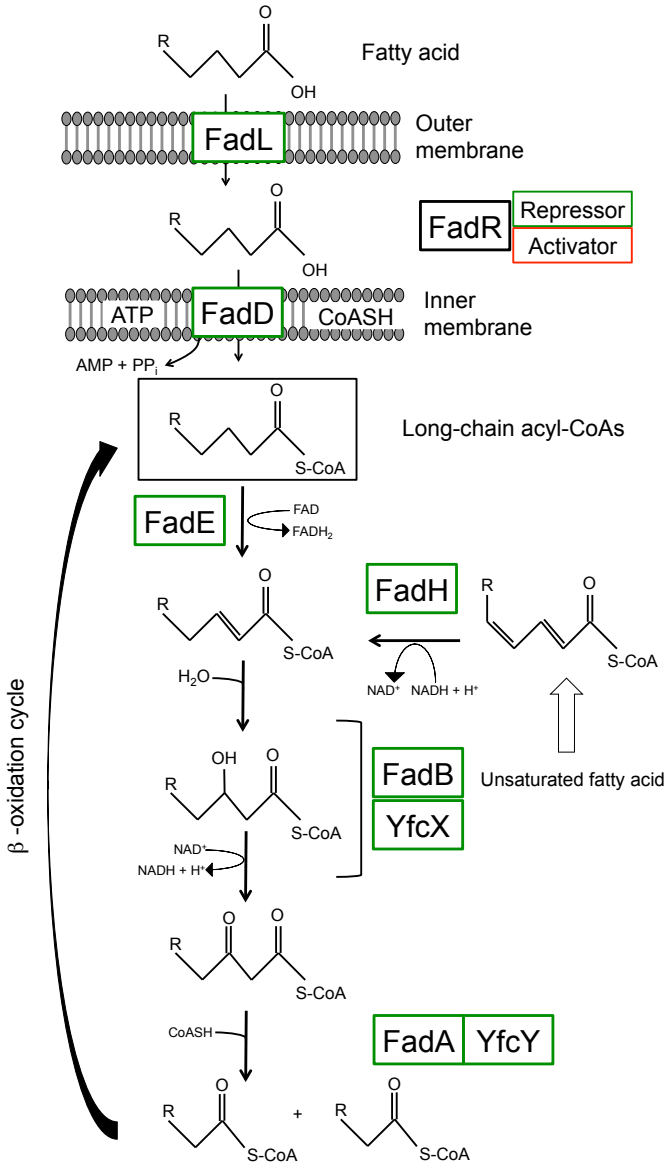

Figure 1.5 Fatty acid uptake and catabolism in Gram-negative bacterium *E. coli* and its regulation through the FadR protein. FadL: long chain fatty acid transporter, FadD: fatty acid CoA ligase, FadR: transcription factor of fatty acid degradation, FadE: acyl-CoA dehydrogenase, FadH: 2,4-dienoyl-CoA reductase, FadB: enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase, FadA: 3-ketoacyl-CoA thiolase, R represent alkyl groups. Reproduced from Fujita *et al.*, 2007.

1.1.4.2. Neutral lipid storage by LAOs

Lipid accumulation is, generally, a stress response mechanism during times of nutrient imbalance, particularly when nitrogen sources are scarce in the environment (Dean *et*

al., 2010). During these periods, carbon substrates are excessively assimilated by cells and are converted into storage TAGs. TAG accumulation has been reported for species mainly belonging to the *Actinomycetes*, particularly *Mycobacterium* spp. (Barksdale & Kim, 1977), *Norcardia* spp. (Alvarez *et al.*, 1997), *Rhodococcus* spp. (Alvarez *et al.*, 1996) and *Streptomycetes* spp. (Olukoshi & Packter, 1994), where TAGs can account up to 80 % of the total cellular dry matter (Alvarez *et al.*, 1996). TAGs are typically stored in spherical lipid bodies, with quantities and diameters depending on the microbial species, growth stage, and cultivation conditions (Wältermann & Steinbüchel, 2005). TAG accumulation has been described to occur in *Acinetobacter* spp., but at one order of magnitude lower than the amount of stored WEs (Kalscheuer & Steinbüchel, 2003). In the Gram-positive bacterium *Rhodococcus opacus* PD630, TAGs and waxes are reported to be accumulated as cytoplasmic inclusions surrounded by a thin boundary layer (Alvarez *et al.*, 1996). Strain PD630 has been considered as an excellent model to study basic aspects of TAG biosynthesis, accumulation and mobilisation in prokaryotes (Alvarez & Steinbüchel, 2002). TAG biosynthesis in *Rhodococci* has been proposed to occur via sequential acyl-CoA-dependent reactions referred to as the "Kennedy pathway", described in yeast and plant model systems (Wältermann *et al.*, 2007; **Figure 1.6**).

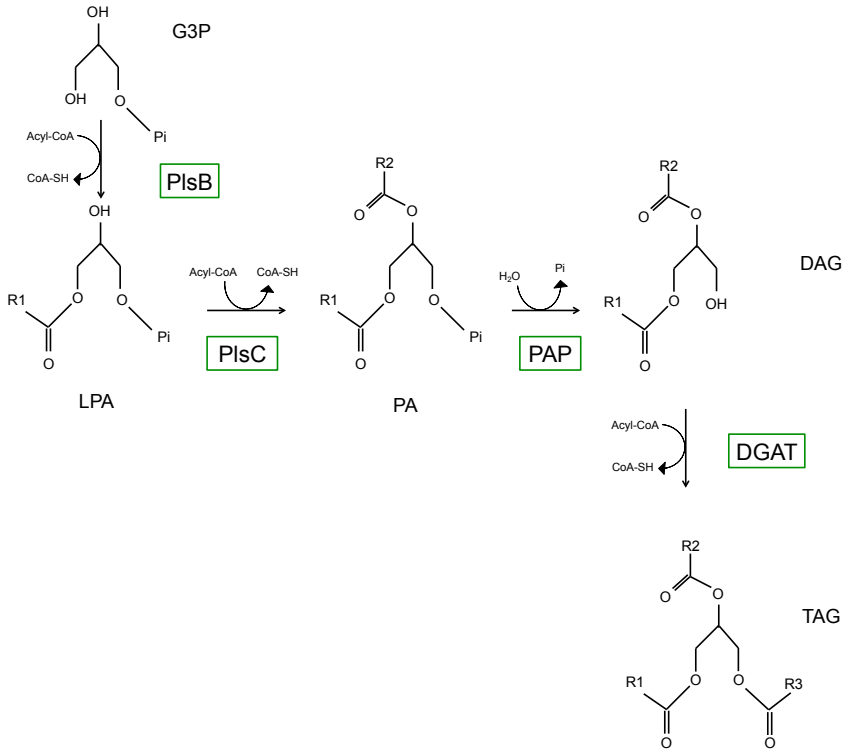

Figure 1.6 TAG synthesis in *Acinetobacter baylyi* spp. G3P: glycerol-3-phosphate, PlsB: acyltransferase, LPA: lysophosphatidic acid, PlsC: lysophosphatidate acyltransferase, PA: phosphatidic acid, PAP: phosphatidic acid phosphatase, DAG: diacylglycerol, DGAT: diacylglycerol acyltransferase. R1, R2 and R3 represent different alkyl groups. Reproduced from Wältermann *et al.*, 2007.

The Kennedy pathway involves the *de novo* synthesis of diacylglycerols (DAGs) by the sequential acylation of glycerol-3-phosphate. The glycerol-3-phosphate acyltransferase gene (*plsB*) catalyzes the first step in DAG synthesis by using a fatty acyl-CoA as substrate, generating lysophosphatidic acid (LPA). The second step involves the acylation of the LPA by the lysophosphatidate acyltransferase (gene *plsC*) yielding phosphatidic acid. The removal of the phosphate group is catalysed by phosphatidic acid phosphatase, prior to a final acylation step. In the third acylation reaction, a fatty acyl-CoA is transferred to the vacant position of DAG, by a diacylglycerol acyltransferase (DGAT or gene *atf*), which represents the final step of

TAG biosynthesis. The third acylation step is a unique enzymatic reaction of TAG biosynthesis as the two first acylation steps are also used for the synthesis of membrane phospholipids. Thus, DGAT is considered to be the key gene in TAG synthesis (Alvarez *et al.*, 2013; Wältermann *et al.*, 2007).

1.2. A microbial community systems biology approach

For many years the microbial world has been studied using methods based on the isolation and cultivation of microorganisms. However, classical culturing techniques have only succeeded in characterising a small group of species (Amann & Ludwig, 2000). Although these culturable microorgansims only represent a small proportion of the microorganisms present in all natural environments, they represent the current basis of the majority of our knowledge about microbial biochemistry (Staley & Konopka, 1985). Progress in molecular biology has made it possible to explore the enormous range of microbial diversity in all natural habitats. These molecular studies, initially based on the sequence of 16S ribosomal RNA genes (Pace, 1997), have revealed the existence of numerous unknown microbial species (bacteria and archaea) of which only a small fraction had been obtained in pure culture. Importantly, the sequencing of 16S rRNA genes has allowed the revelation of phylogenetic relationships between the different species which have been resolved in this way leading to the construction of detailed tree of life (Woese *et al.*, 1990). The direct application of molecular tools such as 16S rRNA sequencing to the study of complex

microbial communities has resulted in a paradigm shift in our understanding of the activities of particular microbial groups in their natural habitat (Head *et al.*, 1998). However, this approach is limited for the in-depth exploration of functional microbial diversity and, thus, community genomics (metagenomics) aims at gaining access to the physiology and genetics of these uncultured organisms through direct characterization of their genomic DNA *in situ* (Handelsman, 2004). The ability for comprehensive sequencing of DNA extracted from microbial communities now forms the foundation for post-genomic approaches which provide detailed functional insight into entire microbial ecosystem function (Raes & Bork, 2008).

1.2.1. Eco-systems biology

1.2.1.1. Systems biology applied to microbial ecology

Eco-Systems Biology is the scientific discipline devoted to the study of the dynamic processes and interactions in an ecosystem, consisting of organisms in a particular habitat and its non-living components, with the explicit aim to model, predict and control the behaviour of the system as a whole (Muller *et al.*, 2013). The term was first coined by Azam and Worden in 2004, in the context of marine ecosystems analysis (Azam & Worden, 2004).

Systems biology investigates the structure and dynamics of cellular and organismal function, rather than the characteristics of isolated parts of a cell or organism to understand their biology at the system level (Kitano, 2002). Thus, biology is advancing from a reductionist approach, which aims at identifying and describing

individual cellular processes, to a holistic approach, which aims to investigate cellular component interactions in their entirety (Sauer *et al.*, 2007).

The systems biology approach, driven by genome sequencing and the generation of high-throughput functional "omic" data, is revolutionising organismal biology (Joyce & Palsson, 2006). These advances in turn are transforming microbial ecology from a descriptive to a quantitative predictive science (McMahon *et al.*, 2007). Through the combined application of community genomics, transcriptomics, proteomics and metabolomics, the field of microbial ecology is currently entering the era of systems biology, which aims to obtain a comprehensive picture of microbial organismal and functional diversity (Raes & Bork, 2008).

1.2.1.2. Multi-omic measurements

Current rapid advances in microbial ecology are driven by the generation of high-resolution systems-level molecular data derived from genomic, transcriptomic, proteomic and metabolomic analyses. The various omic technologies are enabling important progress in our ability to link microbial communities' composition and function. However, no single omic analysis can fully unravel the complexities of fundamental microbial biology (Zhang *et al.*, 2010). While metagenomic data provides gene inventories without any proof of their functionality (Röling *et al.*, 2010), the analysis of community transcripts enables a first assessment of community-wide functions (Helbling *et al.*, 2012), and community proteomics provides a high-resolution representation of the actual phenotypic traits of individual community members (Wilmes & Bond, 2009a). Metagenomic information forms the backbone for the proper interpretation of transcriptomic and proteomic data leading to population-level resolution of individual functions. Recent advances in functional annotation of

metagenomic data and computational processing of inferred metabolic networks enable linkages to be established between the metagenome and the metabolome (Larsen *et al.*, 2011). However, such metagenome-based projections have yet to be validated experimentally by omic analyses in particular metabolomics. Consequently, ideally metagenomic data have to be supplemented by metatranscriptomic, metaproteomic and metabolomics analyses to allow meaningful resolution of community- and population-level roles spanning genetic potential to final phenotype (Muller *et al.*, 2013).

1.2.1.3. Systematic measurements, data integration, analysis and modelling

A meta-omic study aims at identifying a panel of microbial organisms, genes, variants, pathways, or metabolic functions characterising a given microbial community (Segata *et al.*, 2013). However, to be successful, such an approach needs to fulfil the promise of systematic measurement in order to facilitate truthful omic data processing, integration and analysis (Muller *et al.*, 2013). By systematic measurements is meant the standardised, reproducible and simultaneous measurement of multiple features from a single sample (Kitano, 2002). Resulting datasets should be fully integrable and will relate system-wide behaviours (Muller *et al.*, 2013). As such, spatially- and temporally-resolved eco-systematic measurements might provide new insights to improve the understanding of microbial communities (Muller *et al.,* 2013).

1.2.2. Analytical methods for molecular eco-systems biology

1.2.2.1. Metagenomics

In addition to their significant impact on our understanding of human biology and of other organisms, DNA sequencing technologies have enabled new areas of genomics including metagenomics. Metagenomics may be defined as the functional and sequence-based analysis of collective microbial genomes in their natural habitat (Tringe *et al.*, 2005). In the past decade, the application of metagenomics has unravelled novel enzymes and organisms which are biomarkers or drivers of processes relevant to disease, industry and environment (Knight *et al.*, 2012). The field of metagenomics initially started with the cloning of environmental DNA followed by functional expression screening (Handelsman *et al.*, 1998). This functional metagenomic approach allowed, among others, the remarkable discovery of a unique type of photochemistry in marine bacteria (Beja *et al.*, 2000). However, the power of metagenomics to provide a general overview of genetic potential within microbial communities was demonstrated by two landmark studies in 2004 using sequencing-based (shotgun) metagenomics. Venter *et al.,* (2004) exposed the magnitude of microbial organismal and functional diversity in the surface water of the Sargasso Sea, whereas, Tyson *et al.* (2004) demonstrated the possibility to reconstruct genomes of previously uncultured bacteria from a sample with relatively low microbial community diversity. In past years, metagenomics has been successfully applied to different microbial habitats such as agricultural soil (Tringe *et al.*, 2005), the oceans (DeLong *et al.*, 2006), the human distal gut (Gill *et al.*, 2006) and EBPR sludge communities (García-Martín *et al.*, 2006). For metagenomic sequencing, extracted and purified DNA is subjected to library preparation (either single-end or

paired-end librairies may be synthetised) before being subjected to DNA sequencing on a variety of platforms (Cho & Blaser, 2012). Advances in metagenomics have been primarily driven by the substantially reduced cost in next-generation sequencing (NGS) which has greatly accelerated the development of sequence-based metagenomics and it is now becoming a standard tool for studies in microbial ecology. A predominant current NGS technology involves the use of Illumina sequencing technology (Caporaso *et al.,* 2012). To date, 1,962 metagenomes are publicly available on the Integrated Microbial Genome with Microbiome sample (IMG/M; Markowitz *et al.*, 2012) system.

1.2.2.2. Metatranscriptomics

While the discipline of metagenomics allows investigation of the genomic potential of a particular microbial community, metatranscriptomics allows the analysis of transcript levels within a microbial community at a certain point of time (Helbling *et al.*, 2012). Hence, metatranscriptomics is the technique that is typically employed to obtain gene expression profiles under certain environmental conditions (Warnecke & Hess, 2009). Initially, numerous microbial gene expression studies involved the use of microarrays (Bulow *et al.*, 2008; Dennis *et al.*, 2003). However, the main disadvantage of the array-based approach is that target gene sequences and corresponding high-quality annotations are required before a functional microarray chip can be manufactured. Thus, the routine use of functional microarrays technology in the context of microbial ecology is unrealistic since a new array would need to be designed and fabricated for each individual microbial community sample under study (Wagner *et al.*, 2007).

With the advancements of NGS, DNA sequencing methods provide a new method for metatranscriptomics and this involves retro-transcription of the mRNA pool and sequencing of the resulting complementary DNA (cDNA) libraries (RNA-Seq; Wang *et al.*, 2008). Within the first report of an NGS-based metatranscriptomic study, Leininger and colleagues reported that archaea are the predominant ammonia-oxidising in soil environments (Leininger *et al.*, 2006). Since then, this approach has been successfully applied to different microbial communities from varied habitats such as seawater (Frias-Lopez *et al.*, 2008; Gilbert *et al.*, 2008), human gut microbiota (Gosalbes *et al.*, 2011), photosynthetic microbial mats (Liu *et al.*, 2011), wastewater sludge (Yu & Zhang, 2012) and marine sponges (Radax *et al.*, 2012). These studies demonstrate the efficiency of RNA-Seq to provide affordable access to whole-metagenome expression profiling and inference of activity of entire microbial communities (Helbling *et al.*, 2012). As for metagenomic sequencing, Illumina-based sequencing is currently being used for the majority of metatranscriptomic investigations (Leimena *et al.,* 2013).

1.2.2.3. Metaproteomics

Despite the extensive catalogue of metagenomic information that can be *in silico* translated into a list of all possible potential protein products, details about overall community function still require the generation of qualitative and quantitative protein abundance data. Metaproteomics (or community proteomics) has been defined as the study of proteins expressed at a given time within a given microbial ecosystem (Wilmes & Bond, 2004). Community proteomics plays a keystone role in functional microbial community analysis as it provides a representation of the immediate catalytic potential within mixed microbial assemblages. Microbial metaproteomics

has been applied in diverse microbial habitats such as soil (Benndorf *et al.*, 2007), ocean (Sowell *et al.*, 2009), acid mine drainage (Ram *et al.*, 2005), wastewater sludge (Wilmes *et al.*, 2008), human gut (Verberkmoes *et al.*, 2009b) and freshwater (Ng *et al.*, 2010).

After protein extraction, denaturation and reduction, a metaproteomic investigation may involve multiple experimental approaches and measurement platforms. To reduce sample complexity, microbial community proteomics either use gel-based separation or gel-independent liquid chromatography (LC)-based separation, each relying on subsequent mass spectrometric (MS) analysis for the generation of peptide mass fingerprint data (VerBerkmoes *et al.*, 2009a). Each separation approach has its own limitations for comprehensive protein identifications and the integration of both methods is recognised as the more powerful for enlarge protein identification spectrum (Hettich *et al.*, 2013). The last analytical step in community proteomics consists of *in silico* matching of mass spectra to predicted amino acid sequences of annotated protein coding gene sequences derived from metagenomic or isolate genome data for protein identification.

1.2.2.4. Metabolomics

Metabolomics is defined as the comprehensive analysis of the complete set of all metabolites present in and around growing cells at a given time point (Fiehn, 2002). Since metabolites represent the molecular end point of gene expression and cell activity, metabolomics offers a holistic approach for understanding the phenotype of an organism and, thereby, occupies a central role in systems biology approaches. Unlike genomics, transcriptomics and proteomics in the field of functional genomic approaches, metabolomics more closely reflects the activity of cells at a functional

level (Garcia *et al.*, 2008). There is considerable interest in applying metabolomic methodologies to ecological questions, such as metabolic responses to physiological stress, metabolic profiling for ecological risk assessment, environmental monitoring or studying the interactions of gut microbes with the human host (Bundy *et al.*, 2009). Metabolomic measurements require multiple-step methods including, quenching of metabolism, metabolite extraction, derivatisation, metabolite detection by mass spectrometry (MS) and data analysis (Reaves & Rabinowitz, 2011). The application of metabolomics to microbial ecosystems is still in its infancy and the ability to identify and quantify the entire set of intracellular and extracellular metabolites in a natural habitat is challenging (Tang, 2011). In contrast to DNA, RNA and proteins, products generated by metabolic reactions are highly variable in their chemical structure and properties, which contribute to the complexity and resulting analytical challenges for obtaining comprehensive metabolomic overviews. To date, in microbial community metabolomics experiments, only a low fraction of detected metabolites can be identified (Tang, 2011). However, continuing improvements in analytical tools, availability of microbial genomic sequences advances in omics data integration and the development of new databases as well as new bioinformatics standards and models are rendering the application of metabolomics to microbial communities meaningful.

Owing to the varied physiochemical properties and concentrations of metabolites present in natural microbial communities, there is no method that can separate, detect and identify all known metabolites (Garcia *et al.*, 2008). MS is a popular tool, more sensitive than NMR (nuclear magnetic resonance), for the deep detection of metabolites (Viant & Sommer, 2013). MS requires chromatographic separation, typically gas chromatography (GC) and liquid chromatography (LC) to reduce the

complexity of the metabolite mixtures. Particularly, GC coupled to MS (GC-MS), is the most commonly used technique in metabolomics, including for environmental metabolomics (Viant & Sommer, 2013). The main advantages of GC-based separation techniques compared to other techniques include the high separation efficiency (i.e. ability to distinguish isomeric compounds with specific columns), robustness and low cost of the analysis (Mashego *et al.*, 2007). Additionally, the GC-based method yields extensive reproducibility, due to the standardised use of electron impact (EI) for mass spectrometry, which enables trustful metabolite identifications (Garcia *et al.*, 2008). A major drawback of the GC-based separation technique is that metabolites are required to be volatile in nature to be separated. Hence, the derivatisation of metabolites is necessary to increase their volatility prior to analysis and, thereby, restricts the analysis of metabolites smaller than 500 Da (Witting *et al.*, 2012). Other drawbacks include that thermolabile compounds can easily be degraded when exposed to the high temperatures in a GC oven and that metabolites can have diverse affinities with derivatising agents, which can lead to inaccurate quantification, unless standards and data correction strategies are used to remove biases. Consequently, LC is an interesting alternative to GC as LC exhibits the ability to cover a much greater number of metabolites, without prior derivatisation (i.e. not limited to volatile metabolites; Viant & Sommer, 2013). However, LC coupled to MS (LC-MS) has not yet reached a similar level of consistency across instrument types compared to GC-MS and, thus, currently lacks comprehensive databases for metabolite identifications. Nonetheless, targeted LC-MS analysis is the technique of choice for reliable quantification of known, pre-selected metabolites, e.g. lipids (Viant & Sommer, 2013).

1.3. Thesis objective and synopsis of chapters

The overall aim of the present PhD thesis was to obtain unprecedented insights through a molecular Eco-Systems Biology approach into lipid transformations within lipid accumulating organisms (LAOs) located at the air-water interface of anoxic activated sludge wastewater treatment tanks, in view of potentially enriching such communities and harnessing accumulated lipids for biofuel production in the future.

An initial survey was carried out and aimed at the identification of a biological WWTP, which showed the prevalence of LAOs throughout the year. Temporally resolved microbial community profiles and proteomic analyses revealed the existence of a seasonal shift related to dominance of specific LAOs and the corresponding community-wide LAO phenotype (Chapter 2). Winter communities also show a marked enrichment in *Microthrix parvicella* whereas autumn communities are dominated by organisms closely related to *Alkanindiges* spp. and *Acinetobacter* spp.

In order to elucidate mechanistic details of lipid uptake and accumulation by LAOs, part of this thesis involved the development of the required Eco-Systems Biology methodological framework. To account for the inherent heterogeneity of mixed microbial communities and to facilitate meaningful high-resolution molecular data integration, analysis and modelling, a workflow was designed which allowed the isolation of individual biomolecular fractions, comprising DNA, RNA, proteins and metabolites from single, undivided samples (Chapter 3). This represents an important prerequisite for carrying out systematic molecular measurements involving genomics, transcriptomics, proteomics and metabolomics of microbial consortia.

The generation of systematic molecular data using the developed framework allowed the reconstruction of community-level metabolic networks and models onto which gene copy numbers, transcript levels and protein frequency were mapped. Comparative network-level analyses between an LAO-enriched microbial community sampled in winter and one sampled in autumn revealed genes involved in ammonium oxidation and glycerolipid metabolism to play keystone roles in conferring community-wide phenotypes (Chapter 4).

CHAPTER 2: INITIAL CHARACTERISATION OF LIPID ACCUMULATING MICROBIAL COMMUNITIES IN A BIOLOGICAL WASTEWATER TREATMENT PLANT

2.1. Introduction

Within anaerobic/aerobic sludge cycling, lipid-accumulating organisms (LAOs), export highly efficient extracellular lipases, and take up and store resulting long chain fatty acids (LCFAs) as neutral lipids during the anaerobic phase (Section 1.1.3.2; Nielsen al., 2002). Many prokaryotes are able to accumulate large amounts of lipophilic compounds as inclusion bodies in the cytoplasm (Section 1.1.4). Members of most genera synthesize storage polymers such as polyhydroxybutyrates (PHB) or other polyhydroxyalkanoate (PHA), whereas accumulation of triacyglycerides (TAG) and wax esters (WE) in intracellular lipid-bodies is a property of only a few prokaryotes (Wältermann & Steinbüchel, 2005). PHA are polyesters of various hydroxyalkanoates monomers typically synthesised when essential nutrients such as nitrogen or phosphorus are available only in limiting concentrations in the presence of excess carbon sources (Wältermann & Steinbüchel, 2005). Accumulation of storage lipids, serving as endogenous carbon and energy sources during starvation periods, might be a potential adaptation mechanism for coping with nutrient limitations, which is a frequent stress factor challenging bacteria in their natural habitats (Kalscheuer *et al.*, 2007).

In activated sludge, the bacterial biomass has to cope with rapidly changing feast and famine conditions. Populations best able to deal with these will be those able to rapidly assimilate substrates when available and store them intracellularly as carbon and energy stores (Seviour *et al.*, 2008). In domestic wastewater, lipids represent 31 % of COD (Raunkjær *et al.*, 1994) and are dominated by specific LCFAs (Quéméneur & Marty, 1994). Furthermore, extracellular lipases, which catalyse initial lipid hydrolysis, appear to be predominant in specific organisms present in activated

sludge (e.g. *Microthrix parvicella*; Nielsen *et al.*, 2002). However, essential aspects of LAO-driven processes regarding lipid hydrolysis, assimilation and conversion still remain unclear (McIlroy *et al.*, 2013).

The description of the processes involved in conferring the community-wide lipid accumulating phenotype is a fundamental aspect for potentially harnessing these microbial communities for the recuperation of lipids from wastewater and production of biodiesel from the resulting lipid-rich feedstocks. However, in order to understand LAO-enriched microbial communities in intricate detail, it is essential to study them directly in their native enrichment using the whole panoply of high-resolution molecular techniques now available in molecular microbial ecology (Section 1.2). Consequently, an appropriate full-scale WWTP system, which presented the required characteristics to elucidate the details of the fate of lipids within LAO biomass had to be selected.

This chapter presents a comprehensive analysis aimed at the identification of a suitable full-scale WWTP with ample LAO biomass to allow their detailed study.

For the first part of this chapter, detailed analyses of LAO-enriched microbial communities were carried out using microscopic, chemical and molecular biology methodologies which were applied to microbial biomass sampled from the air-water interface of an anoxic activated sludge tank.

For the second part of this chapter, a spatially and temporally resolved deep 16S rRNA gene sequencing survey and a proteomic investigation were conducted to identify possible early linkages between LAO-enriched microbial communities composition and biochemical traits linked to lipid accumulation.

2.2. Materials and methods

2.2.1. Identification of a suitable biological WWTP

At the inception of this project, biological wastewater treatment plant (WWTP) operators in Luxembourg were contacted by telephone in order to enquire about foaming and bulking issues in their respective systems. Foaming and bulking problems are typically associated with high levels of lipid accumulating bacteria within the activated sludge biomass. From this initial survey, it became apparent that transitory foaming and bulking events occur at the WWTP located in Schifflange (Esch-sur-Alzette, Luxembourg; 49°30'48.29"N; 6°1'4.53"E, **Figure 2.1**). The Schifflange biological WWTP is equipped with a set of online multi parameter probes.

Figure 2.1 The Schifflange biological wastewater treatment plant. (**A**) Satellite image of the plant. (**B**) Aerial view of the individual treatment tanks. The aerobic and anoxic tanks are highlighted in red and blue, respectively. Scale bars are equivalent to (**A**) 50 m and (**B**) 5 m. The asterix indicates the anoxic tank from which the LAO biomass subject of the present study was sampled.

2.2.2. Sampling and sample processing

Sampling Floating LAO biomass was sampled from the air-water interface of the anoxic activated sludge tank at the Schifflange biological WWTP. For each sampling date, four distinct "islets" (defined herein as biological replicates; **Figure 2.2**) were sampled using a levy cane of 500 ml. Samples were collected in 50 ml sterile Falcon tubes and then immediately snap-frozen by immersion in liquid nitrogen on site and subsequently stored at -80 °C. Additionally, for each sampling date, a representative fresh LAO biomass sample was used for microscopy analysis.

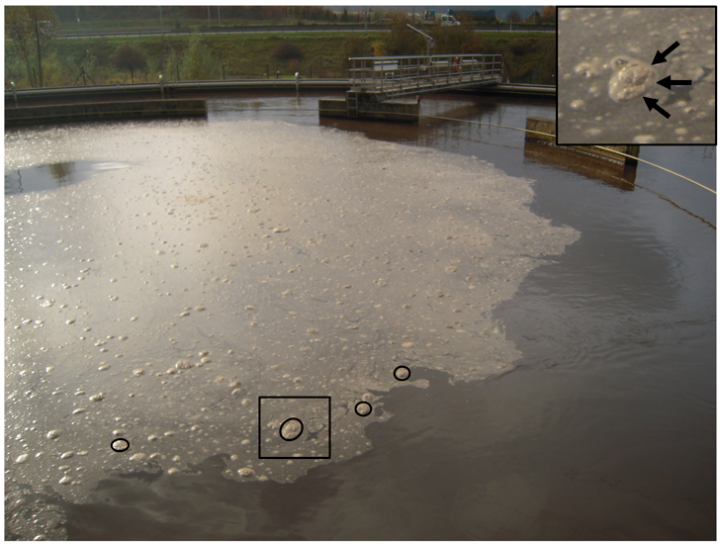

Figure 2.2 Photograph of LAO-enriched microbial communities located at the air-water interface of the anoxic tank (**Figure 2.1, B**) of the Schifflange WWTP. Individual islets (biological replicates, spatially resolved samples) are highlighted by black circles and a magnified area is provided to illustrate delineation of such islets (indicated by arrows).

In order to initially characterise the LAO-enriched microbial communities from the Schifflange WWTP by microscopic and phylogenetic analyses, 3 replicates of 300 mg of LAO-enriched microbial communities (defined herein as technical replicates) from a single biological replicate sample (islet), taken on 27 May 2010, were used. For the comparative phylogenetic and proteomics analyses, 200 mg of LAO-enriched microbial communities were obtained from 4 different islets (biological replicates) from 4 different dates, leading to an overall sample set of 16. These samples were taken on 4 October 2010 (date 1; D1), 25 October 2010 (date 2; D2), 25 January 2011 (date 3; D3) and 23 February 2011 (date 4; D4). This selection was chosen for having a seasonally distinct sample set reflecting a representative water temperature gradient (**Table 2.1).** Wastewater temperatures were measured within the anoxic tank at the time of sampling, using a portable HQ40d multi-parameter measurement device (Hach).

Table 2.1 Air and wastewater temperatures at the time of sampling at the Schifflange WWTP among the four selected sampling dates.

Dates	Abbreviation	Air temperature (°C)	Wastewater temperature (°C)
4 October 2010	D1	15	20.7
25 October 2010	D2	4	18.9
25 January 2011	D3	0	14.5
23 February 2011	D4	-6	13.9

Biomass fixation For each sampling date a representative fresh LAO sample was fixed in paraformaldehyde (PFA) and preserved at -20 °C (Bond *et al.*, 1999). The LAO sample was first centrifuged at 14,000 g for 2 min at 4 °C and the pellet was then resuspended in 250 µl of phosphate buffered saline 1x (PBS, pH 7.2) by vortexing. The 1x PBS solution was prepared with 8 g of NaCl, 200 mg of KCl, 1.44 g of Na_2HPO_4 and 240 mg of PO_4 dissolved within 800 ml of distilled water. The solution was then pH adjusted and sterilised by autoclaving. One volume of the solution was then mixed with three volumes of PFA solution (4 % w/v; Bond *et al*., 1999) and incubated at 4 °C for 3h. The PFA solution was prepared with 2 g of PFA and one drop of 2 M NaOH diluted into 33 ml of deionised water by heating and mixing at 60 °C under the hood for 2 min. The solution was then diluted into 16.5 ml of 3x PBS (pH 7.2) and conserve at 4 °C. The fixed cells were pelleted by centrifugation at 14,000 g for 2 min at 4 °C and washed three times with 250 µl of 1x PBS (pH 7.2) with intermediary centrifugation of 14,000 g for 2 min at 4 °C. The final pellet was re-

suspended in 1 volume of 1x PBS (pH 7.2) and one volume of ice-cold pure ethanol and conserved at -20 °C.

2.2.3. Epifluorescence microscopy

SYBR Green staining For cell visualisation, samples were stained with SYBR Green I 1/10,000 (v/v; Molecular Probes, Invitrogen) in anhydrous Dimethylsulfoxid solution (500 µl DMSO; Molecular Probes, Invitrogen) according to Patel and co-workers (2007). 1 ml of fixed LAO biomass suspension was transferred within a 2 ml reaction tube, 10 µl of a SYBR Green I solution was added to the suspension and mixed by vortexing. The SYBR Green I solution was prepared with 50 µl of SYBR Green I diluted into 450 µl of 1x Tris/EDTA buffer (TE, pH 7.4). The TE buffer was prepared with 10 ml of 1M Tris-HCl and 2 ml of 0.5 M EDTA (pH 8.0) mixed with 988 ml of deionised water. The solution was then pH adjusted and sterilised by autoclaving. 100 µl of stained sample was then diluted into 9.9 ml of 1x PBS buffer. 100 µl of the diluted sample was first filtered onto a 0.8 µm pore size, 25 mm diameter AA mixed-ester membrane filter (Millipore) using vacuum filtration. The resulting filtrate was then filtered onto a 0.02 µm pore size, 25 mm diameter Anodisc Al_2O_3 filter (Whatman) by an additional vacuum filtration run. The filter was then air-dried in the dark and cut in 1 cm^2 pieces. Two drops of mounted and observed oil were added on the filter piece. The membrane was then observed under 1,000X magnification with a Leica DMR Fluorescence Microscope coupled to a DFC500 camera (Leica).

Nile Blue A staining The samples were stained with Nile blue A (Sigma-Adrich) according to the protocol by Ostle & Holt (1982). 5 µl of PFA fixed LAO suspension was heat-fixed onto a microscopic slide and stained using an aqueous solution of 1 %

Nile Blue A in a water bath set to 55 °C for 10 min. This solution was prepared by mixing 1 g of Nile Blue A in 100 ml of deionised water. The microscope slide was washed with an 8 % acetic acid (v/v) solution at 55 °C for 10 min. The slide was then further rinsed with deionised water. A drop of water was added between the glass cover slips. The preparation was then observed at 1,000X magnification with a Leica DMR Fluorescence Microscope coupled to a DFC500 camera (Leica) at 460 nm with F543-CY3 filters.

Fluorescence in situ hybridisation (FISH) For cell permeabilisation, 1 µl of 1/10 diluted fixed LAOs suspension was spread onto a microscope slide. 3 µl of mutanolysin solution was added to the smear on slide and air-dried during 30 min. The mutanolysin solution (5,000 U/ml) was prepared with 1 mg of mutanolysin dissolved in 1 ml of 1x PBS (pH 7.2). The slide was washed twice in 50 ml of deionised water and finally in absolute ethanol. Hybridisation was performed in a hybridation chamber at 46 °C for 2 h using hybridisation buffer prepared with 40 µl of 1M Tris-HCl (pH 8.0), 2 µl of 10 % sodium dodecyl sulfate (SDS), 360 µl of 5 M NaCl, 700 µl of formamide diluted into 898 µl of deionised water (Bond & Rees, 1999). One volume of probe working solution (50 ng/µl) was added to four volumes of hybridisation buffer. The different oligonucleotide probes, used in an equimolar mixture (**Table 2.2**), were labelled with either fluorescein isothiocyanate (FITC) or indocarbocyanine (CY3). A piece of blotting paper was inserted into a 50 ml polyethylene tube and soaked with 1.8 ml of hybridisation buffer. 10 µl of hybridisation solution was added to each bacterial smear and the slide was placed into a 50 ml polyethylene tube (hybridisation chamber; in horizontal position). Following the slide was then immersed in wash buffer and incubated at 48 °C for 15 min. The washing solution was prepared by mixing 50 µl of 10 % SDS, 1 ml of 1M Tris-HCl

(pH 8.0), 700 µl of 5 M NaCl completed with deionised water to a quantity sufficient for 50 ml of solution and pre-heated in water bath at 48 °C. The hybridised slide was counter stained with DAPI mountant solution (1µg/ml), rinsed with deionised water, air-dried and was then observed at 1,000X magnification using a Leica DMR Fluorescence Microscope coupled to a DFC500 camera (Leica).

Table 2.2 Oligonucleotide probes used for fluorescence *in situ* hybridisation.

Probe	Sequence 5'-3'	Specificity	Labeling
EUB338 I – III[a]	I-GCTGCCTCCCGTAGGAGT II-GCAGCCACCCGTAGGTGT III-GCTGCCACCCGTAGGTGT	Bacteria	FITC, CY3
MPA60[b]	GGATGGCCGCGTTCGACT	*Microthrix parvicella* *Microthrix calida*	CY3

[a] Daims *et al.*, 1999
[b] Erhart *et al.*, 1997

2.2.4. Long chain fatty acid analysis

Long chain fatty acid analyses were carried out on 500 mg of frozen LAO biomass and 500 µl of wastewater.

Metabolite extractions For metabolite extractions, a method based on the combination of the selective extraction of polar metabolites by the use of methanol as polar solvent combined with the selective extraction of non-polar metabolites using hexane/isopropanol solution (2/1; v/v) as non-polar solvent was chosen. The frozen LAO sample was thawed on ice for 10 min. The cells were then pelleted by centrifugation at 4,500 rpm during 2 min at 4 °C. The resulting pellet was resuspended in 500 µl of 1x PBS (pH 7.2). Three sonication runs of 10 sec were

followed by a brief incubation on ice. 1 ml of pre-heated methanol (65 °C) was added to the lysate as well as to the effluent wastewater, the solutions were then homogenised by vortexing for 30 sec following a incubation at 65 °C for 3 min. The suspensions were centrifuged at 15,000 rpm for 5 min at 4 °C and the supernatants were then transferred into a new reaction tube. 500 µl of pre-heated isopropanol (65 °C) was then added to the supernatants and homogenised by a quick vortexing following by an incubation at 65 °C for 3 min. The suspensions were centrifuged at 15,000 rpm for 5 min at 4 °C and 750 µl of supernatant was transferred into a new reaction tube. The two supernatants of each suspension were then combined within the same reaction tube and dried using a speed-vacuum at 60 °C. The samples were resuspended into 1 ml of hexane using light sonication and heating in a sonication water bath for 30 min. 50 µl of sample were removed and transferred into GC vial. For derivatisation 100 µl of N,O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA) were added and the vials were heated at 50 °C for 30 min.

Gas chromatography–mass spectrometry (*GC-MS) analysis and metabolite identification* GC for derivatised samples was performed using an Agilent 6890 GC equipped with a 30 m Restek Rxi-5MS capillary column. The GC was connected to an Agilent 5973 Mass Selective Detector (MSD) operating under electron impact ionisation at 70 eV. The MS source was held at 230 °C and the quadrupole at 150 °C. The detector was operated in scan mode and 1 µl of derivatised sample was injected at 250 °C. Helium was used as carrier gas at a flow rate of 1.2 ml/min. The GC oven temperature was held at 80 °C for 2 min and increased to 320 °C at 8 °C/min. The total ion chromatograms were manually interpreted and the main metabolites identified based on retention time, mass spectra and use of the National Institute of Standards and Technology (NIST) mass spectral library.

2.2.5. Biomolecular extractions

DNA extraction for the sequencing of 16S rRNA by amplification and cloning DNA extraction was carried out using the PowerSoil DNA isolation kit (MO BIO Laboratories). Briefly, the protocol consist of cell lysis carried out by bead-beating the sample in a dedicated bead beating tube and lysis buffer, at 30 Hz for 5 min in a Retsch Mixer Mill MM 400 followed by removal by precipitation of diverse PCR inhibitors according to the manufacturer's instructions. Total DNA was captured on a silica membrane incorporated into a chromatographic spin column, washed and then eluted in the dedicated buffer. After extraction, 100 µl of eluted DNA were purified using the PowerClean DNA clean-up kit (MO BIO Laboratories) according to the manufacturer's instructions. For quality control, the isolated genomic DNA was separated by electrophoresis on a 1 % agarose gel containing 4 ‰ ethidium bromide (PlusOne Ethidium Bromide, GE Healthcare). For size estimation, the MassRuler DNA ladder mix (Fermentas) was loaded onto the gels. Agarose gels were visualised on an InGenius gel imaging and analysis system (Syngene).

DNA extraction for the sequencing of the 16S rRNA genes by pyrosequencing and protein extraction DNA and proteins extractions were performed on four biological replicates on each of the four sampling dates using a modified NorgenBiotek All-in-One Purification kit-based biomacromolecular isolation method (NA; Section 3.2.5). For each of the 16 DNA fractions, quality and quantity parameters were obtained using a NanoDrop ND-1000 spectrophotometer and separation on a 1 % agarose gel. Protein fractions were further analysed as described below (Section 2.2.8).

2.2.6. Amplification, cloning and sequencing of 16S rRNA genes

PCR amplification for cloning 16S rRNA genes were directly amplified from purified DNA extracts using universal bacterial primer pairs (Backer *et al*., 2003), in an equimolar mixture of seven 27f primers and a single 1492r primer (**Table 2.3**). PCR reactions for cloning were prepared following the cloning protocol instructions (QIAGEN PCR Cloning plus kit; Qiagen). After an initial 30 sec denaturation step at 95 °C, the DNA was amplified during 30 cycles of 20 sec at 95 °C, 15 sec at 44 °C and 1 min at 72 °C. The last cycle was followed by a final 5 min elongation step at 72 °C. Amplified 16S rRNA gene sequences were separated by electrophoresis on a 1 % agarose gel containing 0.5 ‰ ethidium bromide (Sigma). GeneRuler 100 bp Plus DNA ladder mix (Thermo Scientific) was loaded onto the gels for size estimation. Gels were visualised on an InGenius gel imaging and analysis system (Syngene).

Table 2.3 Oligonucleotide primers used for 16S rRNA amplification by PCR.

Primers	Sequence 5'-3'	Tm (°C)
27f*	YM1 -AGAGTTTGATCATGGCTCAG	54
	YM2 -AGAGTTTGATTATGGCTCAG	50
	YM3 -AGAGTTTGATCCTGGCTCAG	55
	YM4 -AGAGTTTGATTCTGGCTCAG	53
	Bif -AGGGTTCGATTCTGGCTCAG	60
	Bor -AGAGTTTGATCCTGGCTTAG	52
	Chl –AGAATTTGATCTTGGTTCAG	50
1492r*	TACCTTGTTACGACTT	39

*Baker *et al.*, 2003

Cloning For cloning of 16S rRNA amplicons, the QIAGEN PCR Cloning plus kit (Qiagen) were used. According to the manufacturer's instructions, an 8:1 molar ratio of amplified DNA and pDrive cloning vector for ligation was used. Briefly, 1 µl of pDrive cloning vector (50 ng/µl) was mixed with 2 µl of PCR product (27.34 ng/µl), 2 µl of nuclease-free water and 5 µl of dedicated ligation master mix before incubation at 4 °C for 30 min. For transformation, QIAGEN EZ competent cells (Qiagen) were used, according to the manufacturer's instructions. 2 µl of ligation mixture was added into one tube of dedicated competent cells and incubated at 4 °C for 5 min. The tube was the incubated at 42 °C for 30 sec before incubation on ice for 2 min. 250 µl of SOC medium was added to the reaction tube and 100, 80, 60, 40 and 20 µl of the mixture were directly plated onto five LB agar plates containing ampicillin, isopropyl β-D-1-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal). All media were prepared according to the manufacturer's instructions. The plates were incubated at 37 °C during 16 h. A second incubation at for 4 °C during 2 h was performed to enhance blue colour development and, thereby, facilitate differentiation between blue and white colonies.

Plasmid preparation Confirmed transformed white colonies were transferred into 5 ml of LB/Ampicillin medium with a toothpick, the mixture tube was shaken at 350 rpm and incubated at 37 °C overnight. LB/Ampicillin medium was prepared according to the manufacturer's instructions. The culture was centrifuged at 12,000 g during 1 min. The cell pellet was resuspended in 100 µl of Glucose-Tris-EDTA buffer (GTE) and vortexed gently. 200 µl of NaOH/SDS lysis solution was added and the tube was mixed by inversion (6-8 times). 150 µl of 5 M potassium acetate solution (pH 4.8) was immediately added and the sample was centrifuged at 13,000 g during 1 min. The supernatant was transferred into a new 2 ml reaction tube and nucleic

acids were precipitated with 0.5 ml of 80 % isopropanol by incubation on ice for 10 min following a centrifugation at 13,000 g for 1 min. The pellet was resuspended in 400 µl of TE buffer. The presence of the insertion of 16S rRNA sequence within the vector in transformed cells was confirmed by PCR using M13 primers and gel electrophoresis according to the manufacturer's instructions.

Sequencing The sequencing of the resulting PCR products consisted of first specifically amplifying the forward or the reverse side of the 16S rRNA strands. The sequencing reactions were carried out using the BigDye Terminator cycle sequencing kit (Invitrogen). Briefly, 2 µl of the previous PCR product mixture (from confirmation transformant step) was added to the PCR sequencing mixture prepared with 2 µl of the big dye mixture, 2 µl of sequencing buffer 5x concentrated, 0.2 µl of M13 forward or reverse primers (20 µM; Invitrogen) diluted in 13.8 µl of nuclease free water. After an initial 1 min denaturation step at 96 °C, samples were amplified using 25 cycles of 10 sec at 96 °C, 5 sec at 50 °C and 4 min at 60 °C. The PCR sequencing product was purified by the addition of 80 µl ethanol/sodium acetate solution. This solution was prepared by mixing 3 µl of 3 M sodium acetate and 59 µl of pure ethanol with 18 µl of deionised water. The reaction was then left at room temperature for 15 min before being pelleted by centrifugation at 12,000 g at 4 °C for 20 min. The pellet was then re-dissolved in a 70 % ethanol solution before being pelleted again at 12,000 g at 4 °C for 5 min. The ethanol-washing step was repeated two times before a final high-speed centrifugation was carried out at 14,000 g during 30 min. The DNA pellet was then re-dissolve within 20 µl of formamide by vortexing. The resulting DNA solution was injected into the Applied Biosystems 3130 Genetic Analyzer (Applied Biosystems) for sequencing by capillary electrophoresis according to the manufacturer's instructions.

16S rRNA sequencs analyses The resulting 58 full 16S rRNA sequences were aligned using the Seqmatch function (BLAST alignment) of the Ribosomal Database Project (RDP; Cole *et al.*, 2009) and assigned to different levels of taxonomic resolution ranging from genus to phylum with a confidence threshold of 80 %. Phylogenetic trees were generated using the RDP (Cole *et al.*, 2009) release 9, update 28.

2.2.7. Pyrosequencing of 16S rRNA genes

PCR amplification 16S rRNA genes were amplified using the primers Bakt_341F and Bakt_805R (**Table 2.4**) designed for 454 sequencing. These primers were specifically designed for the amplification of 400 bp spanning hypervariable regions from V3 and V4 of the 16S rRNA gene (Herlemann *et al.*, 2011). 16 primer pairs were designed as fusion primers targeting 16S rRNA including 454 adaptors, key and multiplex identifiers (Eurofins MWG Operon). PCR reactions were prepared following the FastStart high fidelity (Roche) protocol instructions, using the methodology described previously by Herlemann and colleagues (Herlemann *et al.*, 2011). The amplicon size and quality were evaluated using the DNA 1000 LabChip kit and 2100 Bioanalyzer (Agilent Technology).

Table 2.4 Oligonucleotide primers used for 16S rRNA region V3V4 amplification by PCR.

Primers	Sequence 5'-3'	Tm (°C)
Bakt_341F[§]	CCTACGGGNGGCWGCAG	55
Bakt_805R[§]	GACTACHVGGGTATCTAATCC	53

[§]Herlemann *et al.*, 2011

Library construction and purification A single library was generated containing amplicons from the 16 individual samples with 16 different multiplex identifiers. 5 µl of each 16S rRNA amplicon set were pooled together resulting in a final volume of 80 µl. The amplicon library was purified twice by using two different technologies. Firstly, small fragments (<200 bp) were removed by using QIAquick columns (Qiagen). The purified library was eluted in 30 µl of MilliQ water. Afterwards, the eluted library was size selected (>250 bp) by using the solid phase reversible immobilisation paramagnetic bead based technology (AMPure XP beads, Beckman Coulter) with a bead/DNA ratio of 1/1 (v/v). The final purified amplicon library was eluted in 20 µl of MilliQ water. All purification steps were carried out according to the manufacturer's instructions. The integrity and the size of the libraries were determined using the DNA 1000 LabChip kit and the 2100 Bioanalyzer (Agilent Technologies). Library quantification was carried out according to the Roche 454 instructions by using the Quant-IT PicoGreen double-stranded DNA assay kit (Invitrogen).

16S rRNA amplicon sequencing Sequencing was carried out on a Roche 454 GS Junior platform. Before pyrosequencing, an emulsion PCR (emPCR) was performed, where DNA fragments derived from the library of enriched target regions were amplified on small beads. The emPCR was carried out according to the manufacturer's instructions by using the GS Junior Titanium emPCR kit (lib-A; Roche). After emPCR, the emulsion was stopped and beads enriched, as well as the sequencing primers annealed. The sequencing was performed according the manufacturer's instructions by using the GS Junior Titanium Sequencing kit (Roche) and the GS Junior Titanium PicoTiterPlate kit (Roche).

Bioinformatic data analysis Data were analysed using the Mothur community analysis pipeline (Schloss *et al.*, 2009), with an average linkage and distance cutoff of 0.03 (97 % of sequence identity). For this analysis, the generated .sff file was used, containing the flowgram from the 454 pyrosequencing platform and the sample barcode sequences. The use of this open-source software package involved four sequential steps. The first step consisted of the preparation and denoising of sequences as well as extraction of the V3-V4 region from 16S rRNA sequences. In this step, low-quality sequences were removed and the frequencies of sequencing and PCR errors reduced. Following the partition of the flowgram data by samples (based on the barcodes), the sequences were trimmed to a specified length range and the flowgram were denoised using the PyroNoise algorithm (Quince *et al.*, 2009). The target region was then specifically located, verified and extracted using the V-Xtractor algorithm (Hartmann *et al.*, 2010), specific for bacterial, archaeal and fungal small subunit rRNA gene regions V3-V4. Then, PCR errors were removed using the Seqnoise algorithm (http://code.google.com/p/ampliconnoise/). Finally, PCR chimeras were removed using the UCHIME algorithm (Edgar *et al.*, 2011) and each sequence file was rendered non-redundant. The second step of the pipeline consisted of clustering sequences into operational taxonomic units (OTUs). OTU binning was completed using a hierarchical clustering algorithm implemented within Mothur. The third step involved assignment of the taxonomic information to sequences and OTUs. Unique fasta sequences were aligning to the Greengenes reference database (DeSantis *et al.*, 2006) using a naïve Bayesian classifier. Then, taxonomic information was assigned to the individual OTU with 80 % of consensus. The analysis was carried out on randomly subsampled OTUs such that each file contained the same number of sequences (1,405). The last step of the use of the Mothur pipeline included some basic

metrics of community structure (α- and β-diversity), using different statistical measurements, such as Shannon diversity and Bray-Curtis indices, analysis of molecular variance (AMOVA) as well as different visualisation approaches for the ordination of results for example through heat maps or principal coordinate analysis (PCoA). The plots were generated using Excel or the R software (R version 2.13.2, http://www.r-project.org/) for the principal coordinate analysis (PCoA; pcoa function) using the ape package.

2.2.8. Proteomic *de novo* peptide sequencing

Protein extraction Protein extraction was performed on four biological replicates on each of the four sampling dates, using the Norgen All-in-one based biomacromolecular isolation method (NA), as described previously (Section 2.2.5).

SDS-PAGE 5 µl of each of the 16 individual protein fractions were subjected to 1D sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE; Bio-Rad Laboratories) in conjunction with staining in LavaPurple protein stain (Fluorotechnics). Precision Plus Protein Stained and Unstained Standard Ladder (Bio-Rad Laboratories) were loaded onto the gels for molecular mass estimation. Gels were visualised using a Typhoon 9400 Variable Mode Imager (GE Healthcare).

MALDI-ToF/ToF analysis A highly dominant protein gel band was cut out and placed into a 1.5 ml Eppendorf tube. Gel reduction was carried out in a mixture of 100 mM of NH_4CO_3 (Sigma) and 10 mM of DTT (Sigma) in Thermomixer (Eppendorf) at 56 °C, 500 rpm during 30 min. Alkylation was carried out in a mixture of 100 mM NH_4CO_3 and 55 mM iodoacetamide (IAA, Sigma) at room temperature for 20 min. Gel destaining was carried out in 50 mM NH_4CO_3 and 50 % methanol (BioSolve) in a

Thermomixer at 37 °C, 500 rpm during 20 min. The protein band was then digested using 80 ng of trypsin (Trypsin Gold, Promega) diluted in 50 mM NH_4CO_3 solution and incubated at 4 °C for 30 min followed by an overnight incubation in a Thermomixer at 37 °C. After two wash steps in 50 % v/v acetonitrile (ACN, BioSolve) / 0.1 % v/v trifluoroacetic acid (TFA, Sigma) solution at 37 °C for 10 min, the isolated peptides were dried in a Speedvac (Heto Lab). The peptides were then again resuspended in ACN/TFA (50 % / 0.1 %) and loaded onto a Matrix-Assisted Laser Desorption Ionization (MALDI) target plate. Peptide tandem mass spectra were acquired on a MALDI-Time of Flight (ToF/ToF), 4800 Proteomics Analyzer (Applied Biosystems).

De novo peptide sequencing Peptide sequences were determined manually by converting the mass differences of consecutive ions in obtained tandem mass spectra to corresponding amino acid sequences. The resulting peptide sequences were then BLAST-searched (Altschul *et al.,* 1997) using default parameters against the NCBInr database.

2.3. Results and discussions

2.3.1. Initial characterisation of LAO communities

From the initial survey of local biological wastewater treatment plants, it was apparent that transitory foaming and bulking events occur at the WWTP located in Schifflange (Esch-sur-Alzette, Luxembourg; 49°30'48.29"N; 6°1'4.53"E, **Figure 2.1**). The Schifflange biological WWTP is an activated sludge plant with alternating anaerobic/aerobic treatment phases operated for phosphorus and nitrogen removal. Informations obtained from the plant operators from the Syndicat Intercommunal à Vocation Ecologique (SIVEC) highlighted the propensity of excess floating biomass on the surface of anoxic tank in winter compared to spring, summer and autumn. Since this WWTP is in close proximity to our laboratory, is of a relatively important size for Luxembourg and routinely carries out multi-parameter online measurements of essential physico-chemical characteristics (**Table 2.5**, **Figure 2.3**; Section 1.1.2.2.3), biomass obtained from this plant was used for the following analyses.

Table 2.5 Overview of the main characteristics of the Schifflange wastewater treatment plant. Measures obtained from the SIVEC (WWTP of Schifflange). *Average values measured in 2012, using multi-parameter online probes.

Schifflange WWTP		
Population equivalent (habitants)		82,834
Rate of flow of WWTP (m^3/day)		19,704
COD (mg/l)*	Influent	652.94
	Effluent	33.94
BOD_5 (mg/l)*	Influent	225.45
	Effluent	2.47

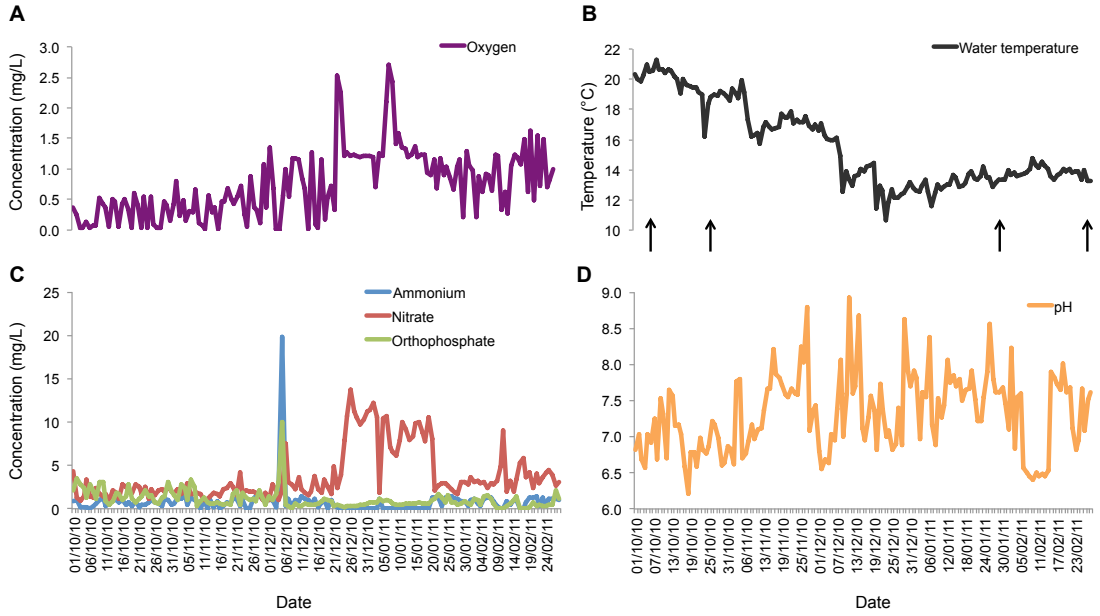

Figure 2.3 Variation of essential physico-chemical parameters at the Schifflange wastewater treatment plant from 1 October 2010 to 28 February 2011. (**A**) Dissolved oxygen. (**B**) Water temperature, arrows indicate sampling time points (Section 2.2.2). (**C**) Nitrate, orthophosphate and ammonium. (**D**) pH measurements. Values obtained from the SIVEC (WWTP of Schifflange) and recorded using multi-parameter online probes.

2.3.1.1. Microscopic analyses

The presence of filamentous organisms in floating sludge sampled from the anoxic treatment tank of the Schifflange WWTP was confirmed by SYBR Green I staining followed by epifluorescence microscopy (**Figure 2.4, A**). The presence of these filamentous organisms has previously been described in activated sludge by several research groups (e.g., Bradford *et al*., 1996; Snaidr *et al.,* 2002; Nielsen *et al.*, 2009; Section 1.1.2.3.2). Filamentous bacteria constitute the backbone of the individual "islets" identifiable with the naked eye at the surface of wastewater treatment tanks (**Figure 2.2**). Their excessive growth can have pronounced effects on biomass settling and foaming, causing serious operational problems for biological wastewater treatment operations (Mielczarek *et al.*, 2012). According to a recent survey focussed on filamentous bacteria, these were identified in 25 EBPR plants over a period of 4 years (Nielsen *et al.*, 2010). Typically, such filamentous bacterial groups are dominated by *Microthrix parvicella* as well as different other organisms belonging to the phylum *Chloroflexi*, *Haliscomenobacter hydrossis* belonging to *Bacteroidetes*, *Curvibacter* spp. belonging to the *Betaproteobacteria* and some filaments belonging to the phylum TM7 and few others (Nielsen *et al.*, 2010).

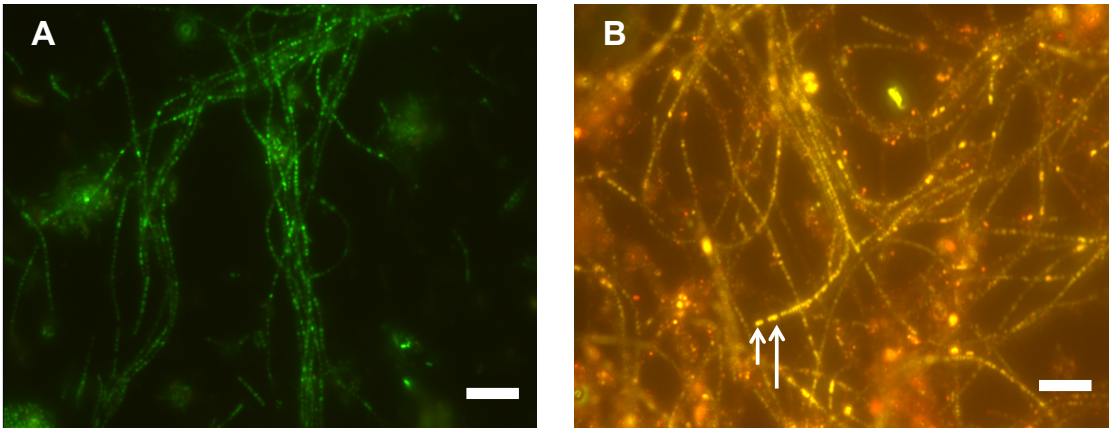

Figure 2.4 Micrographs of the LAO-enriched microbial community sampled from the anoxic tank of the Schifflange biological WWTP following SYBR Green I (**A**) and Nile Blue A (**B**) staining. Pane **B**, the arrows highlight intracellular lipid inclusion. Scale bars are equivalent to 10 µm.

The capacity of the sampled biomass to incorporate intracellular lipids (e.g. poly-3-hydroxybutyrate, PHB) was evaluated by epifluorescence microscopic observation after Nile blue A staining (Ostle & Holt, 1982). Interestingly, most of the filamentous bacteria exhibit considerable lipid inclusions, (**Figure 2.4, B**). Most of the stained lipids are present as lipid inclusions within the filamentous organisms or sorbed to the filamentous cell membrane surface. Several filamentous bacteria species show the capacity to accumulate lipid structures, for example *Microthrix parvicella* (Nielsen *et al.*, 2002), *Meganema perideroedes* (Kragelund *et al.*, 2005) as well as several members of the *Mycolata* group such as *Gordonia* spp. or *Skermania* spp. (Kragelund *et al.*, 2007). The Nile Blue A staining does however not allow differentiation between PHA and lipids. The carbon sources used for PHA synthesis are short chain fatty acids (SCFAs; including acetate, propionate and butyrate and/or glucose). Given the inability to differentiate between PHA and lipids by microscopic analysis, a long chain fatty acid analysis of the floating LAO biomass have been conduct.

2.3.1.2. Long chain fatty acid analysis

An important consideration for using an LAO-based lipid reclamation strategy is the unprocessed uptake and storage of LCFAs from wastewater. In order to assess potential differences in the LCFA fractions contained within the wastewater and in the LAO biomass, non-polar metabolites were extracted from extra- (i.e. reflecting the wastewater) and intracellular (i.e. reflecting the LAOs biomass) compartments and analysed using GC coupled to MS (GC-MS; Section 2.2.4). LCFAs dominate the non-polar metabolite fraction of wastewater (**Figure 2.5, A**) and LAO biomass (**Figure 2.5, B**). Palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1) and linoleic acid (C18:2; Section 1.1.1.1) are the dominant intra- and extracellular LCFAs. Considering previous GC-MS data, these measurements confirm the predominance of LCFAs in a range from C16:0 to C18:0 in domestic sewage sludge (Jardé *et al.*, 2005) as well as in municipal wastewater (Al-Halbouni *et al.*, 2009).

As reported, the enrichment of specific LCFAs in wastewater was demonstrated through this preliminary long chain fatty acids survey and highlighted again the presence of specific LAOs capable of high LCFA accumulation within the floating biomass recovered from the anoxic tank of the Schifflange biological WWTP. Interestingly, it was observed that the LCFAs presented very similar profiles between the two cellular compartments. It is noteworthy that this finding suggests that endogenous LCFAs in the wastewater are taken up directly by LAOs without conversion and, thus, the vast majority of the chemical energy contained within them should be recoverable from the LAO biomass.

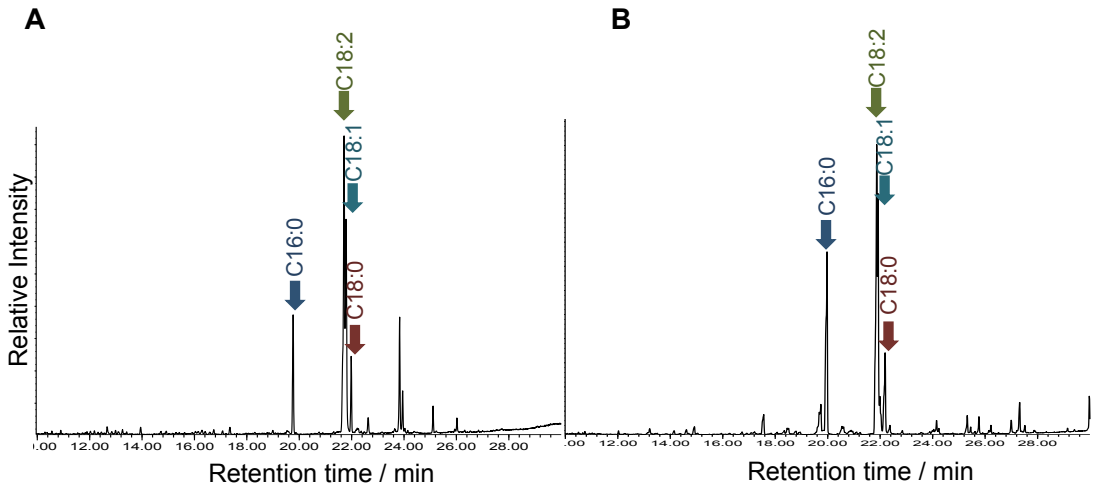

Figure 2.5 Gas chromatograms of non-polar metabolite fractions extracted from wastewater (**A**) and the floating biomass (**B**). The main peak are palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1) and linoleic acid (C18:2).

2.3.1.3. Phylogenetic survey of a single LAO sample

In order to provide a first overview of the overall phylogenetic structure and diversity of the bacterial LAO community within the anoxic phase of the Schifflange biological WWTP, a full-cycle 16S rRNA gene analysis was carried out (Hugenholtz, 2002). This involved the cloning of amplified whole 16S rRNA genes from a given LAO sample, sampled on 27 May 2010, followed by the sequencing of the resulting full 16S rRNA gene sequences using Sanger technology (Section 2.2.6). This analysis revealed the presence of at least 12 different bacterial genera belonging to 3 dominant phyla, the *Actinobacteria, Proteobacteria* and *Bacteroidetes* together comprising 90 % of the LAO community (**Figure 2.6, A**). It is noteworthy that this analysis demonstrated that an organism closely related to the previously described actinobacterial LAO *Microthrix parvicella* (Section 1.1.3.1), dominated the community at the species level (around 30 % of the community; **Figure 2.6, B**). The 16S rRNA gene sequences were closely related (> 98 %) to the gene sequence of

Microthrix parvicella BIO17-1 (Levantesi *et al.*, 2006) suggesting that the dominant population in the LAO-enriched community sample belongs to this species but that clearly different strains are apparent (**Figure 2.6, C**). Finally, the second most abundant species, *Haliscomenobacter*, represented 9 % of the community. It is a common filamentous bacterium belonging to the phylum *Bacteroidetes*, which has been detected worldwide in activated sludge samples (Kragelund *et al.*, 2008).

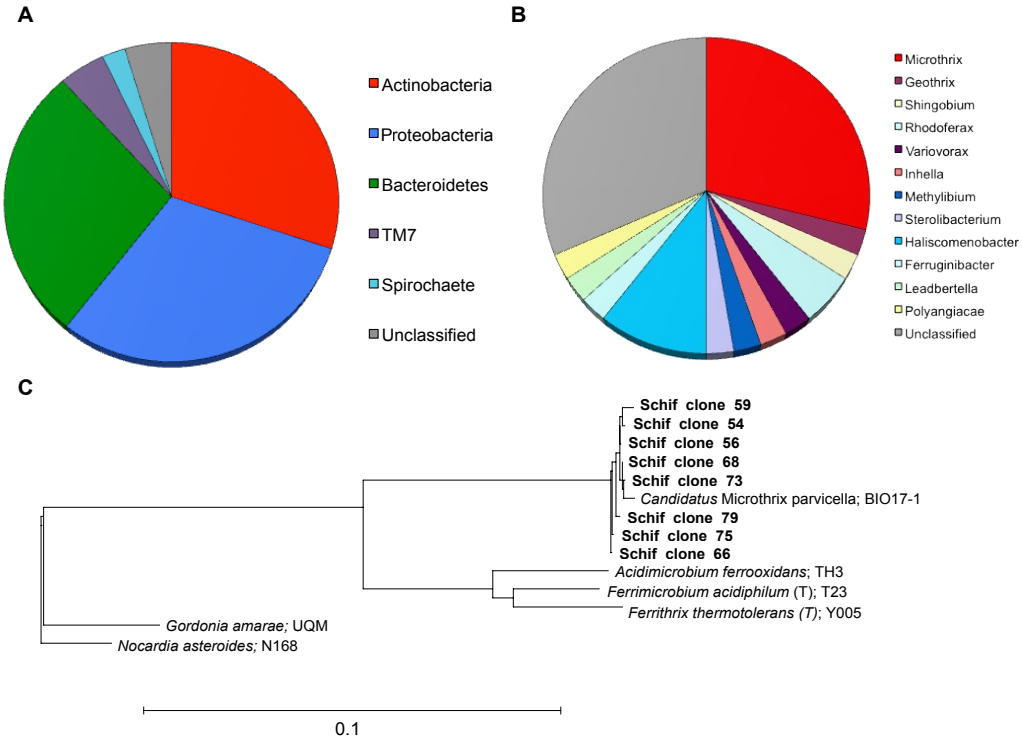

Figure 2.6 Phylogenetic overview derived from entire 16S rRNA gene sequences amplified from DNA extracted from the LAO-enriched microbial community sampled at the anoxic tank of the Schifflange WWTP. (**A,B**) Pie charts representing phylum-level (**A**) and genus-level (**B**) diversity determined from the obtained full 16S rRNA gene sequences. (**C**) Phylogenetic tree at species-level resolution showing some genetic heterogeneity among close relatives of *Microthrix parvicella.* The scale bar indicates 0.1 estimated changes per nucleotide.

To validate these initial findings and provide some *in situ* information on the presence of *Microthrix parvicella* within the LAO-enriched microbial community, a FISH analysis was performed (Section 2.2.3), specifically targeting *Microthrix parvicella* (Levantesi *et al.*, 2006), on a cell permeabilised LAOs suspension, sampled from the Schifflange WWTP (**Figure 2.7**, Section 2.2.3**)**.

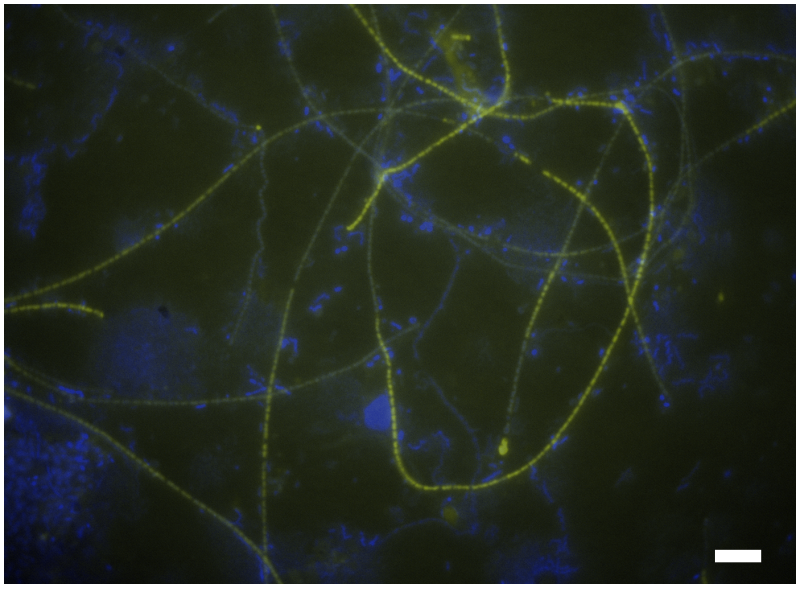

Figure 2.7 Micrograph of the LAO-enriched microbial community sampled from the anoxic tank of the Schifflange biological WWTP following FISH for *Microthrix parvicella*. *Microthrix parvicella* is observable in yellow, surrounded by bacterial of variable morphologies, visible in blue. The scale bar is equivalent to 5 µm.

These two results confirm the presence and the pronounced abundance of *Microthrix parvicella* in the floating LAO-enriched microbial community within the anoxic tank of the Schifflange biological WWTP as encountered in other activated sludge plants throughout the world (Nielsen *et al.,* 2009).

2.3.2. Comparative phylogenetic and proteomic analyses of LAO-enriched microbial communities

In order to fully appreciate the dynamics of the LAO-enriched microbial communities across space and time, a comparative analysis on four different sludges, sampled from the Schifflange WWTP, have been carried out at different sampling dates (**Table 2.2**; D1, 4 October 2010; D2, 25 November 2010; D3, 25 January 2011 and D4, 23 February 2011), considering four biological replicates belonging to four selected dates. The dates were selected in order to sample across a temperature gradient since water temperature is know to have an marked effect on the abundance of LAOs within biological wastewater treatment systems (Jones & Schuler, 2010).

2.3.2.1. Comparative phylogenetic analyses of LAO-enriched microbial communities

The phylogenetic analysis of each of the 16 resulting sludge samples was performed by 16S rRNA gene sequencing using pyrosequencing (Section 2.2.7). This resulted in the identification of 9,729 unique OTUs (**Supplementary Table I**). Compared to the previous 16S rRNA genes (Section 2.3.1.3), this high-throughput analysis allowed the partial sequencing (i.e. 400 bp) per sample of an average of 3,423 16S rRNA gene sequences, from the V3 and V4 hypervariable regions.

Overall, this analysis revealed the presence of 757 bacterial genera belonging to 23 phyla across the 16 samples, representing 96.4 % of the LAO communities. The average α-diversity, (**Figure 2.8**) shows a distinct difference in terms of bacterial population richness and evenness between the 4 October 2010 and 25 October 2010 dates (henceforth, referred to as autumn LAO communities) and 25 January 2011 and

23 February 2011 dates (henceforth referred as winter LAO communities, **Figure 2.8**).

The winter LAO communities (D3, D4) appear to be richer (i.e., containing more different species) and exhibit a more even species distribution (i.e. different species, more evenly distributed) than within the autumn LAO communities (D1, D2). Additionally, a tendency for a higher variability in terms of species richness is apparent between different iselts (biological replicates) in autumn compared to the winter LAO communities.

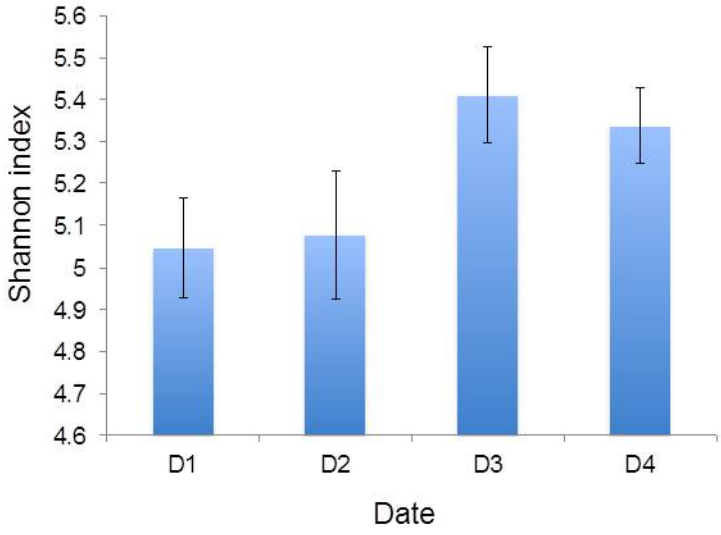

Figure 2.8 Bar charts of average α-diversity based on Shannon diversity indices of bacterial species present in the LAO-enriched microbial communities at the 4 distinct sampling dates. The error bars represent the standard deviation among the 4 analysed samples. D1: 4 October 2010; D2: 25 October 2010; D3: 25 January 2011; D4: 23 February 2011.

To further highlight the difference between the winter and autumn communities, the identity and abundance of the 10 dominant phyla within each single biological

replicate were analysed (**Figure 2.9**). These results confirm the dominance of three major phyla, *Proteobacteria*, *Actinobacteria* and *Bacteroidetes* within all samples, as seen previously from the entire 16S sequences (Section 2.3.1.3). The phylum *Firmicutes* is detected among the LAO community members, but represents only 3.4 % of the total bacterial community in contrast to the predominance of such phyla previously detected in activated sludge from 14 different WWTPs (8.1 %; Zhang *et al.*, 2011). In contrast, the LAO-enriched microbial communities appear to be dominated by *Actinobacteria* (18.9 %) and *Bacteroidetes* (17.6 %) compared to the bacterial populations identified within activated sludge (6.5 and 7 % respectively; Zhang *et al.*, 2011). The data of this doctoral research highlight the stark differences in microbial community composition between submerged and floating sludges. A clear variation between the two most abundant phyla, i.e. *Proteobacteria* and *Actinobacteria,* among dates was observed. Within the autumn LAO communities, *Proteobacteria* and *Actinobacteria* represent 50 and 16 % of the total number of phyla, while in the winter LAO communities, these two phyla represent 40 and 21 % of the total phyla, respectively. Consequently, there exists a clear phylum-level shift between both seasons.

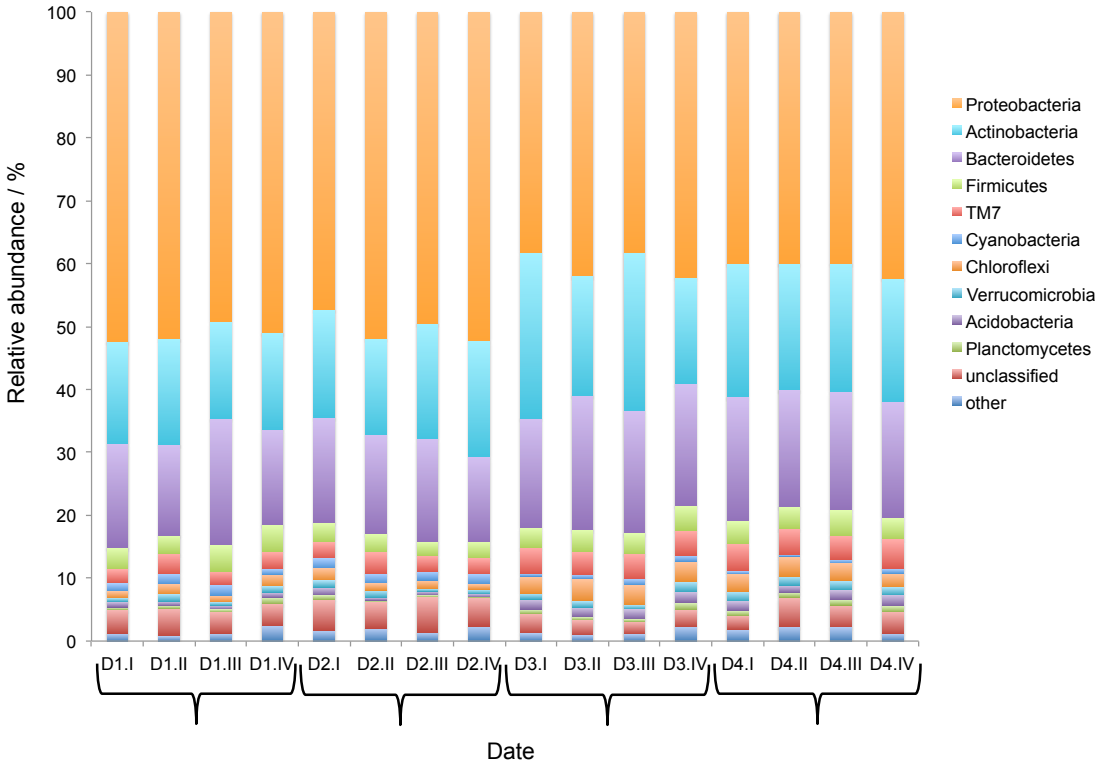

Figure 2.9 Bar chart of phylum-level distribution of bacteria presents in 4 different isles sampled on 4 different dates. D1: 4 October 2010; D2: 25 October 2010; D3: 25 January 2011; D4: 23 February 2011. I, II, III, IV represent the individual islets sampled on each date.

To further measure and reveal the level of similarity and dissimilarity in community membership and structure in space and time, β-diversity was calculated using the Bray-Curtis index based on the OTU diversity and the results visualised using PCoA (**Figure 2.10**). The obtained results further point to a seasonal shift in community structure between the autumn and winter communities (**Figure 2.8**).

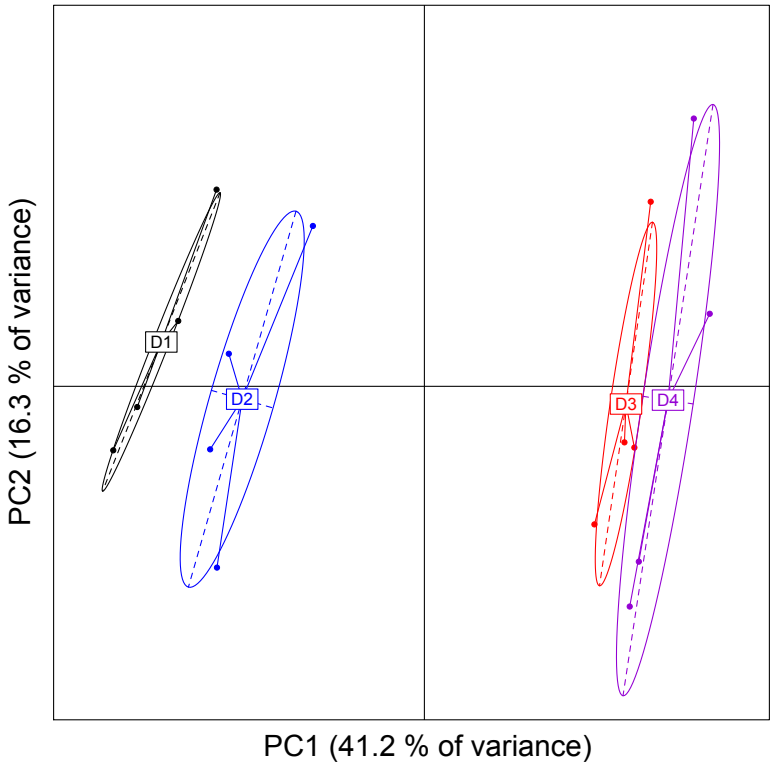

Figure 2.10 Principal coordinate analysis (PCoA) of Bray-Curtis dissimilarity indexes calculated on the based on OTU abundances obtained on 4 biological replicates sampled on the 4 different dates. D1: 4 October 2010, D2: 25 October 2010, D3: 25 January 2011 and D4: 23 February 2011. The centre of gravity for each date cluster is marked by a rectangle and the coloured ellipse covers 67 % of the samples belonging to the cluster.

Finally, as reflected by the area of the different ellipses among the four different LAO-enriched microbial communities, the biological replicates present in D2 (25 October 2010) and D4 (23 February 2011) exhibit the most pronounced compositional heterogeneity (**Figure 2.10**).

Differences in community structure among groups were further assessed by statistical analysis using molecular variance (AMOVA). AMOVA is a non-parametric test analogous to traditional analysis of variance measures. This method is widely used in population genetics to test the hypothesis that genetic diversity within two populations

is not significantly different from that which would result from pooling the two populations (Excoffier *et al.*, 1992). The AMOVA statistics (**Figure 2.11**) were calculated using the Bray-Curtis distance matrix.

Phylogenetic diversity analyses confirmed the presence of significant differences among the winter and autumn LAO communities and the taxa clustering previously observed (**Figure 2.10)**. The LAO-enriched microbial communities sampled on D1 (4 October 2010) and D3 (25 January 2011) are the most dissimilar (F-statistic= 6.68, p-value < 0.001).

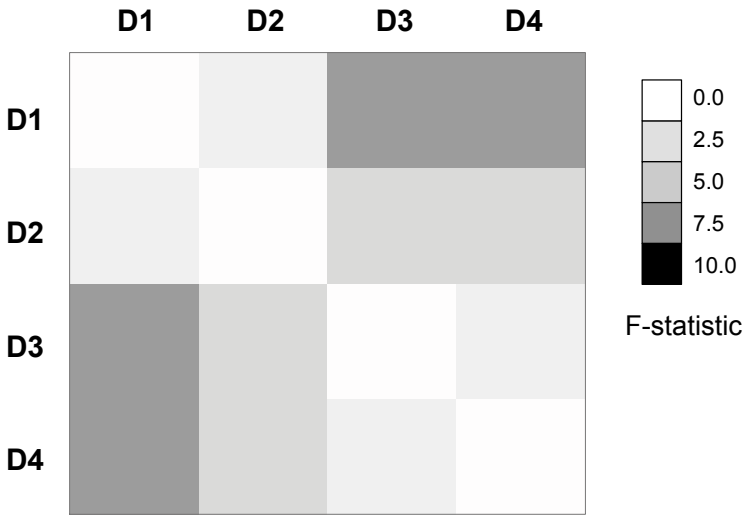

Figure 2.11 Matrix of pairwise comparisons of the phylogenetic diversity (AMOVA, F_{ST}) of LAO-enriched microbial communities recovered from each of the four sampling dates. D1: 4 October 2010, D2: 25 October 2010, D3: 25 January 2011 and D4: 23 February 2011.

Within the LAO-enriched microbial communities, the dominant OTU, considering the overall 16S rRNA genes sequences analysed, across the 16 samples analysed,

correspond to *Microthrix parvicella*, which represent on average 8.38 % of the total detected species. The overall dominance of *Microthrix parvicella* confirmed the results obtained using the 16S rRNA phylogenetic analysis carried out on the single sample (Section 2.3.1.3). The second most dominant OTU, which was not detected in the previous phylogenetic analyses (Section 2.3.1.3), is related to the *Alkanindiges* genus from the *Moraxellaceae* family. Representative members from the *Alkanindiges* genus were previously isolated from oilfield soils (Bogan *et al.*, 2003). They are recognised as obligate oil degraders and have previously been implicated in foaming in activated sludge systems (Klein *et al.*, 2007). Interestingly, WEs accumulation has been observed in *Alkanindiges illinoisensis* cells (Manilla-Pérez *et al.*, 2010). Additionnaly, *Acinetobacter* spp. also belonging to the *Moraxellaceae* family have been identified as the third most abundant genus in the LAO-enriched microbial communities. *Acinetobacter* spp. is well known to occur in activated sludge biomass (Carr *et al.,* 2003). As *Alkanindiges* spp., have been found to be oil degraders in soil and sediments contaminated with crude oil or fuel oil (Bordenave *et al.,* 2007). *Acinetobacter* spp. have also been shown to be able to accumulate WEs (Hanson *et al.,* 1997). Interestingly, *Acinetobacter baylyi*, strain ADP1 is known to naturally accumulate storage lipids, including TAG (Santala *et al.,* 2011) and has been extensively described to be able to carriy out the final step in TAGs and WE biosynthesis using WE synthase/acyl-CoA:DAG acyltransferase (WS/DGAT) which catalyzes the acylation of DAG by using acyl-CoA as the substrate (Kalscheuer & Steinbüchel, 2003; Stöveken *et al.*, 2005).

However, members of the *Alkanindiges* and *Acinetobacter* genera have not yet been described as organisms capable of accumulating exogenous lipids. Therefore, it remains to be seen if these members are actual competitors of *Microthrix parvicella*

over lipids in wastewater.

It is noteworthy that each of these three most abundant OTUs belong to the two main dominant phyla. To further appreciate the variation in abundance of these OTUs and their participation to the phylum variation measured among seasons, the average proportion of each of the individual species within the four different dates was calculated (**Figure 2.12**).

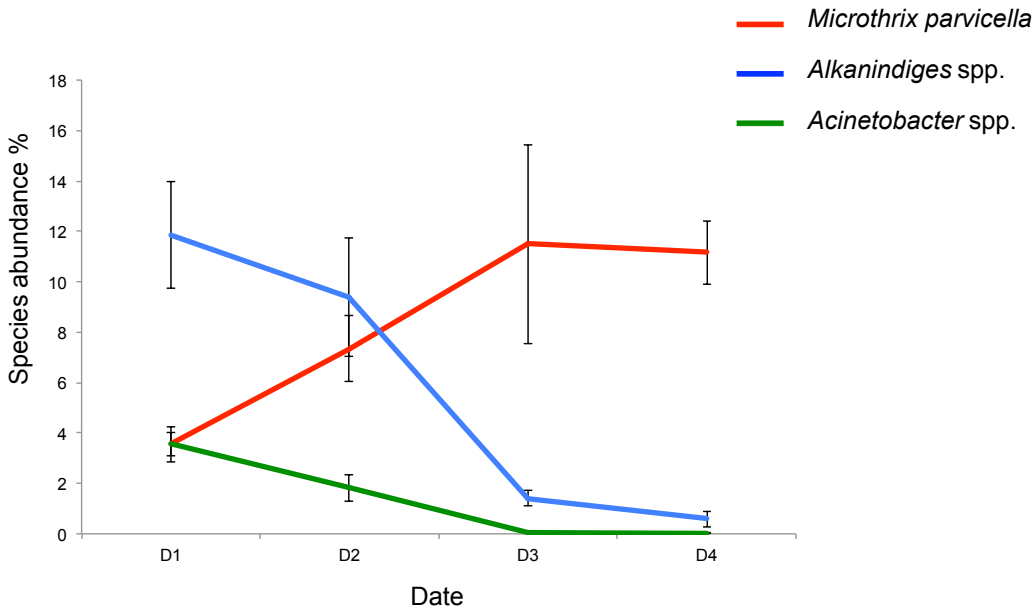

Figure 2.12 Average species abundance of the three most abundant OTUs among the 4 sampling dates. Error bars represent the standard deviations between the four biological replicates (islets). D1: 4 October 2010, D2: 25 October 2010, D3: 25 January 2011 and D4: 23 February 2011.

From the overall phylogenetic profiles across the four distinct dates, a shift in the dominant community members is apparent (**Figure 2.12**). Autumn communities are dominated by organisms closely related to *Alkanindiges* spp. and *Acinetobacter* spp. from the *Moraxellaceae* family, whereas winter LAO communities are dominated by *Microthrix parvicella.* The dominance of *Microthrix parvicella* in winter may be

related to their preference for low temperatures (Rossetti *et al.*, 2005). Importantly, by calculating correlation between axes score and OTU abundances it is apparent that *Alkanindiges* spp., *Acinetobacter* spp. and *Microthrix parvicella* are the main drivers of community structure observed, explaining the differences observed between autumn and winter clusters for 97.5, 93.5 and - 82.8 % of PC1 axis of the PCoA (**Figure 2.10)**. Consequently, *Alkanindiges* spp., *Acinetobacter* spp. and *Microthrix parvicella* are the main community members linked to the LAO community structure dissimilarities observed between autumn and winter (**Figure 2.10**). The abundances of *Microthrix parvicella* and *Alkanindiges* spp. represent, individually, approximately 12 % of the total detected species members within the respective seasonal community considered. Such dominance of individual species among the community contributes significantly to the unevenness of the community but does facilitate the community genomic analyses since both populations should be well resolvable from community genomic datasets.

2.3.2.2. Comparative proteomics

In order to appreciate further the functional outcome of the phylogenic shift observed between the autumn and winter LAO communities, a preliminary comparative proteomic study between the four different dates was performed. The analysis was based on single-dimensional SDS-PAGE gel band profile comparison across dates, followed by targeted peptide sequencing of a band of highest interest.

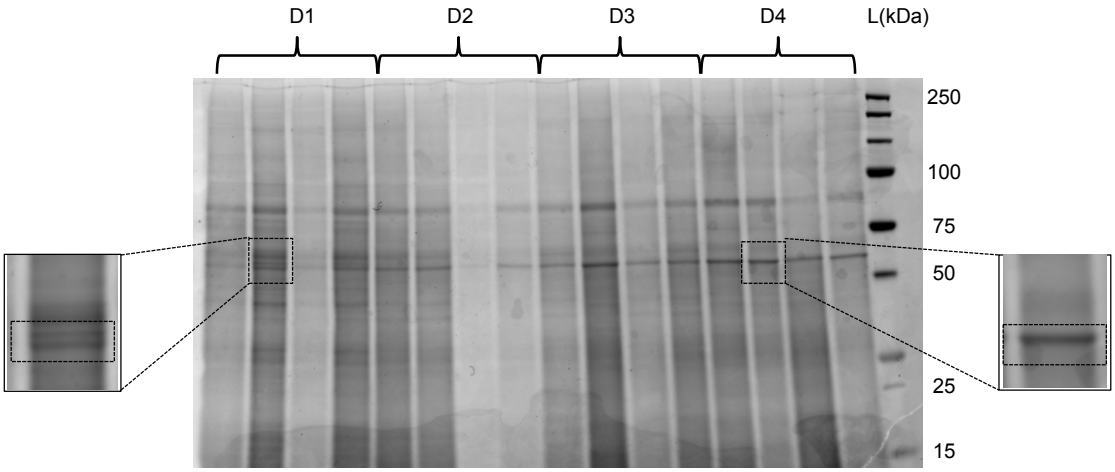

Figure 2.13 SDS-PAGE image of representative protein fractions from each of the 4 biological replicates originating from the four different sampling dates. From left to right, magnified regions of the protein banding profile detected between 100 and 50 kDa for the autumn and winter microbial communities, respectively. Right pane represents the dominant gel band from the D4 (23 February 2011) protein fraction, which was submitted to tryptic digestion and MALDI-ToF/ToF analysis. Left pane highlights the targeted band from the D1 (4 October 2010) protein fraction. D1: 4 October 2010, D2: 25 October 2010, D3: 25 January 2011 and D4: 23 February 2011.

In the protein banding profile presented for the winter communities (**Figure 2.13, right pane**), a strong protein band is apparent around 65 kDa, which is less dominant in the protein pattern observed for the autumn community-wide protein profiles (**Figure 2.13, left pane**). This particular gel band was excised, tryptically digested and subjected to MALDI-ToF/ToF analysis, followed by a *de novo* peptide sequencing of dominant spectra (Section 2.2.8). The resulting peptide sequences (**Table 2.6**) were then searched against current genomic databases.

Table 2.6 BLASTP output from NCBI search of *de novo* peptide sequencing following MALDI-ToF/ToF analysis.

Peptides fragment sequences	Sequences producing significant alignment	Score (e-value)	Query cover (%)
1_SDSEDWCNKFIGTGTPVNDFR **2**_FSANEGDGPGGAFIR **3**_VKSNTAPVVVTENGTAQR **4**_RGAGTAPGVFSN **5**_DVDVVTR **6**_WFIGER	Exported hypothetical protein [*Candidatus* Microthrix parvicella]	92.4 (2e-04)	100

Following de novo peptide sequencing, the peptide sequences have been confidently identified as protein annotated as an exported hypothetical protein encoded by *Microthrix parvicella*, strain RN1 (Mcllroy *et al.,* 2013). It is noteworthy that this result confirms further the high level of activity of *Microthrix parvicella* in the LAO-enriched microbial community (**Table 2.6**, **Figure 2.14**). Further detailed biochemical characterisation of this highly abundant protein is required in order to elucidate its function in *Microthrix parvicella*.

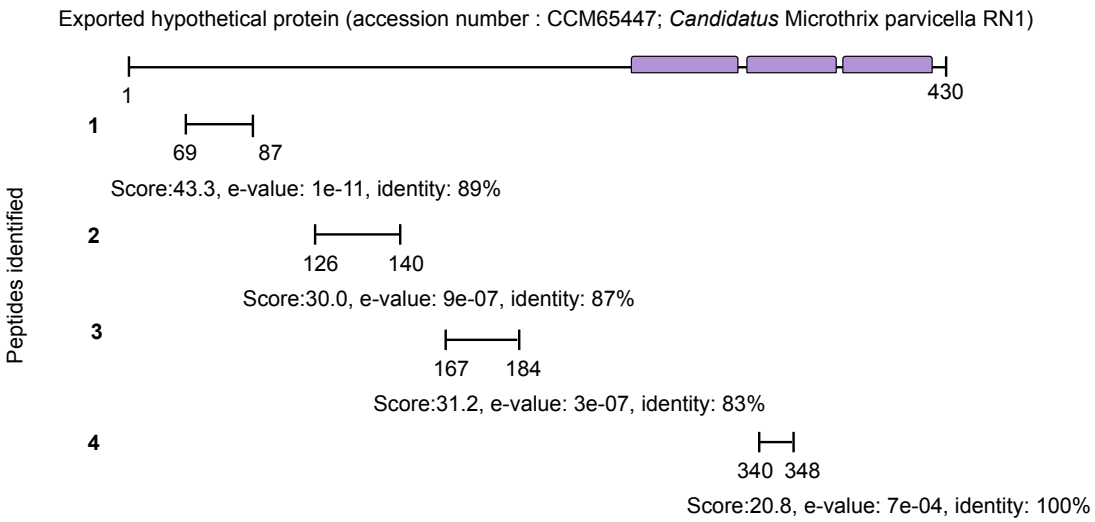

Figure 2.14 Schematic representation of the peptides mapped to the exported hypothetical protein of *Microthrix parvicella* and respective alignment scores from BLASTp. The purple area, on the query sequence represents the motif S-layer homology domain, characterisitic for an exported protein identified using an InterproScan search. 1,2,3 and 4 denotes the peptide numbers as annotated in **Table 2.6**.

2.4. Conclusion and perspectives

The initial results presented in this chapter confirm the choice of the WWTP located at Schifflange as an ideal model system for the study of lipid processing and uptake by lipid accumulating microbial communities in biological WWTP. It is apparent from the results, that the LAO-enriched microbial communities are dynamic over time with a strong seasonal variation. The comparison of the diversity between autumn and winter communities emphasises the presence of multiple different LAO-enriched microbial community compositions, characterised by a switch in the dominant species. The community structures are most likely influenced by temperature variation. Especially, two sampling dates, 4 October 2010 and 25 January 2011, appear to show the most pronounced difference in phylogenetic variation. Winter LAO communities are vastly dominated by *Microthrix parvicella*, a known specialized lipid accumulating organism. In contrast, autumn communities are dominated by organisms closely related to *Alkanindiges* spp. and *Acinetobacter* spp. Interestingly, the communities dominated by *Microthrix parvicella* exhibit marked expression of a protein annotated an exported hypothetical protein encoded by *Microthrix parvicella*. The function of this protein is currently unknown and is a prime target for future in-depth biochemical characterisation. However, such analyses were beyond the scope of the present doctoral research project.

To further characterise and understand the genetic and functional differences in lipid accumulation potential in winter compared to autumn LAO communities, the application of a comparative molecular eco-systematic analytical approach was chosen (Chapter 4). Designated autumn and winter representative LAO communities were selected among the biological replicates of 4 October 2010 and the 25 January

2011 sampling dates, as these presented the highest and largest relative differences in the proportion of *Microthrix parvicella*. A comparative eco-systems level integrated omic analysis comprising community genomics, transcriptomics, proteomics and metabolomics applied to autumn and winter LAO communities should thus, reveal the link between genetic potential and functionality in lipid accumulation. Such an approach should highlight the processes involved in the transformation of lipids within the biomass and should, specifically, reveal the biochemical transformations of LCFAs by *Microthrix parvicella*.

CHAPTER 3: A BIOMOLECULAR ISOLATION FRAMEWORK FOR ECO-SYSTEMS BIOLOGY

This section is published in:

Roume H, Muller EEL, Cordes T, Renaut J, Hiller K and Wilmes P (2013). A biomolecular isolation framework for eco-systems biology. *The ISME Journal* **7**: 110-121.

Roume H, Heintz-Buschart A, Muller EEL, and Wilmes P (2013). Sequential isolation of metabolites, RNA, DNA and proteins from the same unique sample. *Methods in Enzymology***, in press**.

3.1. Introduction

Natural microbial communities play fundamental roles in the Earth's biogeochemical cycles as well as in human health and disease, and provide essential services to mankind. They represent highly complex, dynamic and heterogeneous systems (Denef *et al.*, 2010). The advent of high-resolution molecular biology methodologies, including genomics (Tyson *et al.*, 2004), transcriptomics (Frias-Lopez *et al.*, 2008), proteomics (Ram *et al.*, 2005) and metabolomics (Li *et al.*, 2008), is facilitating unprecedented insights into the structure and function of microbial consortia *in situ*. Beyond isolated biomolecular characterisation, integrated omics combined with relevant statistical analyses offer the ability to unravel fundamental ecological and evolutionary relationships, which are indiscernible from isolated omic datasets (Wilmes *et al.*, 2010). Furthermore, space- and time-resolved integrated omics have the potential to uncover associations between distinct biomolecules which, in turn, allows for discovery of previously unknown biochemical traits of specific microbial community members (Fischer *et al.*, 2011). Consequently, in order to elucidate mechanistic details of lipid uptake and accumulation by LAOs, high-resolution data generated in a truly systematic fashion including robust sampling, sample preservation and biomolecular isolation methodologies are essential to facilitate meaningful data integration and analysis.

In *Foundation of Systems Biology*, Kitano defines the ideal systematic measurement as the "simultaneous measurement of multiple features for a single sample" (Kitano, 2001). This consideration is particularly important in eco-systems biology as a vast microbial diversity is a hallmark of natural settings (Caporaso *et al.*, 2010b) and within-population genetic heterogeneity is also very pronounced (Wilmes *et al.*,

2009b). Only if high-resolution omic measurements are obtained from single unique samples will the subsequent integration of high-resolution omic data enable true systems-level understanding of community-wide, population-wide and individual-level processes (Muller *et al.,* 2013). No methodology existed for the isolation of all concomitant small molecules (metabolites) and biomacromolecules (DNA, RNA and proteins) from single unique biological samples.

A well-established method for the isolation of concomitant biomacromolecules is based on the simultaneous addition of a monophasic mixture of phenol and guanidine isothiocyanate, commercially available as TRIzol and TRI Reagent (TR), and chloroform to biological samples in order to obtain an aqueous phase containing primarily RNA, an interphase containing DNA and an organic phase containing proteins (Chey *et al.*, 2011; Chomczynski, 1993; Hummon *et al.*, 2007; Xiong *et al.*, 2011). Individual biomacromolecular fractions are purified from the respective phases and can be subjected to specialised downstream analyses. Adaptation of the standard TR protocol for the additional extraction of small molecules was carried out on plant material by Weckwerth and co-workers (Weckwerth *et al.*, 2004). For this, a solvent mixture of methanol, chloroform and water is first used for the fractionation of small molecules into polar and non-polar metabolites, and the precipitation of biological macromolecules. Polar and non-polar metabolites are retrieved from the aqueous and organic phases, respectively. RNA and proteins are isolated from the remaining pellet following extraction in dedicated buffers and phenol, respectively. However, no genomic DNA fraction was obtained using this method, a need that is particularly important in microbial communities which exhibit extensive genetic heterogeneity (Wilmes *et al.*, 2009b) and for which genomic context is, thus, essential for meaningful interpretation of functional omic data.

Because of the hazardous chemicals involved and the methods, therefore, being unsuited for routine high-throughput laboratory use, chromatographic spin column-based procedures have been introduced for the isolation of concomitant biomacromolecules (Morse *et al.*, 2006; Radpour *et al.*, 2009; Tolosa *et al.*, 2007). These methods rely on the pH- and salt-concentration-dependent adsorption of nucleic acids and proteins to solid phases such as silica or glass. The solid phases are washed and the biomacromolecules of interest are sequentially eluted. Such methods are now available as commercial kits from Qiagen (AllPrep DNA/RNA/Protein Mini kit; QA) and Norgen Biotek (All-in-One Purification kit for large RNA, small/microRNA, total proteins and genomic DNA; NA). The latter has the advantage of offering the ability to deplete the total RNA fraction of small RNA (< 200 nt), which can be analysed separately. To the knowledge of the author, chromatographic spin column-based biomacromolecular isolation has yet to be combined with the extraction of concomitant small molecules.

Here, sample processing, sample cryopreservation, cell disruption by cryomilling, small molecule extraction and biomacromolecular isolation based on chromatographic spin columns are combined to result in a universal methodological framework which allows the standardised isolation of extracellular and intracellular polar and non-polar metabolites, genomic DNA, RNA (divided into large and small RNA fractions) and proteins from single LAO community samples. The performance of the methodology was validated by comparison to widely used exclusive and simultaneous biomolecular isolation methods, and proved its general applicability to diverse mixed microbial communities environmental and biomedical research interest, i.e. river water filtrate and human faeces.

3.2 Materials and methods

3.2.1. Sampling and sample processing

When collecting biological samples for systematic molecular analyses, it is crucial to immediately fix the samples in a state which best preserves the information contained within the system at the time of sampling. If care is not given to this step, the information contained within the DNA, RNA, protein and metabolite fractions may change, due to specific and non-specific degradation, changes in gene expression and protein modifications. Sample fixation directly after sampling is, therefore, a key step in the experimental workflows presented for the three different microbial communities described below.

LAO-enriched mixed microbial community For the development and assessment of the devised biomolecular extraction protocol (Section 3.3.4), a single islet sample taken on 13 December 2010 was used. For the comparative metabolomic analysis of the 3 microbiomes (Section 3.3.5), three technical replicates from D4 islet 4 (D4_IV), sampled on 23 February 2011, were used (Section 2.2.1). For each extraction, technical replicates of 200 mg were sub-sampled from the collected islet using a sterile metal spatula, while at all times guaranteeing that the samples remained in the frozen state.

Following weighing out, 200 mg of LAO-enriched microbial community samples were briefly thawed on ice followed by centrifugation at 18,000 g for 10 min at 4 °C to separate the supernatant (~150 µl; extracellular fraction) from the biomass (intracellular fraction). The intracellular fraction was then immediately refrozen in

liquid nitrogen prior to homogenisation by cryomilling (**Figure 3.1, A**). In contrast, the extracellular fraction immediately underwent metabolite extraction.

Freshwater mixed microbial community Forty litres of river water were collected at a depth of around 1 m from the Alzette river (Schifflange, Luxembourg; 49°30'28.04"N; 6°0'11.48"E). Cells were concentrated by tangential flow filtration using Sartocon silica cassettes (Satorius Stedim Biotek) with a filtration area of 0.1 m^2 and molecular weight cut-offs of 10 kDa at a flow rate of ~1.5 l/min. The concentrated cells were pelleted by ultracentrifugation at 48,400 g for 1 h at 4 °C using a Beckman-Coulter Optimal L-90K ultracentrifuge with a Type 45 Ti rotor (Analis). Following ultracentrifugation, each of the resulting pellets was resuspended in 1 ml of supernatant and transferred into 2 ml Eppendorf centrifugation tubes. Concentrated biomass pellets were then obtained by performing an additional centrifugation step at 14,000 g for 5 min at 4 °C. Pellets were then snap-frozen and stored at -80 °C until the cryomilling step (**Figure 3.1, A**).

Human faecal sample In order to guarantee the integrity of the RNA fraction, a one third dilution (w/v) of faecal samples was carried out immediately after addition of RNAlater solution (Ambion). The samples were homogenised by bead-beating using three stainless steel milling balls (diameter of 4 mm; Retsch) for 5 min at 10 Hz in a Retsch Mixer Mill MM 400. The supernatants obtained by low-speed centrifugation (700 g for 1 min at 4 °C) of the human faeces homogenate resulted in microbial suspensions which were then further processed. Biomass pellets were then obtained from the RNAlater suspension by centrifugation at 14,000 g for 5 min at 4 °C (Fitzsimons *et al.*, 2003). Pellets were stored at -80 °C until the cryomilling step (**Figure 3.1, A**).

Each of the three different microbial community samples were homogenised by cryomilling for 2 min at 30 Hz using two 5 mm and five 2 mm stainless steel milling balls (Retsch) in an Mixer Mill MM 400 (Retsch; **Figure 3.1, A**). The cryomilling (or cryogenic grinding) step involves breaking up the structure of the biological material and allows homogenisation of the sample. The frozen homogenous powder obtained at the end of this step guarantees a sufficiently large surface area for the solvents to efficiently extract the small molecules. The processing, described below, is of fundamental importance for efficient metabolite extractions from cellular biomass and subsequent representative metabolic profiling, thereof (Lolo *et al.*, 2007).

3.2.2. Assessment of sample heterogeneity by metabolomics

For the determination of sample heterogeneity by metabolomics (Section 3.3.1), 4 replicates of 200 mg of LAO-enriched microbial communities (technical replicates, i.e. different sub-samples derived from the same islet after sample splitting) were obtained from 4 different islets (biological replicates, i.e. space resolved LAO islets) for 4 different dates (D1, 4 October 2010; D2, 25 October 2010; D3, 25 January 2011; and D4 23 February 2011; Section 2.2.1), leading to an overall sample set of 64. The different sampling dates were chosen to reflect different LAO-enriched microbial community compositions (Section 2.3.2.1).

3.2.3. Determination of cell numbers

Three different smears of LAO-enriched and freshwater mixed microbial communities were prepared and cell counts were obtained for 10 fields of view per smear. The samples were stained with SYBR Green I (Section 2.2.3). The sampled LAO-enriched

microbial communities, typically, contained $6.63\ 10^{8} \pm 2.04\ 10^{8}$ microorganisms per ml, whereas the concentrated river water samples contained $1.48\ 10^{9} \pm 1.08\ 10^{7}$ microorganisms per ml.

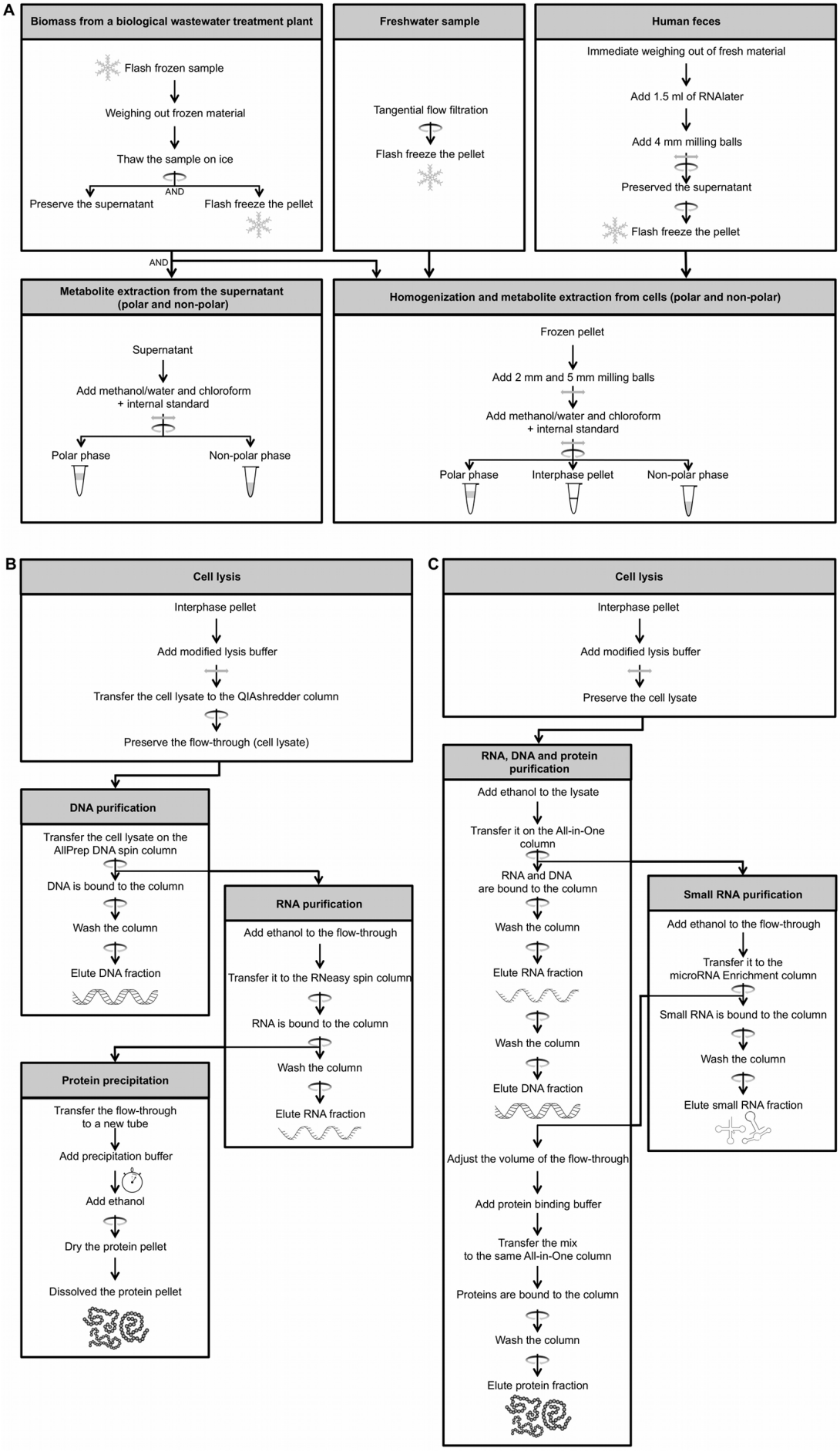
A
Biomass from a biological wastewater treatment plant
Flash frozen sample
Weighing out frozen material
Thaw the sample on ice
AND
Preserve the supernatant
Flash freeze the pellet
Freshwater sample
Tangential flow filtration
Flash freeze the pellet
Human feces
Immediate weighing out of fresh material
Add 1.5 ml of RNAlater
Add 4 mm milling balls
Preserved the supernatant
Flash freeze the pellet
AND
Metabolite extraction from the supernatant (polar and non-polar)
Supernatant
Add methanol/water and chloroform + internal standard
Polar phase
Non-polar phase
Homogenization and metabolite extraction from cells (polar and non-polar)
Frozen pellet
Add 2 mm and 5 mm milling balls
Add methanol/water and chloroform + internal standard
Polar phase
Interphase pellet
Non-polar phase
B
Cell lysis
Interphase pellet
Add modified lysis buffer
Transfer the cell lysate to the QIAshredder column
Preserve the flow-through (cell lysate)
DNA purification
Transfer the cell lysate on the AllPrep DNA spin column
DNA is bound to the column
Wash the column
Elute DNA fraction
RNA purification
Add ethanol to the flow-through
Transfer it to the RNeasy spin column
RNA is bound to the column
Wash the column
Elute RNA fraction
Protein precipitation
Transfer the flow-through to a new tube
Add precipitation buffer
Add ethanol
Dry the protein pellet
Dissolved the protein pellet
C
Cell lysis
Interphase pellet
Add modified lysis buffer
Preserve the cell lysate
RNA, DNA and protein purification
Add ethanol to the lysate
Transfer it on the All-in-One column
RNA and DNA are bound to the column
Wash the column
Elute RNA fraction
Wash the column
Elute DNA fraction
Small RNA purification
Add ethanol to the flow-through
Transfer it to the microRNA Enrichment column
Small RNA is bound to the column
Wash the column
Elute small RNA fraction
Adjust the volume of the flow-through
Add protein binding buffer
Transfer the mix to the same All-in-One column
Proteins are bound to the column
Wash the column
Elute protein fraction

Figure 3.1 Overview of the described methodological workflow. (**A**) Pre-processing and metabolite extraction, (**B**) Isolation of DNA, total RNA and proteins using the Qiagen AllPrep-based procedure, (**C**) isolation of DNA, total RNA and protein using the Norgen Biotek All-In-One-based procedure. The circular arrows denote a centrifugation step and the horizontal arrows symbolise a bead-beating step.

3.2.4. Metabolite extractions

Depending on the physical characteristics of the sample and the pre-processing, metabolite extractions can be performed separately on extracellular and/or on intracellular cell compartments.

Extracellular metabolite extractions Extracellular metabolite extractions were only carried out on supernatant from the LAO-enriched microbial communities (Section 3.3.4**)**. For the river water filtrate and human faecal samples, supernatants were not obtainable, because of the need for concentrating the river water sample by tangential flow filtration and the very limited liquid content in the human faecal samples, respectively. Extracellular metabolite extraction was performed on 150 µl of supernatant. The extraction involved the addition of 300 µl of chloroform and 150 µl of methanol (both at -20 °C; high purity; VWR and BioSolve). The methanol contained ribitol (Sigma-Aldrich) at 0.12 µg/ml as an internal standard for the polar phase metabolomic analyses. The mixture was vortexed at 800 rpm for 10 min at 4 °C in a Thermomixer (Eppendorf) and centrifuged at 14,000 g at 4 °C for 5 min. For the LAO-enriched microbial communities, extracellular proteins are present in the supernatant after the high-speed centrifugation and can be recovered from the

interphase pellet after extracellular metabolite extraction by trichloroacetic acid TCA/acetone precipitation (Wilmes & Bond, 2004).

Intracellular metabolites extraction Intracellular metabolites were extracted by sequentially adding 300 µl of a 1:1 (v/v) methanol/pure water mixture (containing ribitol at 0.12 µg/ml) at -20 °C and 300 µl of chloroform at -20 °C to the cryomilled biomass pellet (**Figure 3.1, A**). The extraction was performed by cold (4 °C) bead-beating the samples using stainless steel balls (the same as previously used for sample homogenization, section 3.2.1) for 2 min at 20 Hz in a Retsch Mixer Mill MM400. After centrifugation at 14,000 g for 10 min at 4 °C, a separation was observed for three phases. Metabolite extractions resulted in an upper phase comprising polar metabolites, an interphase pellet comprising genomic DNA, large and small RNA, proteins and non-lysed cells, and a lower phase containing non-polar metabolites. Defined volumes (Section 3.2.7) of both polar and non-polar metabolites extracts were dried in specific GC glass vials before metabolomic analyses (Section 3.2.7).

3.2.5. Sample processing and biomacromolecular isolations

After removal of the respective metabolite fractions (Section 3.2.4), the interphase pellet (along with the steel milling balls) was kept on ice for the subsequent total RNA (enriched in large RNA), genomic DNA, small RNA and protein sequential isolations and purifications using the different methods as specified below. As the described metabolite extraction is a major modification of the typical extraction workflow, the interphase pellet was lysed in the respective lysis buffers by bead-beating at 30 Hz for 30 sec at 4 °C (Retsch Mixer Mill MM 400) with stainless steel balls (the same as previously used for sample cryomilling and metabolite extraction steps, Sections 3.21

& 3.2.4, **Figure 3.1, A**). Cell disruption, comprising both sample homogenisation and cell lysis, is an early and fundamental step in any biomolecular isolation methodology. Both chemical and mechanical/physical methods are available for cell disruption. However, in natural microbial communities, due to the presence of many different microorganisms with vastly differing cellular properties and due to the presence of different interfering matrixes, chemical and/or enzymatic lysis, by themselves, are typically ineffective at comprehensive and reproducible cell lysis. In the, herewith, presented methodology, the mechanical method of cryogenic grinding was combined with chemical lysis to result in indiscriminate cell lysis. These steps are essential to guarantee the efficiency of cell lysis, as well as the quality of the obtained biomolecular fractions.

NorgenBiotek All-in-One Purification kit-based method (NA) The NorgenBiotek All-in-One Purification kit-based biomacromolecular isolation method (NA**,** NorgenBiotek Corporation; **Figure 3.1, C**) was applied to the interphase pellet according to the manufacturer's instructions with a few important modifications. The standard NA lysis buffer was modified by the addition of β-mercaptoethanol at 10 µl/ml to prevent RNA degradation. One volume of 1x Tris-EDTA was further added to three volumes of the NA lysis buffer included in the kit. Lysis was carried out by bead-beating the interphase pellet in the modified lysis buffer using the stainless steel balls (the same as used previously in the metabolite extraction step, Section 3.2.4). The lysate was then mixed with 100 µl of pure ethanol (analytical grade; Sigma-Aldrich) and was loaded onto an All-in-One chromatographic spin column. The increased quantity of ethanol added resulted in a higher efficiency of nucleic acid binding according to our own experiments. In this step, total (mainly large) RNA, genomic DNA and a small part of the proteins are bound to the column

while RNA smaller than 200 nt and the majority of proteins are found in the flow-through. The bound biomolecules (mainly large RNA and genomic DNA) were sequentially eluted off the column two times with dedicated solutions according to the manufacturer's instructions resulting in the total (mainly large) RNA and DNA fractions, respectively. The first flow-through containing proteins and small RNA was loaded onto the NA-specific microRNA Enrichment Column allowing for the purification of RNA smaller than 200 nt (small RNA fraction). The flow-through from the microRNA Enrichment Column was adjusted to pH 3 and loaded back onto the first spin column in order to bind the proteins to the column. The bound proteins were finally washed and eluted from the column two times using the dedicated solution (protein fraction).

The Qiagen AllPrep DNA/RNA/Protein Mini kit-based method (QA, Qiagen, **Figure 3.1, B**) was applied to the interphase pellet according to the manufacturer's instructions. Following the cryomilling and metabolite extraction steps, a lysate was obtained by bead-beating the interphase using the same stainless steel balls as used previously (metabolite extraction step) in 600 µl of the QA lysis buffer supplemented with β-mercaptoethanol (10 µl/ml of buffer). The lysate was first passed through a QIAshredder column. The resulting flow-through was subsequently passed through an Allprep DNA spin column, which allows the selective binding of genomic DNA. The genomic DNA was then eluded using the dedicated QA buffer (DNA fraction). According to the manufacturer's instructions, 400 µl of ethanol were added to the flow-through and the mixture was loaded onto the membrane of an RNeasy spin column. The RNA was eluded using the dedicated buffer (RNA fraction). The supplied aqueous protein precipitation solution was then added to the flow-through

from the RNeasy spin column for the isolation of the total protein fraction. The resulting protein pellet was re-dissolved in the dedicated buffer (protein fraction).

The TRI Reagent-based method (TR, Sigma-Aldrich) was directly applied to the interphase pellet according to the manufacturer's instructions with a few important modifications. Following the metabolite extraction steps, cell lysis was performed on the interphase pellet following the addition of TR reagent and bead-beating using the stainless steel balls (the same as used previously in the metabolite extraction step). The lysate was mixed with chloroform and centrifuged at 12,000 g for 15 min at 4 °C, which yielded three fractionation phases. RNA was precipitated from the aqueous phase (top layer) by addition of 500 µl of isopropanol, washed and redissolved in RNase free water (RNA fraction). DNA was precipitated from both the interphase (middle layer) and organic phase (lower layer) by addition of 300 µl of ethanol. The precipitated DNA was washed and then re-dissolved in 8 mM NaOH solution (DNA fraction). Proteins were subsequently precipitated from the supernatant phenol-ethanol phase by the addition of isopropanol, washed and dissolved in a mixture of urea-tris-HCl/1 % SDS (1:1; v/v; protein fraction; Hummon *et al.*, 2007).

3.2.6. Reference methods

In order to qualitatively and quantitatively assess the biomolecular fractions obtained through the sequential and simultaneous biomolecular isolation protocols, widely used dedicated biomolecular extraction and purification methods were used as references. In each case, the reference methods were applied to 200 mg of LAO-enriched biomass of a single biological replicate sampled on 13 December 2010.

DNA extraction DNA extraction was performed using the PowerSoil DNA isolation kit (Section 2.2.5).

RNA extraction. RNA extraction was performed using an RNeasy Mini kit (Qiagen) including the optional DNase treatment, to eliminate contaminating genomic DNA according to the manufacturer's instructions. Samples were lysed by bead-beating for 2 min at 30 Hz using two 5 mm and five 2 mm stainless steel milling balls (Retsch) in an Mixer Mill MM 400, in 600 µl of the dedicated lysis buffer, modified by the addition of highly denaturing guanidine-thiocyanate containing buffer supplemented with β-mercaptoethanol (10 µl/ml of buffer), which contributes to the inactivation of RNases. The lysate was homogenised by first passing it through a QIAshredder column. This column simultaneously removes insoluble material and reduces the viscosity of the lysates by disrupting gelatinous material. 400 µl of 70 % ethanol was then added to the flow-through to ensure appropriate binding conditions for the total RNA to a silica-based membrane. An optional DNase treatment was carried out to eliminate contaminating genomic DNA. The total RNA bound to the membrane was washed, and addition of a high-salt buffer allowed the elution of RNA longer than 200 nt.

Protein extractions Protein extractions were carried out using a metaproteomic extraction method for activated sludge (Wilmes & Bond, 2004). Briefly, an LAO-community sample was harvested by centrifugation at 10,000 g for 5 min at 4 °C and washed two times with 0.9 % (w/v) NaCl solution. The resulting pellet was resuspended in 50 mM Tris-HCl, pH 7.0. After centrifugation, the cell pellet was mixed with 1 ml Urea-Thiourea-CHAPS (UTCHAPS) sample buffer and placed on ice for 1 hour. Cells were subsequently lysed by three passes through a French Press

(Thermo Scientific) at 900 psi. The lysate was spun down and the proteins were isolated from the supernatant by trichloroacetic acid (TCA, 20 %, w/v) precipitation, then re-suspension in the low stringency buffer (Wilmes & Bond, 2004).

3.2.7. Metabolomics

Here, identical procedures were used for the analysis of both intracellular and extracellular metabolite fractions. For a more detailed description of procedures used please refer to Section 1.2.2.4.

Determination of sample heterogeneity For the determination of sample heterogeneity by comparative metabolomics and other omic analyses (Section 3.3.1), temporally and spatially resolved (Section 2.2.2) LAO-enriched microbial communities were obtained from 4 different dates and one single date, GC-MS measurements were performed on intracellular polar extracts from the LAO-enriched microbial communities, obtained as specified above. A pool, from which analytical replicates were obtained, was also prepared by combining 100 µl of each of the 64 polar extracts (Section 3.2.2). Aliquots of 40 µl of intracellular polar extracts and of the pool were dried under vacuum at -4 °C using a refrigerated CentriVap Concentrator (Labconco Corp) and analysed after resuspension and derivatisation. For all samples, the polar extracts were evaporated in specific GC glass vials (Chromatographie Zubehör Trott), whereas, the non-polar phases were dried under an extractor hood overnight at room temperature. Each pool aliquot represented an analytical replicate. For GC-MS analysis, the sample sequence in the autosampler was randomised and between each four samples analytical replicates and blanks were introduced.

Comparative metabolomic analysis For the comparative metabolomic analysis of the three different microbial communities (Section 3.3.5), 50 µl, 10 µl and 5 µl of the intracellular polar and non-polar phase extracts of the LAO-enriched microbial community, river water filtrate and diluted human faecal samples, were dried and analysed, respectively. To prevent overloading of the GC column, a 1 in 10 dilution in a 1/1 (v/v) methanol/water mixture was previously carried out on the raw polar phase extracts (Section 3.2.4) derived from the human faecal samples.

Metabolomic analyses of the representative LAO-enriched microbial community samples For the metabolomic analyses of the representative LAO-enriched microbial community samples (Section 3.3.4), 50 µl of the polar and non-polar phase extracts of the extracellular and intracellular compartments were dried, as specify above. Metabolite derivatisation was performed using an Agilent 7693 Autosampler (Agilent Technologies).

Derivatisation Dried polar metabolites were dissolved in 15 µl of 2 % methoxyamine hydrochloride in pyridine (Pierce) at 45 °C. After 30 min, an equal volume of MSTFA (2,2,2-trifluoro-N-methyl-N-trimethylsilyl-acetamide) + 1 % (v/v) TMCS (chloro-trimethyl-silane) was added and held for 30 min at 45 °C. Dried non-polar metabolites were dissolved in 20 µl MSTFA, including 1 % TMCS and incubated for 30 min at 45 °C before analysis.

GC-MS analysis GC for polar and non-polar samples was performed using an Agilent 6890 GC equipped with a 30 m DB-35MS capillary column. The GC was connected to an Agilent 5975C MS operating under electron ionization at 70 eV. The GC oven temperature was held at 80 °C for 6 min and increased to 300 °C at 6 °C/min. After 10 min, the temperature was increased to 325 °C at 10 °C/min for 4 min. The MS source

was held at 230 °C and the quadrupole at 150 °C. The analyser was operated in scan mode and 1 µl of derivatised sample was injected in splitless mode at 270 °C. Helium was used as carrier gas at a flow rate of 1 ml/min. The run time for one sample was 59 minutes.

For the present work, the acquired GC-MS data, were deconvoluted and matched to a dedicated in-house metabolite library and the Golm Metabolome database (GMD) using the MetaboliteDetector software (Hiller *et al.*, 2009). The raw GC-MS data were filtered so that only metabolites present in more than 70 % of technical replicates were considered. This stringent cut-off was chosen to initially result in an even number of detected metabolites for the individual biological replicates and, thus, restrict inclusion of metabolites resulting from measurement noise.

3.2.8. Biomacromolecular quality and quantity assessment

Isolated genomic DNA was separated by electrophoresis on an agarose gel (Section 2.2.5). PCR amplifications were carried out for each of the triplicate DNA fractions obtained using the NA-based method from the different microbial community samples, without any additional DNA purification steps, using previously defined 16S rRNA primers (Section 2.2.6). PCR reactions were prepared in 20 µl volumes containing 10 µl of 2x Phusion Master Mix (Finnzymes), 500 nmol/l of the forward and the reverse primers and 20 ng of DNA. After an initial 3 min denaturation step at 98 °C, samples were amplified during 35 cycles of 20 sec at 98 °C, 15 sec at 44 °C and 1 min at 72 °C, where the last cycle was followed by a final 7 min elongation step at 72 °C. Amplified DNA was separated by electrophoresis on a 1 % agarose gel containing 0.5 ‰ ethidium bromide (Sigma). MassRuler DNA ladder mixture

(Fermentas GmbH) was loaded onto the gels for size estimation. Gels were visualized on a BioDocAnalyze (BDA) gel imaging and analysis system (Biometra).

For RNA quality assessment and quantification, an Agilent 2100 Bioanalyser (Agilent Technologies) was used. The Agilent RNA 6000 Nano kit and Agilent Small RNA kit for prokaryotes were used. In order to compare different RNA fractions, traces obtained using the Agilent 2100 Bioanalyser, i.e. fluorescent units (FUs), were normalised by the total integrated peak intensities. To correlate the RNA migration time to nucleotide size, marker and ladder peaks were used as reference points. The polynomial trend curves of the second degree for the Nano kit and of the third degree for the Small RNA kit were used to transform migration times into molecular size (nucleotides).

Genomic DNA and RNA fractions were quantified and assessed using a NanoDrop Spectrophotometer 1000 (Thermo Scientific).

Protein extracts were separated using 1D SDS-PAGE (Section 2.2.8). The protein fractions were quantified using a 2-D Quant kit (GE Healthcare).

Peptide tandem mass spectra were obtained on a MALDI-ToF/ToF analyser, after tryptic digestion of a dominant protein band excised from a SDS-PAGE gel on which an NA-based LAO-enriched microbial community protein extract had been separated (Section 2.2.8**)**.

The mass to charge (m/z) ratios of each spectrum were converted into discrete m/z-values and intensities in order to reduce the size of the dataset and computational time (Armengaud, 2013). Selected mass spectra were *de novo* annotated using the PepSeq algorithm (Chaurand *et al.*, 1999).

3.2.9. Determination of intact cells versus cells with a compromised cell membrane

To evaluate the efficiency of cell lysis and variation between biomolecular extraction protocols on representative LAO-enriched microbial community samples, the percentage of damaged cells was determined by epifluorescent microscopic quantification after staining using the Live/Dead BacLight Bacterial Viability kit (Invitrogen). Cells were washed once with phosphate buffer saline solution (1x PBS, pH 7.2) prior to staining. Following staining bacteria with intact cell membranes display green fluorescence under UV light, whereas, bacteria with damaged membrane exhibit red fluorescence. Cells were first observed at 1,000X magnification on a Leica DMR Fluorescence Microscope coupled to a DFC500 camera (Leica). Images were captured using the Leica Application Suite software package. For the determination of the percentages of intact cells versus cells with a compromised membrane, the red and green fluorescence micrographs were obtained and processed using the *daime* software package (Daims *et al.*, 2006a). Ten images were analysed from each sample having undergone the different treatments, i.e. either the freeze-thaw cycle (FT), the additional metabolite extraction (M) and/or different lysis treatments (NA, QA, TR, RM-A, RM-B and RM-C). The percentage of red pixels was determined as follows: in the red channel, biomass was automatically segmented from the background by the edge detection algorithm (dark threshold set as 50). The created object layer was transferred into the green image channel and these pixels were excluded for the next step. Green pixels were then selected using the Magic Wand tool with a tolerance of 50 %. The numbers of red and green pixels in the respective channels were then counted using the objects measurement tool.

3.2.10. Assessment of sample heterogeneity by different molecular analyses

To assess the extent of heterogeneity at each biomolecular information layer (Section 3.3.1), a comparative experiment was carried out on spatially resolved LAO-enriched microbial community samples obtained from 4 biological replicates and four technical replicates of a single biological replicate (islet) sampled on a single date (D4; 23 February 2011).

16S rRNA amplicon sequencing The phylogenetic composition was assessed by 16S rRNA amplicon sequencing using a Genome Sequencer FLX instrument (454 Life Sciences) using specific primers for the amplification of 400 bp and spanning hypervariable regions from V3 and V6 of the 16S rRNA gene. PCR reactions were prepared in 50 µl volumes containing 40 µl of PCR reaction mix, 400 nmol/l of the forward and the reverse primers (**Table 3.1**) and 18 ng of DNA. 10x PCR buffer without $MgCl_2$ (Invitrogen), 2.5 mM $MgCl_2$, 0.5 mM dNTP mix, 0.067 U/µl Platinum Taq DNA Polymerase (Invitrogen), and molecular grade H_2O using the following thermocycling conditions: 90 sec at 95°C for initial denaturation and uracil-N-glycolase inactivation, 30 sec at 95°C for denaturation, 30 sec at 62°C for annealing, 30 sec at 72°C for extension, with the annealing temperature decreasing by 0.3°C for each subsequent cycle for 19 cycles, followed by 10 cycles of amplification consisting of 30 sec at 95°C for denaturation, 30 sec at 45°C for annealing, 30 sec at 72°C for extension, and a final extension for 7 min at 72°C and cool down to 15°C. The resultant fusion PCR product were analysed using 1% E-Gel 96 Agarose (Invitrogen) to confirm PCR amplification and product band size. The pooled barcoded 16S rRNA

gene amplicon library underwent emulsion PCR, bead enrichment and recovery, and pyrosequencing analysis on the Genome Sequencer FLX instrument.

Table 3.1 Oligonucleotide primers used for 16S rRNA region V3V6 amplification by PCR.

Primers	Sequence 5'-3'	Tm (°C)
Forward[§]	CCTACGGGDGGCWGCA	59.6
Reverse[§]	CTGACGACRRCCRTGCA	47.8

[§]Liu *et al.*, 2013

The bioinformatic pipeline consisted of first sorting the FASTA file, clustering the sequences at a 99 % threshold using the cluster utility and executing the UCHIME utility in *denovo* mode in order to identify chimeric clusters. Both clustering and UCHIME are utilites present in the USEARCH package (Edgar *et al.*, 2011). Once chimera checking was completed, all the non-chimeric clusters were selected and pooled into a single FASTA sequence. The resulting sequences were trimmed and filtrated using the split_libraries.py utility from the Qiime pipeline (Caporaso *et al.*, 2010a), and classified to the lower taxonomic level until genus level at the bootstrap confidence levels of ≥ 95 % using a web service for the Naïve Bayesian Classifier, made available by the Ribosomal Database Project (Cole *et al.*, 2009), release 10, update 28 and uploaded to an inhouse SQL database for future access and queries. The resulting data were normalised by the total read counts before statistical testing.

Metagenomic Functional metagenomic was carried out using raw metagenomics Illumina-sequence data (average length of 101 bp, Section 4.2.4) loaded into MG-RAST (Meyer *et al.*, 2008), filtered and trimmed, then annotated with KOs (KEGG data base; orthologous genes; e-value 10^{-3}, 50 % similarity, 15 bp for minimum

alignment length cutoff). The normalised KOs abundance was considered for statistical analysis.

Metabolomics Metabolic data were generated by gas chromatography coupled to mass spectrometry (GC-MS; as described before, Section 3.2.7).

3.2.11. Data treatment and statistical analyses

In order to determine sample heterogeneity by comparative metabolomics, metabolite spectral intensities were normalised by the sum of the total intensity measured for each replicate sample. Only metabolites which were detected in each of the 16 analytical replicates (pool samples) were further considered. To account for instrument drift, sample intensities were further normalised by dividing through the mean intensity of two preceding and two subsequent analytical replicates (pool samples). Samples were then re-normalised by the sum of the total intensity. Outliers were defined as technical replicates having a negative within-biological-replicate-group concordance correlation coefficient (CCC) and removed from the dataset. For the sample heterogeneity dataset (**Supplementary Table II**), this filter resulted in the removal of a single technical replicate from islet 1 of sampling date 2 (D2_I1).

The ade4 package of the R software (R version 2.13.2, http://www.r-project.org/) was used for the principal component analysis (dudi.pca function). The vegan package in R (R version 2.13.2, http://www.r-project.org/) was used to calculate Bray-Curtis dissimilarity matrices (vegdist function) and the data was represented by principal coordinate analysis (PCoA; pcoa function) using the ape package. The epiR package was used for calculating the concordance correlation coefficients (epi.ccc function). Hierarchical clustering was performed in Multiple Experiment Viewer (MeV) v4.6 on

mean centred metabolite abundances, using the Pearson product moment correlation coefficient and average linkage clustering.

Statistical significance (*P* values) was calculated by Kruskal-Wallis one-way analysis of variance using R with an alpha level of 0.05 and *n* variables. This non-parametric test was chosen due to the relatively low number of variables for each condition.

To highlight potential signature metabolites for the three analysed microbial communities, intensity values for polar and non-polar metabolites for each microbiome were combined (**Supplementary Table VI**), the mean intensity values calculated and converted to percentages of the respective community metabolomes. The percentages were subjected to a three-way comparison using the triangle.plot function of the ade4 package in R.

3.3. Results and discussions

3.3.1. Analysis of sample heterogeneity by comparative omics

The metabolome represents the final output that results from the cellular interactions of the genome, transcriptome, proteome and environment and, thus, is the most sensitive indicator of cellular activity and sample-to-sample variation. In order to assess the extent of heterogeneity within microbial communities, an initial metabolomics experiment was performed on spatially and temporally resolved LAO-enriched microbial communities (Section 2.2.1).

For the four different sampling dates (D1, D2, D3 and D4; as defined in Section 2.2.2), intracellular polar metabolites were extracted from four different biological replicates, i.e. different islets (**Figure 2.2**), and per islet, four technical replicates (different sub-samples derived from the same islet) were extracted. In addition, an analytical reference sample for quality control and normalisation, consisting of a pool of all polar metabolite extracts (Dunn *et al.*, 2011), was prepared using the remaining polar extracts from all the samples (Section 3.2.7).

The resulting metabolite extracts were analysed by GC-MS (Section 3.2.7) and the peak intensities of detected metabolite were integrated using specialised software and databases (Section 3.2.7, **Supplementary Table II**). The data obtained for sixteen analytical replicates highlights the high reproducibility of the GC-MS method [mean concordance correlation coefficient (Lin, 1989) = 0.993 ± 0.006; mean CCC ± s.d.].

Because of the apparent unevenness of total integrated peak intensities in the metabolomics data, which results from inherent sample heterogeneity, a unit vector

normalisation strategy, analogous to that commonly used for shotgun proteomics experiments (Florens *et al.*, 2006), was chosen. This involves dividing each metabolite peak intensity by the total peak intensity of the sample (Wilmes *et al.*, 2010). To further exclude experimental variation, the metabolomics dataset was filtered according to presence in all analytical pool replicates. Finally, in order to account for instrument drift, the data were normalised, using the pool replicate data (Dunn *et al.*, 2011) and finally the data were re-normalised to the total intensity (Section 3.2.10).

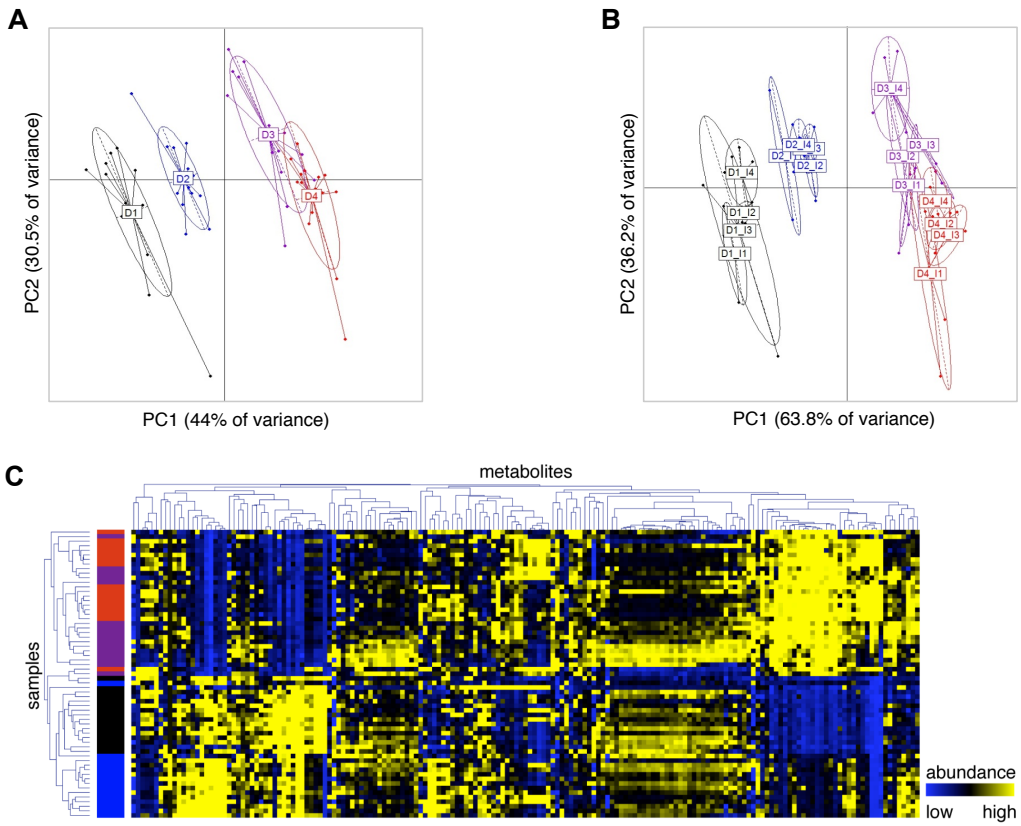

Figure 3.2 Metabolome heterogeneity within LAO-enriched microbial community samples. (**A**) Scatter plot of the two first principal components obtained using Principal Component Analysis (PCA) of the normalised metabolomics data derived from the four biological replicates (islets; I1-I4), for each of the four distinct sampling dates (D1-D4). Each microbial community metabolome is indicated by a dot and colour-coded according to sampling date. (**B**) Between-class PCA of the individual technical replicates for each biological replicate (islets; I1-I4). (**A**,**B**) The centre of gravity for each date/islet cluster is marked by a rectangle and the coloured ellipse covers 67 % of the samples belonging to the cluster. (**C**) Hierarchical clustering of the normalised metabolomics data using the Pearson product moment correlation coefficient.

Using the normalised metabolomics data (Section 3.2.10), extensive sample-to-sample variation is apparent for both biological (islets) and technical replicates. Although most samples from specific dates are distinguishable by their metabolomic profiles (**Figure 3.2, A**), extensive overlap between biological and technical replicates is

apparent (**Figure 3.2, B**), with numerous samples clustering outside of their respective replicate groups (**Figure 3.2, C**). These results indicate that sample heterogeneity is an important consideration for integrated omic studies of natural microbial consortia.

Furthermore, in order to assess if this heterogeneity was only observable at the metabolome level, intracellular polar metabolites and genomic DNA from 4 biological replicates were extracted from a unique sampling date (D4, 23 February 2011 as defined in Section 2.2.2), including one biological replicate (D4_I) split into 4 technical replicates. From these analyses, sample-to-sample variation is apparent at each level of biomolecular information (**Figure 3.3**). Interestingly, the ability to discriminate between biological and technical replicates decreases from phylogenetic profiles (**Figure 3.3, A**) to metagenomes (**Figure 3.3, B**) and, lastly, metabolite profiles (**Figure 3.3, C**). This finding reflects increased complexity at each level of biomolecular information. The resulting fine-scale heterogeneity underlines the requirement for the isolation of concomitant biomolecules to fulfill the premise of systematic measurements and enable meaningful integration of space- and time-resolved microbial community omic datasets.

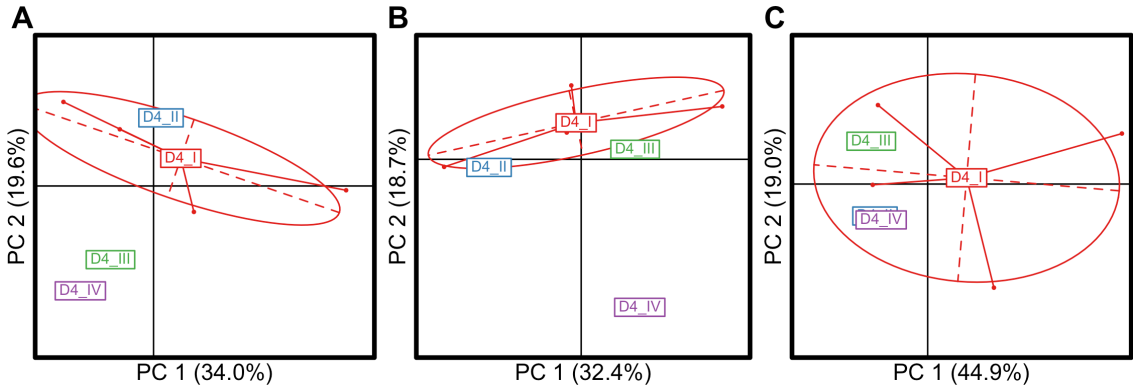

Figure 3.3 Within- and between-sample heterogeneity using biological and technical replicates of LAO-enriched microbial communities. Scatter plots of the results of principal coordinate analysis of Bray-Curtis dissimilarity indices of 4 biological replicates (D4_I, D4_II, D4_III, D4_IV), from a unique sampling date, including one biological replicate (D4_I) split into 4 technical replicates (in red). Each biological replicate is indicated by a rectangle, and each technical replicate is represented by a dot. The ellipse covers two-thirds of the technical replicates which are derived from the same biological replicate, with the rectangle indicating their common center of gravity. (**A**) Community compositions of the samples as revealed by 16S rRNA amplicon sequencing (Section 3.2.2**, Supplementary Table III**). (**B**) Corresponding functional capacities determined using KEGG orthologous group annotations of metagenomic sequence data (Section 3.2.2**, Supplementary Table IV**), and (**C**) Corresponding polar metabolomes (Section 3.2.7**, Supplementary Table II**).

3.3.2 Conceptualisation of a biomolecular isolation framework

In the context of dealing with extensive sample-to-sample and endogenous sample heterogeneity, understanding a complex system's structure and dynamics requires integrated analyses of its parts which, in the present case, demands the isolation of concomitant metabolites, nucleic acids and proteins from single unique samples prior to the individual fractions being subjected to specialised metabolomic, genomic, transcriptomic and proteomic analyses (**Figure 3.1**). Major considerations to allow

comprehensive and reproducible extraction and purification of biomolecules from mixed microbial communities are (i) standardised and representative sample preservation, (ii) indiscriminate cell lysis, and (iii) retrieval of high quality biomolecular fractions. These constraints were taken into account when devising the biomolecular isolation framework and resulted in a methodology, which is based on (i) immediate snap-freezing of mixed microbial community samples in liquid nitrogen following sampling (or after sample-specific cold preprocessing) and preservation of the biomass at a minimum of -80 °C, (ii) mechanical (cryogenic) and chemical cell lysis, and (iii) reliance on robust methods for biomolecular extraction and purification. The resultant methodological framework offers a fully integrated workflow which can easily be adjusted for specific samples and which allows flexibility for use of sequential or dedicated biomolecular extraction and purification protocols. For this doctoral research, the methodology was developed and validated on a representative LAO-enriched microbial community sample.

3.3.3. Cell lysis efficiency

Conservation of cell integrity before sample processing and representative cell lysis prior to biomolecular extraction are essential considerations for the methodology to result in reproducible and representative biomolecular fractions. Following snap-freezing in liquid nitrogen, sample preservation at -80 °C and thawing on ice, almost all cells are left intact (**Figure 3.4, A&D**). The conceptualised methodological framework relies on an initial cryomilling homogenisation and lysis step followed by polar and non-polar metabolite extractions. These steps result in indiscriminate and comprehensive cell disruption and lysis (**Figure 3.4, B&D**). Following the further

addition of dedicated lysis buffers, the vast majority of cells (90.20 ± 6.46 %; NA-based method mean ± s.d.) are lysed (**Figure 3.4, C&D**). Importantly, as highlighted in Figure 3 D, the lysis efficiencies of the methods allowing sequential biomolecular isolations (NA and QA) compare favourably to widely used reference methods for exclusive biomacromolecular isolation, i.e. no statistically significant differences are apparent (Kruskal-Wallis, $P = 0.189$, $n = 10$), and they significantly outperform the standard TR-based simultaneous biomolecular isolation method (Kruskal-Wallis, $P = 0.002$, $n = 10$). These results validate the chosen approach for sample preservation, homogenisation and cell lysis.

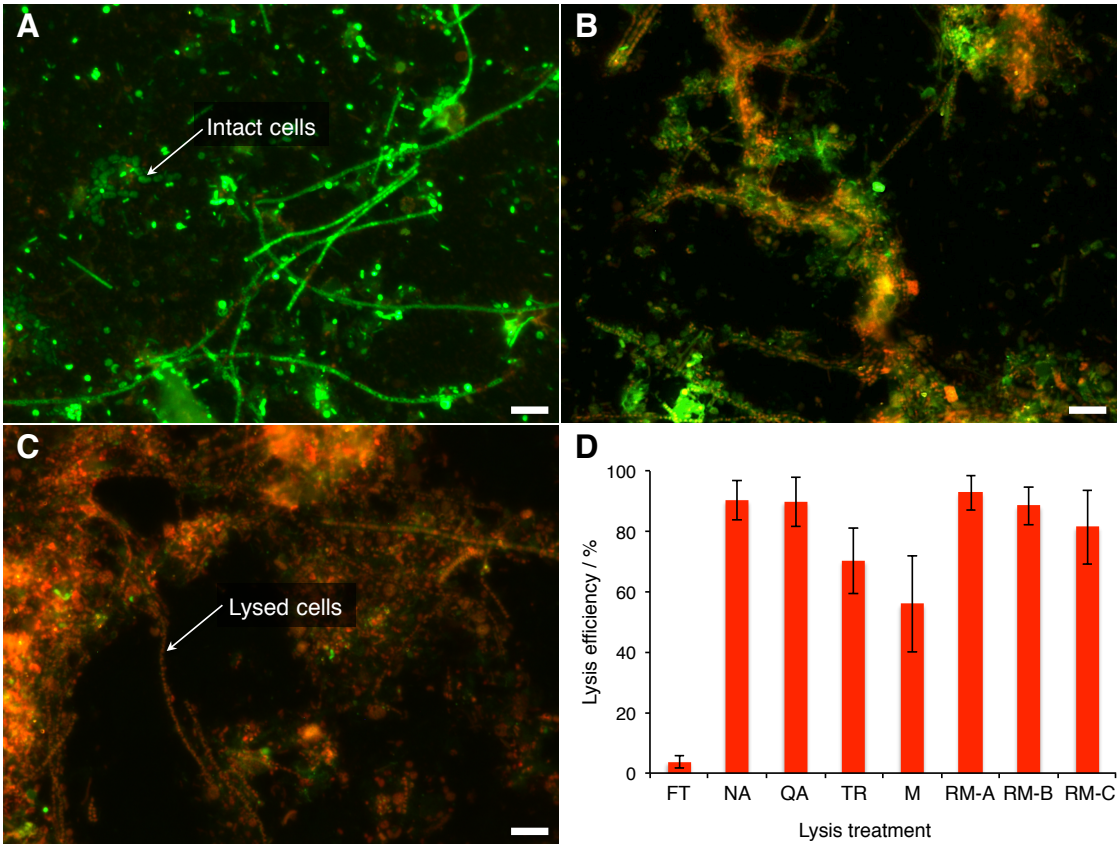

Figure 3.4 Efficiencies of different cell lysis methods. (**A**,**B**,**C**) Representative epifluorescence micrographs of microbial cells from a representative LAO-enriched microbial community sample stained with the Live/Dead stain (intact cells highlighted in green, disrupted cells with a compromised cell membrane in red). Scale bar is equivalent to 10 µm. (**A**) Sample having undergone a single freeze-thaw cycle. (**B**) Sample having undergone polar and non-polar metabolite extractions. (**C**) Sample having undergone the additional mechanical and chemical lysis step using the NA kit's modified lysis buffer. (**D**) Bar chart highlighting the percentages of cells disrupted by the different treatments (n = 10; error bars represent s.d.). x-axis legend: FT, sample having undergone a single freeze-thaw cycle, reflecting pane a; NA, sample having been subjected to cell lysis in the modified NA lysis buffer, reflecting pane c; QA, QA lysis buffer; TR, TR-based lysis; M, sample after polar and non-polar metabolite extractions reflecting pane b; RM-A, sample having been subjected to the dedicated DNA extraction reference method; RM-B, dedicated RNA extraction method; RM-C, dedicated protein extraction method.

3.3.4. Quality and quantity of obtained biomolecular fractions

In addition to the need for efficient cell lysis, the most important consideration is the requirement for obtaining representative high-quality biomolecular fractions. Because metabolites are extracted first, the methodology results in identical results for metabolomic analyses if subsequent biomacromolecular isolations are carried out or not. Metabolomic analysis of the representative LAO-enriched microbial community sample result in clear and reproducible total ion chromatograms (**Figure 3.5**). These allowed the robust detection, using 70 % cut-off (Section 3.2.7) and semi-quantification of 267 polar and 242 non-polar intracellular metabolites as well as 268 polar and 176 non-polar extracellular metabolites (**Supplementary Table V**). Using a spectral reference library, this enabled the combined identification of 62 polar and 30 non-polar metabolites. Consequently, the vast majority (> 85 %) of detected metabolites were not identified in line with earlier metabolomics results obtained on mixed microbial communities (Wilmes *et al.*, 2010).

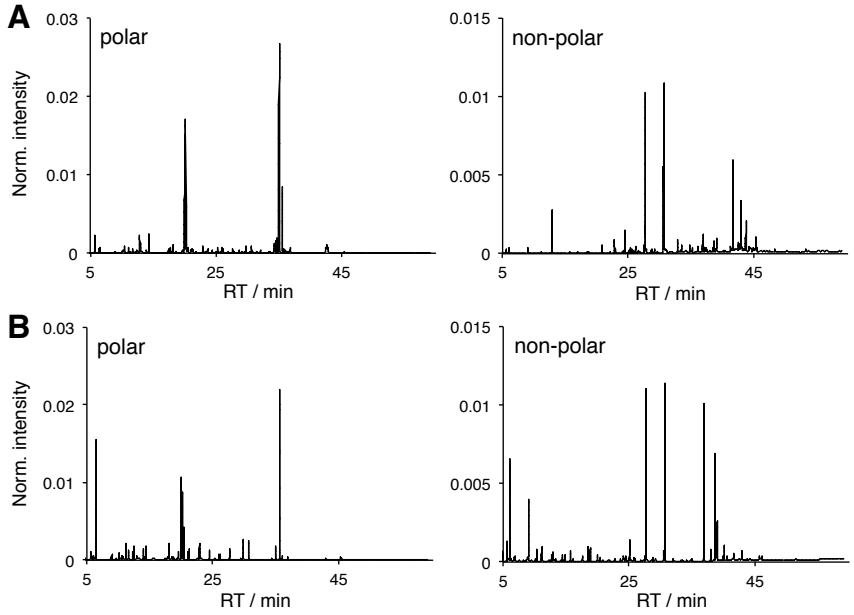

Figure 3.5 Representative GC-MS total ion chromatograms of polar and non-polar metabolite fractions obtained from the (**A**) intracellular and (**B**) extracellular compartments of representative LAO-enriched microbial community samples. Norm., normalised; RT, retention time.

The nucleic acid fractions obtained using the different extraction protocols were firstly analysed by NanoDrop spectrophotometer, the absorbance ratios at 260/280 nm in particular reflecting purity of the respective fractions obtained (**Table 3.2**). The mean absorbance ratios at 260/280 nm for the DNA fractions were between 1.9 and 2.1 for the methods used, generally accepted as "pure" (Manchester, 1995). The only exception is the DNA fraction obtained using the simultaneous biomolecular isolation using the TR-based method for which a poor mean ratio of 1.5 was measured, highlighting the superior performance of the chromatographic spin column-based methods.

Table 3.2 Quality and quantity parameters of the nucleic acid fractions obtained using the different isolation methods applied to a representative LAO-enriched microbial community sample.

Isolation method	RNA			DNA		
	RIN [1]	23S/16S [1]	Quantity / µg [2]	A260/A280 [2]	A260/A230 [2]	Quantity / µg [2]
Reference method [3]	6.60 ± 0.88	1.17 ± 0.21	2.04 ± 0.62	2.02 ± 0.09	1.05 ± 0.34	1.81 ± 0.71
NA-based method, first elution	7.03 ± 1.20	1.32 ± 0.05	8.80 ± 6.87	2.03 ± 0.09	0.11 ± 0.05	5.90 ± 2.91
NA-based method, second elution	5.70 ± 2.16	0.73 ± 0.63	6.64 ± 2.84	1.99 ± 0.11	0.09 ± 0.01	5.46 ± 0.98
QA-based method	9.68 ± 0.05	1.51 ± 0.09	42.93 ± 15.60	2.12 ± 0.11	0.92 ± 0.55	14.43 ± 10.65
TR-based method	7.44 ± 0.28	1.37 ± 0.07	8.86 ± 1.52	1.50 ± 0.43	1.12 ± 0.24	13.67 ± 12.76

[1] The ratio of 23S/16S rRNAs and RIN (RNA integrity number) were determined using the Agilent 2100 Bioanalyzer. RIN values range from 1 (completely degraded) to 10 (intact).

[2] Nucleic acid quantity and absorbance ratios were determined using a NanoDrop spectrophotometer.

[3] The Qiagen RNeasy Mini kit was used as reference method for RNA extraction and the MO BIO PowerSoil DNA isolation kit was used as reference method for DNA extraction.

The absorbance ratios at 260/280 nm of all RNA fractions are between 1.8 and 2.1 (**Table 3.2**), indicating that overall high quality RNA was extracted using the different protocols. The Agilent Bioanalyzer 2100 electropherograms (**Figure 3.6, A**) show distinct peaks between 100 and 4,500 nt representing the total RNA fractions obtained using the different methods. The 23S/16S rRNA ratios vary depending on the RNA extraction method used from 0.7 to 1.5 (**Table 3.2**). However, integrity of the isolated RNA was mainly assessed using the RNA integrity number (RIN) score, which is now commonly accepted as a better RNA quality indicator (Fleige & Pfaffl, 2006; Jahn *et al.*, 2008). The RIN scores obtained with the NA-based method (7.03 ± 1.20) are similar to those obtained using the Qiagen RNeasy Mini kit-based reference method (6.60 ± 0.88; Kruskal-Wallis, $P = 0.806$, $n = 5$) and the sequential biomolecular extraction based on the TR-based method (7.44 ± 0.28; Kruskal-Wallis, $P = 0.352$, $n = 5$) but lower than those obtained for the QA-based method (9.68 ± 0.05; Kruskal-Wallis, $P = 0.011$, $n = 5$), which had a very high and consistent score. The RNA fraction obtained using the TR-based method is particularly enriched in small RNAs, indicative of extensive RNA degradation, which results in a non-representative RNA fraction (**Figure 3.6, A**). Overall, the RNA fractions derived using the simultaneous and sequential biomolecular isolation methods are of sufficient quality for downstream ribosomal RNA removal, reverse transcription and high-throughput cDNA sequencing (He *et al.*, 2010).

The NA-based method allows for the additional subfractionation of the total RNA extracted into a small RNA fraction (**Figure 3.6, B**). As expected, the major components of the small RNA fractions are transfer RNAs (tRNA) with a mean size of around 60 nt and other small RNAs of larger size, represented by multiple peaks around 120 nt, e.g. 5S rRNA. A “miRNA”-like region is observed as a broad peak

ranging from 10 to 40 nt (**Figure 3.6, B**). The "miRNA" region relative to small RNA content ("miRNA"/small RNA ratio) is 26.43 ± 1.97 % for the NA-based method. Because of important regulatory roles fulfilled by small bacterial regulatory RNAs, the small RNA fraction, only provided by the NA-based method, adds an important additional level of information to the integrative molecular analyses which can be carried out using the reported methodology.

The size, the quality (degraded versus intact) and semi-quantitative amounts of DNA extracted were determined by gel electrophoresis (**Figure 3.6, C**). Rather expectably, the dedicated reference method results in the best DNA quality extract (**Figure 3.6, C**, lane RM). The genomic DNA fraction isolated using the sequential biomolecular extraction protocol based on NA provides the most similar results to the reference method. Sequential biomolecular isolation based on QA also results in DNA extracts exhibiting intense and large bands. Importantly, the TR-based extraction method results in poor quality DNA extracts (**Figure 3.6, C**), in concordance with the low absorbance ratios measured at 260/280 nm (**Table 3.2**) and, thus, is rather ill-suited for comprehensive biomolecular isolations.

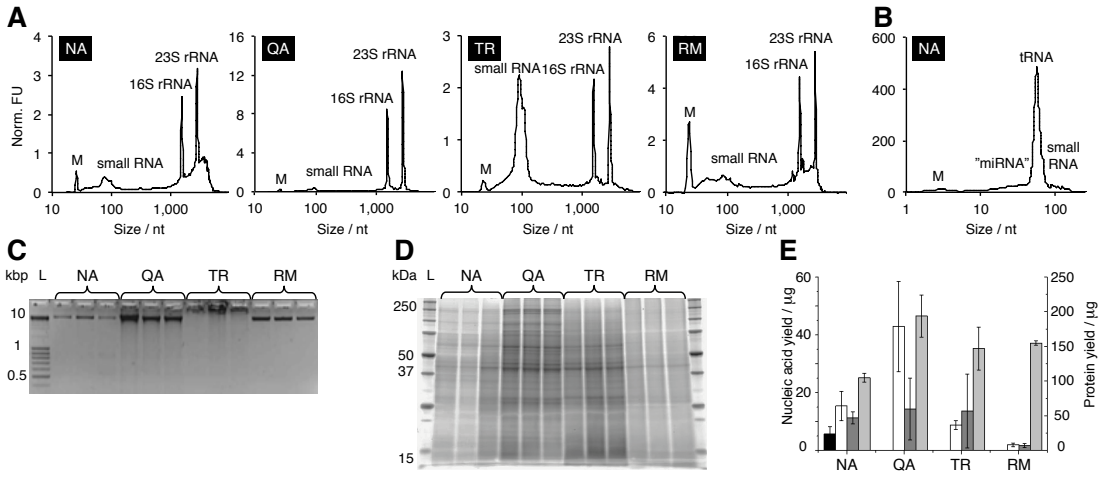

Figure 3.6 Quality and quantity of biomacromolecular fractions isolated from the representative LAO-enriched microbial community sample using either the NA-, QA- and TR-based methods (following prior metabolite extractions) or using the reference methods (no metabolite extractions were carried out before the respective extractions). (**A**,**B**) Representative Agilent Bioanalyzer 2100 electropherograms of the total RNA and small RNA fractions, respectively. (**C**) Agarose gel image highlighting the representative genomic DNA fractions obtained [Mean amount (n = 3) loaded in μg ± st. dev., from right to left; NA: 0.35 ± 0.17; QA: 0.35 ± 0.08; TR: 0.83 ± 0.85; RM: 0.08 ± 0.04] and (**D**) SDS-PAGE image of representative protein fractions [Mean amount (n = 3) loaded in μg ± st. dev., from right to left; NA, first elution: 3.20 ± 0.19; QA: 5.44 ± 1.06; TR: 3.88 ± 0.30; RM: 4.62 ± 0.09]. The arrow and the box represent the dominant gel band which was submitted to tryptic digestion and MALDI-ToF/ToF analysis. (**E**) Biomacromolecular yield obtained for the small RNA, total and large RNA, DNA and protein (first elution) fractions (n = 5, error bars represent s.d.). Abbreviations: NA, NA-based method; QA, QA-based method; TR, TR-based method; RM, reference methods; Norm., normalised; FU, fluorescent unit; M, marker; nt, nucleotides; L, ladder.

Gel electrophoresis of protein fractions (SDS-PAGE) provides a visual representation of the community metaproteomes derived from the representative LAO-enriched microbial community samples (**Figure 3.6, D**). Importantly, in terms of band diversity and clarity, the efficiency of protein extraction is superior for the sequential and

simultaneous isolation protocols (**Figure 3.6 D**, lanes NA, QA and TR) compared to those obtained using the reference method (**Figure 3.6 D**, lane RM). This is most likely due to the removal of "contaminant" biomolecular fractions (e.g. small molecules, nucleic acids) from the protein fraction during the sequential or simultaneous isolation methods. In contrast, some of these biomolecules may be retained in the extracts obtained with the reference method. Consequently, the developed methodology results in higher quality protein extracts than the dedicated protein extraction method.

To further assess the compatibility of the protein fractions obtained with the NA-based method with subsequent downstream analyses, a protein band of approximately 45 kDa was excided from an SDS-PAGE gel (**Figure 3.6, D**) and analysed by MALDI-ToF/ToF following tryptic in-gel digestion. A *de novo* peptide sequence was derived from a clear tandem mass spectrum (**Figure 3.7**). The peptide (*de novo* sequence: VESITAPVVVTEDQTQR) was putatively identified as an exported hypothetical protein isolated from *Microthrix parvicella* (GI: 501187104, e-value = 0.97, score = 34.1, query covery = 100 %, BLASTP against NCBInr). This sequence is similar to the sequence obtained previously (Section 2.3.2.2). This result highlights the predominance of *Microthrix parvicella* in the LAO community as well as the prevalence of this exported protein of unknown function.

Overall, these results highlight the ability to carry out proteomic analyses based on either gel- or liquid chromatography-based separation followed by mass spectrometry on the obtained protein fractions.

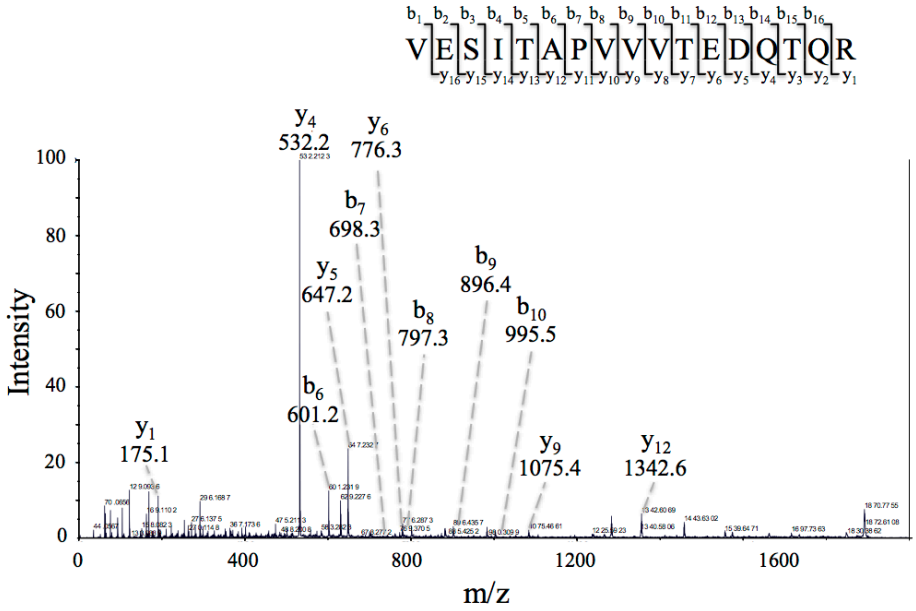

Figure 3.7 Representative mass spectrum obtained by MALDI-ToF/ToF mass spectrometry which allows *de novo* peptide sequencing of the excised protein band derived from the representative LAO community sample. A clean b- and y-ion series is apparent for the peptide VESITAPVVVTEDQTQR (m/z 1872.00).

In terms of yields, larger quantities of nucleic acids were obtained using the chromatographic spin column-based methods (NA and QA), compared to the reference methods (Kruskal-Wallis, $P = 0.009$, $n = 5$; **Figure 3.6, C&E and Table 3.2)**. An analogous finding was previously highlighted in a comparative analysis of RNA yields obtained using a dedicated RNA extraction method and the simultaneous biomacromolecular extraction method based on TR (Chomczynski, 1993). In terms of protein quantity (**Figure 3.6, D&E**), the overall yield was highest for the QA-based method, followed by the exclusive isolation method. However, extensive variation in terms of the quantities of obtained biomolecular fractions is apparent for the QA-based isolation methodology. In contrast, the NA-based method results in the most consistent results for the tested sequential/simultaneous extraction methods.

3.3.5. Broad applicability of the methodological framework

The broad applicability of the developed methodology was demonstrated by applying the NA-based method to two additional mixed microbial community samples of environmental and biomedical research interest, i.e. river water filtrate and human faeces. Some minor sample-specific pretreatments, e.g. RNAlater treatment for the preservation of RNA of the faecal samples, were necessary to guarantee sufficient recovery of biomolecules. Analyses of the respective biomacromolecular fractions' quality and quantity resulted in no apparent differences, compared to those obtained from the LAO-enriched microbial community (Kruskal-Wallis, $0.054 < P < 0.729$, $n = 5$, **Figure 3.8**, **Table 3.3 and Supplementary Table VI**).

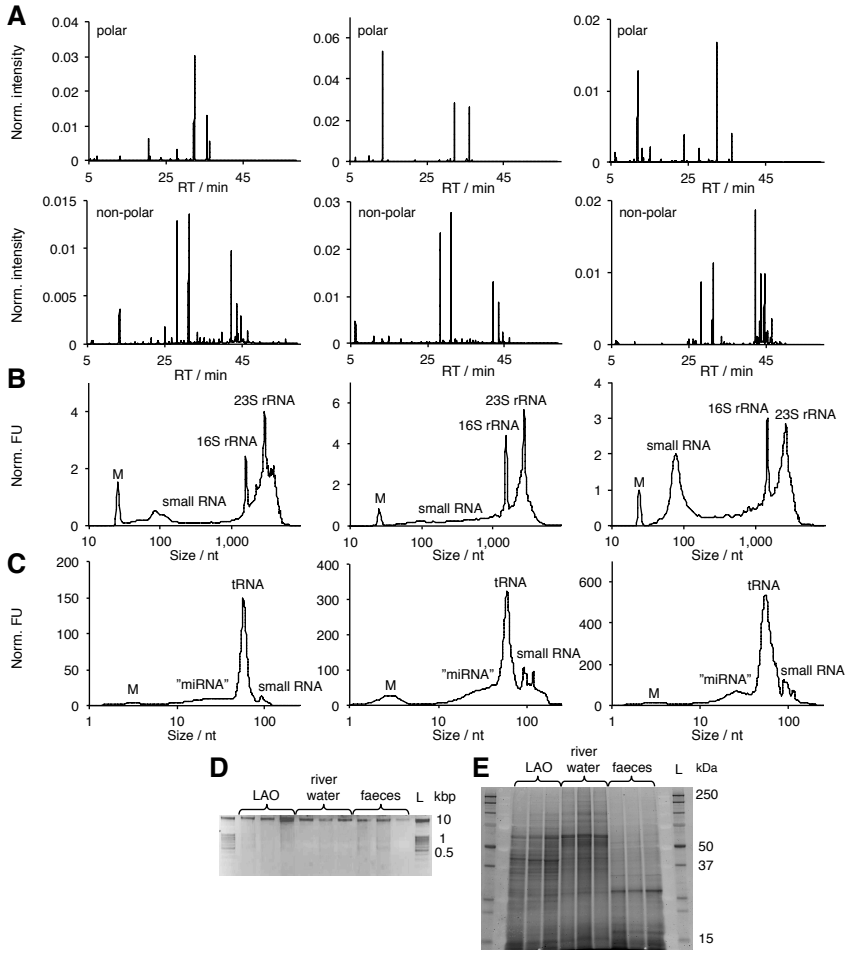

Figure 3.8 Application of the developed biomolecular isolation methodology to an LAO-enriched microbial community (n = 3), river water filtrate (n = 3) and human faeces (n = 3). (**A**,**B**,**C**) Left-hand panes represent LAO-enriched microbial communities, middle panes represent river water filtrate and right-hand panes represent human faeces. (**A**) Representative GC-MS total ion chromatograms of polar and non-polar metabolite fractions. Representative Agilent Bioanalyzer 2100 electropherograms of the (**B**) total RNA fractions and (**C**) small RNA fractions. (**D**) Agarose gel electrophoresis image of the genomic DNA fractions [Mean amount loaded in µg ± st.dev., left to right; LAO: 0.63 ± 0.28; River water: 0.35 ± 0.03; Faeces: 0.61 ± 0.26] for each of the three technical replicates considered. (**E**) SDS-PAGE image of protein fractions, first elution [Mean amount loaded in µg ± st.dev., right to left; LAO: 1.19 ± 0.40; River water: 1.70 ± 0.63; Faeces: 1.19 ± 1.21] for each of the three technical replicates considered. Abbreviations L, ladder; RT, retention time.

Table 3.3 Summary of metabolomics data, nucleic acid qualities and quantities, and protein quantities obtained using the developed NA-based method applied to an LAO-enriched microbial community, river water filtrate and human faeces.

Microbial community	Number of detected metabolites (% identified)		Quantity / μg			RNA quality parameters [1]	
	Polar	Non-polar	Protein	RNA	DNA	RIN	"miRNA"/ small RNA (%)
LAO-enriched microbial communities	246 (12.20)	158 (12.18)	60.59 ± 2.91	22.68 ± 7.06	28.79 ± 9.65	5.66 ± 0.15	27.28 ± 10.91
River water filtrate	75 (18.67)	70 (17.72)	84.33 ± 26.81	16.03 ± 3.36	11.28 ± 1.09	6.70 ± 1.49	31.00 ± 1.73
Human faeces	162 (9.88)	133 (11.28)	99.74 ± 23.63	18.48 ± 3.62	24.57 ± 8.31	6.03 ± 1.79	30.79 ± 1.52

[1] RIN (RNA integrity number) for the large RNA fraction and the ratio of "miRNA"/small RNA for the RNA fraction < 200 nt were determined using the Agilent 2100 Bioanalyzer software.

In order to ascertain the quality of the obtained DNA for further downstream analyses, each DNA fraction from the three environmental samples, was subjected to PCR amplification with primers targeting the 16S rRNA gene. For each microbial community DNA extract, clear and distinct bands with the correct molecular size were obtained (**Figure 3.9**), highlighting the possibility for specialised downstream analyses without the need for additional purification steps.

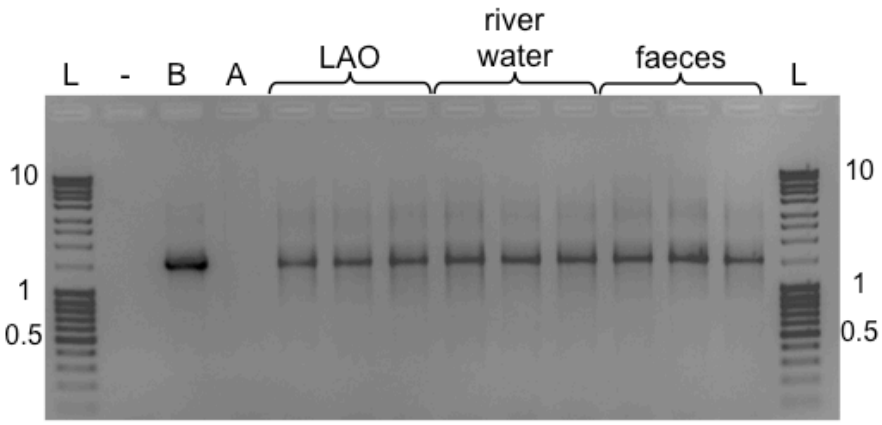

Figure 3.9 Agarose gel electrophoresis image highlighting the amplification of 1,300 nt of the 16S rRNA gene from DNA fractions obtained using the NA-based method on the three different microbiomes, i.e. LAO-enriched microbial community, river water filtrate and human faeces. Abbreviations: L, ladder; -, negative control; B, bacterial (*E. coli*) genomic DNA; A, archaeal (*Halobacterium* spp. NRC-1) genomic DNA.

Use of the developed standardised methodology further allows comparative analysis of different microbiomes. For example, it has not escaped the attention of the author that an apparent enrichment in smallRNA exists in human faecal samples (**Figure 3.8 B**). Microbial smallRNA has since been discovered in the blood of human subjects (Wang *et al*., 2012) opening up the possibility that smallRNA may be involved in human-microbiome interactions. Furthermore, in order to highlight potential signature

metabolites for the three analysed microbial communities, the relative abundances of detected metabolites were contrasted for each microbiome. For this, the metabolomic datasets from the respective microbiomes were subjected to a three-way comparative analysis. Clear differences in the metabolite composition of the three microbiomes are apparent, allowing the identification of specific signature metabolites for the three microbiomes (**Figure 3.10**, **Supplementary Table VI**). For example, from the confidentally identified metabolites, the enrichment of docosanol, a saturated fatty alcohol, in the LAO-enriched microbial community from the wastewater treatment plant may be explained by the wide use of this molecule as an emollient, emulsifier and thickener in cosmetics, nutritional and pharmaceutical products. Consequently, its presence may be expected in domestic wastewater and through its chemical properties would be enriched in LAOs. The pronounced enrichment of mannose, a simple sugar, in river water filtrate may be a direct result of photosynthesis by the dominant phototrophic organisms. Finally, a comparatively strong enrichment in citric acid was found in the human faecal samples. An increase of all metabolites involved in energy metabolism including citric acid was previously found in the serum of conventionally reared mice versus germ-free mice (Velagapudi *et al.*, 2010). The present results indicate that this elevated level may directly result from the metabolic activities of microbial communities in the gastrointestinal tract.

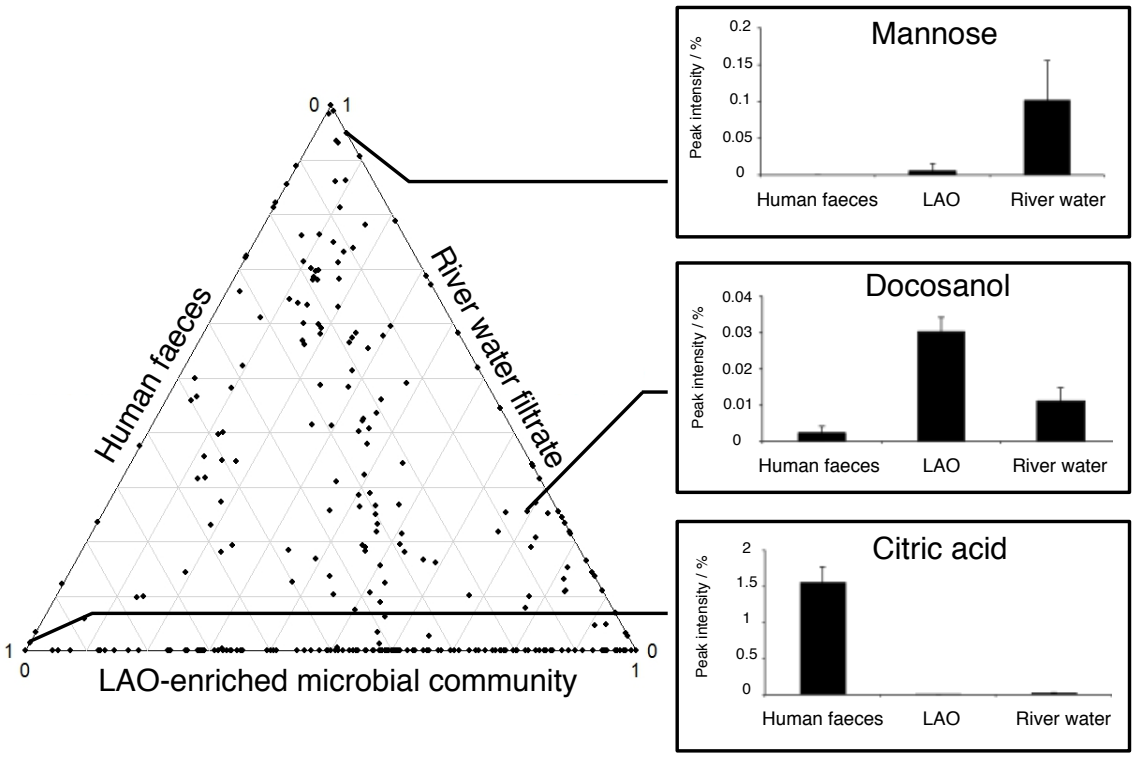

Figure 3.10 Three-way comparison of microbial community metabolomes obtained from an LAO-enriched microbial community (n = 3), river water filtrate (n = 3) and human faeces (n = 3). Each dot represents the relative abundance of a metabolite in the three different environmental samples. Exemplary metabolite signatures for each microbiome are highlighted with their respective abundance levels. Error bars represent s.d.

By applying the NA-based workflow to these two additional mixed microbial communities, it is demonstrated that the methodological framework can be applied to a range of different samples and, thus, represents a standardised biomolecular isolation procedure for comparative eco-systems biology investigations on a range of different microbial communities in the future.

3.4. Conclusion and perspectives

In microbial ecology, an increasing number of studies aim to integrate multi-omic datasets in order to obtain comprehensive high-resolution overview of microbial community structure and function. Molecular eco-systems biology investigations, centred around the integration of omic data, are faced with major challenges arising from the complexity, dynamics and heterogeneity of microbial consortia. Considering recent technological improvements and the accompanying decrease in cost of high-throughput molecular methods as well as associated progress in computational and statistical methodologies, these types of investigations will dramatically intensify in the coming years (Muller *et al.,* 2013).

In the present chapter, it is demonstrated that microbial community metabolomes do not allow certain biological and technical replicates to be discriminated, which is a direct consequence of the extensive heterogeneity encountered in these systems. Additionally, sample-to-sample variation is extensive for microbial communities at each biomolecular level. Consequently, major artefacts may be introduced into coupled high-resolution omic experiments by sample splitting before dedicated isolation of the, respective, individual biomolecular fractions (Muller *et al.,* 2013). Such approaches will result in disparate omic results and, thus, will not fulfill the premise of systematic measurements. Furthermore, they will not allow meaningful *in silico* reconstructions and modelling of community-wide processes and interactions. Therefore, the described methodological framework was developed, which allows for the sequential extraction and purification of all known biomolecular fractions from single unique samples. The methodology is based on the combination of chromatographic spin column-based biomacromolecular isolation with prior

extraction of concomitant polar and non-polar metabolite fractions. In the present study, it is demonstrated that the devised methodology yields equivalent or better results, compared to those obtained with dedicated extraction protocols, and that it is applicable to a range of different microbial community samples. Furthermore, the methodology may prove successful on a range of other biological samples, in particular those that are precious or which are characterised by extensive within-sample heterogeneity, e.g. human tumor samples. The utilisation of spin column chromatography for biomacromolecular separation minimises the risk of exposure to harmful chemicals and, thus, allows standardised routine laboratory use.

As already discussed in Section 2.4, the standardised and systematic application of community genomics, transcriptomics, proteomics and metabolomics to autumn and winter LAO communities (Chapter 2, Conclusion and perspectives) should reveal the link between genetic potential and functionality in lipid accumulation. Equipped with the biomolecular isolation framework described in this chapter, a comparative community-wide integrated omics analysis was carried out on the representative autumn and winter LAO communities (Chapter 4).

CHAPTER 4: COMPARATIVE INTEGRATED OMICS HIGHLIGHT KEYSTONE GENES IN MICROBIAL COMMUNITY-WIDE METABOLIC NETWORKS

4.1. Introduction

In recent years, our ability to study microbial communities in their natural environments has improved dramatically, due to rapid advances in high-throughput DNA sequencing technologies and developments in various omics analysis (Section 1.2). While metagenomics data provide gene inventories, without any proof of their functionality (Röling *et al.*, 2010), the analysis of community transcripts enables a first assessment of community-wide functions (Helbling *et al.*, 2012), and community proteomics provides a high-resolution representation of the actual phenotypic traits of individual community members (Wilmes *et al.*, 2009a). Nevertheless, the generation of metagenomic information is indispensable since it forms the backbone for the proper interpretation of transcriptomic and proteomic data ideally leading to population-level resolution of individual functions following data integration and analysis.

Metabolomics through resolving the final and intermediate products of cellular metabolism, theoretically, should be the most sensitive indicator of community-wide phenotypes. However, metabolomic methods are still somewhat limited in the number of metabolites, which can be measured and only a fraction of these can typically be positively identified (Roume *et al.*, 2013). To overcome current limitations that hamper comprehensive probing of community-wide metabolism by metabolomics, reconstruction of metabolic networks based on genomic data presents an alternative approach to assess overall differences in metabolism between different communities. With the progress of genome-sequencing capabilities, the reconstruction of metabolic pathways has led to the formulation of genome-scale metabolic models (Oberhardt *et al.*, 2009). The inception of this cell-level approach, which formally captures the

metabolic capabilities of the entire metabolic network of an organism, has so far allowed the reconstruction of more than 59 genome-scale metabolic models from 39 distinct organisms (Chindelevitch *et al.*, 2012). However, at present, we still know very little about the metabolic contributions of individual microbial players within microbial communities because of the prevalence and complexity of interactions among them (Zomorrodi & Maranas, 2012). The traditional approaches used to transit from single to multi-species metabolic network reconstructions is to consider the metabolic networks of individual species as an input-output system to build network-based (Cottret *et al.*, 2010; Freilich *et al.*, 2010) or constraint-based (Stolyar *et al.*, 2007; Wintermute & Silver, 2010) models of species interaction. These approaches demonstrate tremendous potential for elucidating relationship between microbial species and their environment or predict species interactions and subsequently inferring inter-species metabolic transfer. However, these multi-species models, rather limited to a few species, fail to explain how variations in gene or species composition affect the overall metabolic activity of ecosystems (Greenblum *et al.*, 2013). Given the complexity of microbial communities as well as the inability to isolate and sequence representative single cultures of all organisms within a community, such bottom-up approaches may be limited by the inherent impossibility to extrapolate community-wide networks and their behaviour from individual isolate omic datasets. In other words, community-wide networks are expected to be more than, merely, sums of their respective parts (Muller *et al*., 2013). Ideally, top-down and bottom-up approaches should be combined to identify links between microbial community structure and function, thereby, bridging the gap between individual metabolic networks and the larger community-wide networks to ultimately build a systems-level interaction model between species (Borenstein, 2012).

Given the challenges associated with reconstructing the individual metabolic networks of individual community members and then interfacing these, an alternative approach is to reconstruct community-wide metabolic networks based directly on metagenomic data thereby ignoring the contribution of individual species. This approach was recently pioneered by Greenblum and colleagues (Greenblum *et al.*, 2012) and through this population-free metabolic network reconstruction they identified enzyme-coding genes, either enriched or depleted, in individuals with obesity or inflammatory bowel disease. The topological analysis of these disease-associated enzymes revealed an unexpected metabolite exchange within the human gut environment in the context of disease.

Here, is introduced a unique framework for comparative eco-systems biology, which integrates meta-omic data within a community-level metabolic network reconstruction based on combined metagenomic, metatranscriptomic and metaproteomic data. The resulting networks allow description of the function of entire microbial communities by resolving gene expressions, protein frequencies, as well as the associated systems-level network topological features, to understand the dynamics and evolution of entire microbial ecosystems. In particular, this methodology was applied to further characterise and understand lipid accumulation in LAO-enriched microbial communities and allowed for a complete systems-level overview of the processes involved in the transformation of lipids within the LAO biomass.

4.2. Materials and methods

4.2.1. Sampling and sample processing

For the development of our comparative eco-systems biology analytical framework, one sample among four (defined, herein, as a biological replicate; Chapter 2 Materials and methods) was selected for 4 October 2010 (defined, herein, as the autumn LAO community, **Figure 4.1, A; Table 4.1**) and 25 January 2011 (defined, herein, as the winter LAO community, **Figure 4.1, B; Table 4.1**). These samples were selected, because they were found to be very dissimilar and at both extremes of the LAO-enriched microbial community-wide phenotypes witnessed at the Schifflange WWTP (Section 2.3.2.1). Thus, these particular biological replicates and sampling dates were selected with the intention to maximise the sample-to-sample variation in phylogenetic composition, particularly, in the abundance of *Microthrix parvicella* (Section, 2.3.2.1) and the variation within corresponding metabolite levels (Section 3.3.1) within the LAO communities.

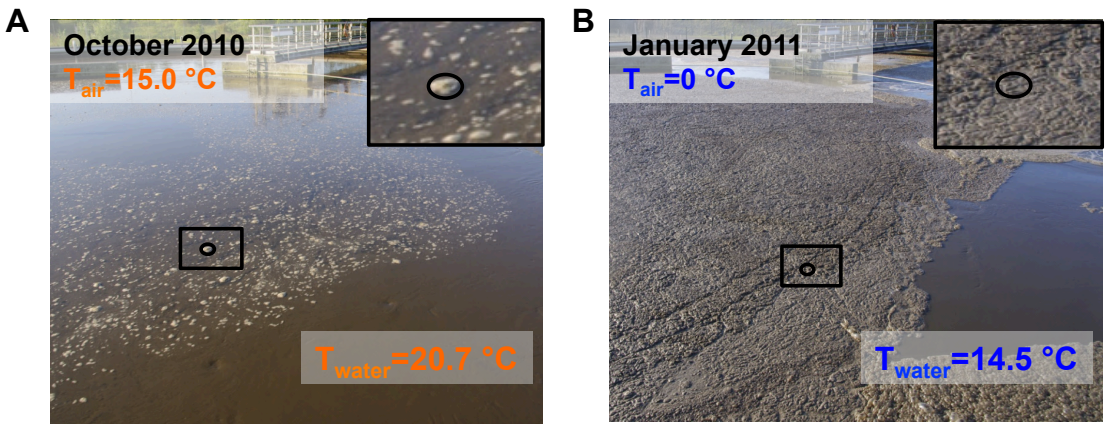

Figure 4.1 Photographs of the LAO-enriched microbial communities located at the air-water interface of the anoxic tank at the Schifflange wastewater treatment plant (**A**) in autumn and (**B**) in winter, subject of the present study. The sampling date, the air temperature (T_{air}) and the water temperature (T_{water}) are indicated on each photograph.

Table 4.1 Physico-chemical characteristics of the wastewater at the time of the sampling the anoxic tank of the Schifflange WWTP (**Figure 2.3**).

Sampling dates	Suspended solids (g/l)	pH	NO_3^- (mg/l)	O_2 (mg/l)	NH_4 (mg/l)	PO_4 (mg/l)	Water temperature (°C)
4 October 2010	2.78	6.94	2.77	0.64	0.6	2.06	20.7
25 January 2011	3.24	7.01	3.37	1.13	1.72	1.02	14.5

4.2.2. Biomolecular extractions

Biomolecular extraction The previously described biomolecular isolation framework for eco-systems biology (Sections 3.2.4 & 3.2.5) were applied to sequentially extract polar and non-polar metabolites, total RNA, genomic DNA, and proteins from LAO-enriched mixed microbial communities using the Qiagen AllPrep DNA/RNA/Protein Mini kit-based method (QA, Qiagen, Venlo, The Netherlands). For each sampling date, three samples (defined, herein, as technical replicates; Section 2.2.2) of 200 mg were used for biomolecular extraction. Each of the individual molecular fractions isolated from three technical replicates were combined into a pool sample in order to yield sufficient biomolecular quantities for subsequent high-throughput analysis and to reduce the influence of fine-scale sample-to-sample heterogeneity.

Biomolecular quality assessment and pre-processing for high-throughput analyses From the metabolite pools (Section 3.2.7), aliquots of 100 µl of extra- and intracellular polar and non-polar fraction were transferred into GC vials, immediately dried by evaporation and stored at -80 °C until analysis by GC-MS (Section 3.2.7). The quality and quantity of isolated biomacromolecules were assessed using dedicated measurements as described previously (Section 3.2.8; **Figure 4.2, Table 4.2**).

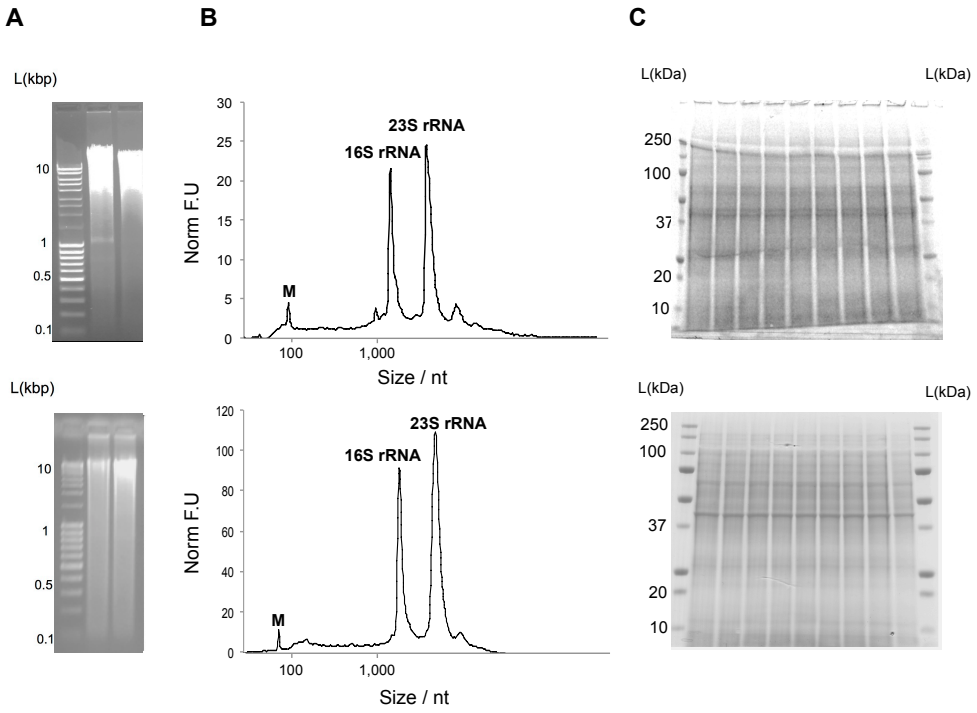

Figure 4.2 Quality assessment of biomacromolecular fractions isolated from the LAO-enriched microbial community samples. Top panels: autumn sample; lower panels: winter sample. (**A**) Agarose gel image highlighting representative high-molecular weight genomic DNA fractions. (**B**) Representative Agilent Bioanalyzer 2100 electropherograms of the total RNA. (**C**) SDS-PAGE image of the obtained protein fractions. Abbreviations, bp: base pair, M: marker, L: ladder.

Table 4.2 Quantitative and qualitative analyses of biomacromolecular fractions sequentially isolated from the LAO-enriched microbial community samples.

Sampling dates	DNA		RNA		Protein
	A260/A280	Quantity / µg	RIN	Quantity / µg	Quantity per gel line / µg
4 October 2010	2.20	15.44	8.9	81.97	20.1
25 January 2011	2.03	6.11	9.2	75	21.8

DNA The DNA pool sample was snap-frozen in liquid nitrogen in the elution buffer and stored at -80 °C until library preparation and sequencing (Section 4.2.4).

RNA The volume of the RNA pool sample was adjusted to 180 µl using RNase free-water. Then, a mixture of 18 µl of a 3M (w/v) sodium acetate solution, 2 µl of a glycogen solution (10 mg/µl) and 600 µl of ice-cold 96 % (v/v) ethanol were added to the RNA solution. The RNA solution was gently mixed by inversion and precipitated at -20 °C for, at least, 1 h. Following centrifugation at 10,000 g for 30 min at 4 °C, the RNA pellet was washed three times with ice-cold 70 % (v/v) ethanol at 4 °C. The pellet was then air dried for 5 min at room temperature overlaid with 100 µl of RNAlater (Ambion) and placed at -20 °C until shipment. For reclamation of the RNA, the solution was centrifuged at 14,000 g for 10 min at 8 °C. After removal of the supernatant, the pellet was washed three times with 100 µl of 80 % (v/v) ethanol, followed by high speed centrifugation for 3 min at 8 °C. The pellet was then air dried until all visible ethanol had evaporated. The dried RNA pellet was then resuspended in the appropriate volume of 1 mM sodium citrate buffer (pH 6.4) to reach a concentration of 55 ng/µl and 145 ng/µl for winter and autumn RNA sample, respectively.

Protein Following 1D-SDS-PAGE electrophoresis and staining with Imperial protein stain (Thermo Scientific; Section 2.2.8), the protein gel was conserved at 4 °C in a plastic bag under vacuum. Prior to further analyses, entire lanes were cut into 1mm slices using a grid cutter (MEE-1x5, Gel Company), yielding approximately 70 slices per lane. Two 1 mm slices were combined in a single well of a 96 well V-bottom plate with a hole introduced into the bottom using a 30 gauge lancet needle (Becton Dickson). The wells contained size 11 black hexagonal glass beads (SB3656, Fusion

Beads) to prevent the gel pieces from clogging the hole. For in-gel digestion, an automated liquid handling system (Tecan EVO) was used for reduction, alkylation, tryptic digestion and peptide extraction from the gel pieces. After extraction, the peptide solution was dried and reconstituted in 20 µL of a solution of 0.1 % (v/v) [acid trifluoroacetic (TFA, 5 %; Sigma)/ acetonitrile (ACN, 95 %; BioSolve)] in Milli-Q H_2O in a round bottom polypropylene 96 well plate (Greiner) and placed into an Eksigent Nano 2D plus system autosampler (ABSciex) for analysis.

4.2.3. Comparative metabolomic and lipidomic analyses

Identical analytical procedures were used for the identification of metabolites from the extra- and intracellular compartments of both temporally resolved metabolite extracts.

Metabolomics For the detection of total polar and non-polar metabolites, GC-MS measurements were performed in triplicates and matched to a dedicated in-house metabolomics library using the same methodology described for the metabolomic analyses of the representative LAO-enriched microbial community samples (Section 3.2.7). Following normalisation of the peak intensities of detected metabolite features, the data were analysed using statistical methods i.e., PCA and hierarchical clustering. The ade4 R package (R version 2.13.2, http://www.r-project.org/) was used for the principal component analysis (dudi.pca function).

Triacylglycerol analyses The identification and quantification of triacylglycerols (TAGs) was carried out on the non-polar metabolite fractions using liquid

chromatography coupled to tandem mass spectroscopy (LC-MS/MS). Dried metabolite extracts were dissolved in 200 µl of a chloroform/methanol mixture (2:1 v/v). An Agilent 1290 Infinity LC with a 2.1 (i.d.) x 100 mm, 3.5 µm Agilent ZORBAX Eclipse XDB-C18 Narrow Bore Rapid Resolution column heated to 30 °C was used with a binary solvent system and a flow rate of 500 µl/min. The system was equilibrated with solvent A [5 mM ammonium acetate dissolved in methanol/water (99:1 v/v)]. A 5 µl injection volume of the non-polar metabolite extract was applied to the column, followed by a 30 min linear gradient to 80 % solvent B [5 mM ammonium acetate dissolved in isopropanol/water (99/1 v/v)], and held there for 2.5 min. An Agilent 6530 qTOF equipped with an Agilent ESI source was used for accurate mass analysis of the LC eluent. Positive-ion data were generated by operation of the mass spectrometer with a capillary voltage of 3,500 V, nebulizer of 35 pounds per square inch, drying gas of 8 l/min, gas temperature of 300 °C, fragmentor of 120 V, skimmer of 65 V, and octopole radio frequency voltage of 750 V. Mass spectra data were acquired in 2 GHz Extended Dynamic Range mode at a rate of 1 spectra per second, and data were collected as profiled spectra over a mass range of 150 to 3,200 m/z. Mass calibration was performed with an external reference mass ion of m/z 922.009798 to enable accurate mass determination. Data were collected with the Agilent MassHunter WorkStation Data Acquisition software, version B.04.00. LC-MS data files were processed with the MassHunter Qualitative Analysis Software version B.03.01. Molecular features (MFs) were extracted from the raw MS1 data using the Molecular Feature Extraction (MFE) algorithm with a peak filter of ≥ 10,000 counts. The resulting MFs were then identified with MassHunter by searching against a custom database containing 1,382 triacylglycerol (TAG) ions corresponding to $[M+Na]^+$ and $[M+NH_4]^+$ (M+: positive ion mode) adducts using a

±10 ppm search window. For quantitative purposes, TAG lipid groups were compared using the $[M+NH_4]^+$ molecular ion height.

Long chain fatty acid analysis. The identification and quantification of long chain fatty acids (LCFAs) were carried out using GC coupled to tandem mass spectroscopy (GC-MS/MS). Dried non-polar metabolite extracts were dissolved in different volume of dichloromethane (**Supplementary Table IX**) and derivatised using the following procedure: 40 µl of re-dissolved non-polar metabolite extract were transferred into a GC vial and mixed with 40 µl of BSTFA (N,O-bis-trimethylsilyl-trifluoroacetamide) and TMCS (N,O-bis-trimethyl-trifluoroacetamide) 99/1 (v/v). The mixture was then heated during one hour at 70 °C prior to analysis by GC-MS/MS. The analysis was carried out on a Thermo Trace GC and a Thermo TSQ Quantum XLS triple-quadrupole MS. Samples were injected in PTV splitless mode and separated on a Restek Rxi-5Sil MS (0.18mm x 0.18µm x 20m) column. Helium was used as a carrier gas at a constant flow rate of 1.0 ml/min. Detection was performed in Multiple Reaction Monitoring (MRM) mode, with 2 MRM transitions per target compound. The absolute quantification of LCFAs was obtained by external calibration against standard mixtures of the pure compounds.

4.2.4. Next generation sequencing

Identical analytical procedures were used to sequence total genomic DNA and the rRNA substracted RNA using Illumina HiSeq technology for both temporally resolved samples considered. Isolated high-purity genomic DNA (Section 4.2.2) was subjected to the sequencing protocol whereas isolated RNA (Section 4.2.2) was first

depleted of rRNA and then subjected to complementary DNA synthesis as details below.

RNA-Seq In order to enrich the RNA fraction in mRNA, the rRNA was subtracted from the total RNA fraction using the Ribo-Zero rRNA Removal Kit (Meta-Bacteria; Epicentre). This was followed by cDNA synthesis and amplification using the ScriptSeq™ v2 RNA-Seq library preparation kit (Epicentre).

rRNA removal The procedure consisted of an initial washing and re-suspension of the Ribo-Zero microspheres with dedicated solutions. Following treatment of the total RNA sample with Ribo-Zero removal solution, the RNA sample solution was added to the re-suspended Ribo-Zero microspheres and selected by hybridisation. The RNA-microsphere solution was then removed and the rRNA-depleted sample was purified by ethanol precipitation.

Library preparation The cDNA synthesis and amplification were performed using the ScriptSeq™ v2 RNA-Seq library preparation kit (Epicentre). The procedure consisted of initial fragmentation of the RNA, followed by the annealing of the cDNA synthesis primer. Briefly, 4 µl of cDNA Synthesis Master Mix was added to fragmented RNA solution and incubated at 25 °C for 5 min followed by 42 °C for 20 min. cDNA Synthesis Master Mix was prepared combining 3 µl cDNA Synthesis PreMix (ScriptSeq v2 RNA-Seq library preparation kit, Epicentre), 0.5 µl of DTT (100 mM) and 0.5 µl of StartScript Reverse Transcriptase solution (Epicentre). After cooling at 37 °C, 1 µl of finishing solution (Epicentre) was added to the cDNA synthesis solution and incubated at 37 °C for 10 min, before an incubation at 95 °C for 3 min. 8 µl of terminal tagging master mix was then added to each solution and incubated at 25 °C for a further 15 min, following 95 °C incubation for 3 min. Terminal tagging

master mix was prepared using 7.5 µl of Terminal tagging premix (Epicentre) and 0.5 µl of DNA polymerase. The cDNA 3'-terminal tagged was then purified using the AMPure XP system (Beckman Coulter). The purified cDNA strand was then amplified by PCR, thereby, generating the second strand of cDNA, adding the Illumina adaptator sequences and incorporating specific barcodes. Finally, the RNA-Seq library was purified using AMPure XP system (Beckman Coulter). The size distribution of the RNA-Seq library was assessed using the 2100 Bioanalyzer (Agilent).

DNA/cDNA sequencing DNA/cDNA was prepared according to the modified instructions from The Wellcome Trust Sanger Institute (Kozarewa & Turner, 2011). The metagenomic sequencing protocol used a 96-well library preparation and the molecular barcoding method for Illumina library construction. The barcodes were designed using Hamming codes, which allow single nucleotide sequencing errors to be corrected and single indels (insertions/deletions) to be detected without ambiguity. Several optimisations were employed in the indexing protocol. The tags used were 8 bp long, which allowed the design of larger number of barcodes with error-correcting capability. The bar codes were introduced in a regular PCR, which simplifies the PCR step and allows using as few as six cycles. Before pooling, the relative concentration of each sample library was measured by quantitative PCR (qPCR), which then allowed the pooling of libraries together, accurately, and improved the uniformity of their representation.

DNA fragmentation 1-5 µg of DNA was used for the fragmentation step, as determined by gel analysis, and re-suspended in 75 µl of 10 mM Tris-HCl, pH 8.5. The sample was then sheared for 150 sec in 100 µl Covaris microtubes. The following

programme was used: duty cycle 20 %, intensity 5, cycle burst 200, power 37 W, temperature 7 °C and mode Freq sweeping. The sheared DNA samples were then transferred into a MicroAmp optical 96-well plate. Six samples, chosen at random, were run on an Agilent DNA 1000 chip to check the quality of the fragmentation. A 150-250 bp smear was detected and 70-90 % of the initial DNA amount was recovered.

Size selection A large fragment plate was prepared by dispensing 150 µl of beads into wells of a round-bottom Costar plate for each sample. Additionally, a small-fragment plate was prepared by dispensing 60 µl of beads into wells of a separated Costar plate for every sample. A large-fragment Costar plate was placed on a magnetic stand, allowing beads to be collected and 45 µl of 10 mM Tris-HCl buffer were removed from large-fragment wells, leaving all beads in the well. The plate was removed from the magnetic stand and 190 µl of sample from sonication tubes was added to the large fragment wells on the Costar plate. Following 8 min of incubation at room temperature, using two magnetic stands, the large-fragment Costar plate was placed on a magnetic stand to collect beads. Following a 5 min incubation at room temperature, 58 µl of bead buffer were removed from the small fragment wells, leaving behind beads and ~2 µl of bead buffer. Small-fragment wells were then removed from the magnetic stand. 300 µl of supernatant were transferred from large-fragment wells into small-fragment wells. After transfer, the large-fragment plate was discarded. Samples were then incubated in small-fragment wells for 5 min and the plate was placed on a magnetic plate to collect beads. 300 µl of supernatant were removed and replaced by 300 µl of 80 % ethanol without disturbing beads. Following 30 sec of incubation, ethanol was removed. This last step was repeated two times. Following ethanol removal by air-drying for 5 min, the small fragment plate was

removed from the magnetic stand and 45 µl of pre-warmed 10 mM Tris-HCl (pH 8.5) were dispensed into the wells. After re-suspension of beads by vigorous mixing, the solution was incubated for 2 min and placed on a magnetic plate to collect beads. Following a futher 3 min incubation, 42.5 µl of supernatant containing the size-selected products were transferred to a new plate.

All enzymes and reaction buffers used for the end-repair, dA tailing, ligation and PCR amplification were provided from the KAPA Library Preparation Kits with Standard PCR Library Amplification/Illumina series (KapaBiosystems).

End-repair 100 µl of end-repair master mix, prepared as below, were added to each well. To the supernatant containing the sheared DNA, 10 µl of end-repair buffer were added to 5 µl of end-repair enzyme solution. The plate was then covered, vortexed briefly and spun down before being incubated for 30 min at 20 °C in a thermal cycler. The samples were then cleaned using Agencourt AMPure SPRI beads (Beckman Coulter). 90 µl of SPRI beads were added into each well, the plate was covered and vortexed for 30 sec. The reaction plate was placed on the magnetic SPIRPlate for 10 min and beads separated from solution. Following a 10 min incubation, the solution was completely clear and discarded. 200 µl of 70 % (v/v) ethanol were added to each well and incubated for 30 sec at room temperature. The ethanol washes were repeated two times. The reaction plate was dried for 15 min in a thermocycler and each sample was eluted with 43.5 µl of 10 mM Tris buffer (pH 8.0).

A-tailing 50 µl of A-tailing master mix, prepared as below, were added to each well. 30 µl of end-repaired DNA were added to 5 µl of 10x A-tailing buffer and 3 µl of A-tailing enzyme. The plate was then covered, vortexed briefly and spun down before being incubated for 30 min at 30 °C in a thermocycler. The sample was then cleaned

up using the Agencourt AMPure SPRI bead method with 90 µl of SPRI beads placed in each well. Each sample was then eluted with 36 µl of 10 mM Tris buffer (pH 8.0). The same six samples, randomly chosen for the DNA fragmentation step, were run on an Agilent DNA 1000 chip and the average sample concentration was calculated using the "integrated peak" function.

Adapter preparation and ligation 50 µl of adaptor ligation reaction mixture, prepared as below, were added to each well. 30 µl of A-tailed DNA were then added to 10 µl of 5x ligation buffer, 5 µl of DNA ligase and 5 µl of DNA adaptor (30 µM). The plate was then incubated at 20 °C (room temperature) for 15 min. The sample was cleaned up using Agencourt AMPure SPRI beads by addition of 40 µl of SPRI beads to each well. Each sample was then eluted with 30 µl of 10 mM Tris buffer (pH 8.0) / 0.05 % (v/v) Tween 20. The same six samples were randomly chosen for the DNA fragmentation step and run on an Agilent DNA 1000 chip to check the success of the ligation; the smear obtained had an average molecular size of 50 to 200 bp larger than before ligation.

PCR amplification Library amplification was carried out according to a modified reaction setup as defined in the KAPA Library Preparation kit instruction (Illumina). 25 µl of 2x Kapa HIFI Hotstart Mix were spiked with 1M betaine. The addition of betaine aids in amplification of high GC % regions and reduces biases. 1 µl of the supplied PCR primers were added to each tube and a sufficient quantity of water was added to reach 50 µl of solution per PCR. The tubes were then transferred into a PCR thermocycler and run with the recommended KAPA library preparation cycling programme. The samples were then cleaned using the SPRI beads method with 40 µl of SPRI beads. Each sample was eluted using 50 µl of 10 mM Tris-HCl / 0.05 % (v/v)

Tween 20. The same six samples, randomly chosen DNA fragmentation step, were run into an Agilent DNA 1000 chip to check the success of the indexing enrichment PCR.

Final library quantification by qPCR The KAPA Library Quant kit (KapaBiosystems) for final library quantification was used (Illumina). Briefly, after an initial 1:1000 dilution in 10 mM Tris-HCl, pH 8.0 + 0.05 % (v/v) Tween 20, the 2x KAPA SYBR FAST qPCR master mix was used to amplify the DNA library with six other standards on an ABI 7900 thermocycler. The qPCR step was conducted using the following cycling conditions: 95 °C for 5 min followed by 35 cycles at 95 °C for 30 sec and at 60 °C for 45 sec. The concentration of the library was established using the standards. According to the manufacter's instruction recommendations. Both the standards and the libraries with unknown concentration were run in triplicate. Prior to sequencing, the libraries were diluted to the required concentration (e.g. 4.5 pM) following the Illumina cluster generation protocol.

A total of 16.2 gigabases (Gb) of shotgun metagenomic sequence data as well as 38.6 Gb of metatranscriptomic sequence data from the two LAO communities, autumn and winter samples, were generated.

4.2.5. Sequence assembly, gene annotation and determination of KO abundances

Sequence assembly For each of the two dates, the raw paired-end read sequences from the metagenome and metatranscriptome datasets (Section 4.2.4, **Figure 4.3)**, were processed separately and together using the PAired-eND Assembler for Illumina sequences (PANDAseq, **Figure 4.3, step 1**; Masella *et al.*, 2012). Briefly, PANDAseq determines the location of the amplification primers, identifies the optimal overlap between reads, corrects sequencing errors and checks for various constraints such as length and base quality. PANDAseq was run with a score threshold of 0.9 and 25 nt minimum overlap requirement (-o 25 -t 0.9 –F). The reads selected by the PANDAseq assembler were extracted from the raw sequence files using in-house Perl scripts. The remaining non-redundant paired-end reads were trimmed using the trim-fastq.pl script from the PoPoolation package (Kofler *et al.*, 2011) using a quality-threshold of 20 (1 % probability of miscall) and a min-length of 40 nt. Terminal Ns were removed, thereby, preserving trimmed reads of at least 75 nt. The PANDAseq and the single-end reads were then combined into a single fastq file. In total, this resulted in two paired-end (R1 and R2) and one single end file for each of the four libraries. These files were then used as input for the MOCAT assembly pipeline (**Figure 4.3, step 2;** Kultima *et al.*, 2012) using default parameters. The MOCAT pipeline can process metagenomic assemblies in a standardised and automated way. Briefly, the pipeline uses state-of-art programs to quality control map and assemble reads from sequence files at a depth of several billion base pairs and predicts protein-coding genes of the assembled metagenomes. The assembly part uses SOAP*denovo,* version 1.06. The assembled contigs were filtered with minimum length threshold of 150 nt (**Figure 4.3, step 3**). To enhance the final assembly by

reads that were assembled by PANDAseq, but not used by the MOCAT assembly pipeline, all PANDAseq reads were mapped onto the MOCAT contigs using SOAP*denovo* (Li *et al.*, 2008), with the following parameter settings: -r 2 -M 4 -l 30 -v 10 -p 8 (**Figure 4.3, step 3**). SOAP*denovo* has been specifically designed to align large amount of short reads generated by Illumina sequencing. The unmapped PANDAseq assembled reads, with a minimum length of 150 bp, were extracted and added to the contigs. The final contig files were made non-redundant using Cd-hit (-c 1.0; Huang *et al.*, 2010) by clustering identical sequences.

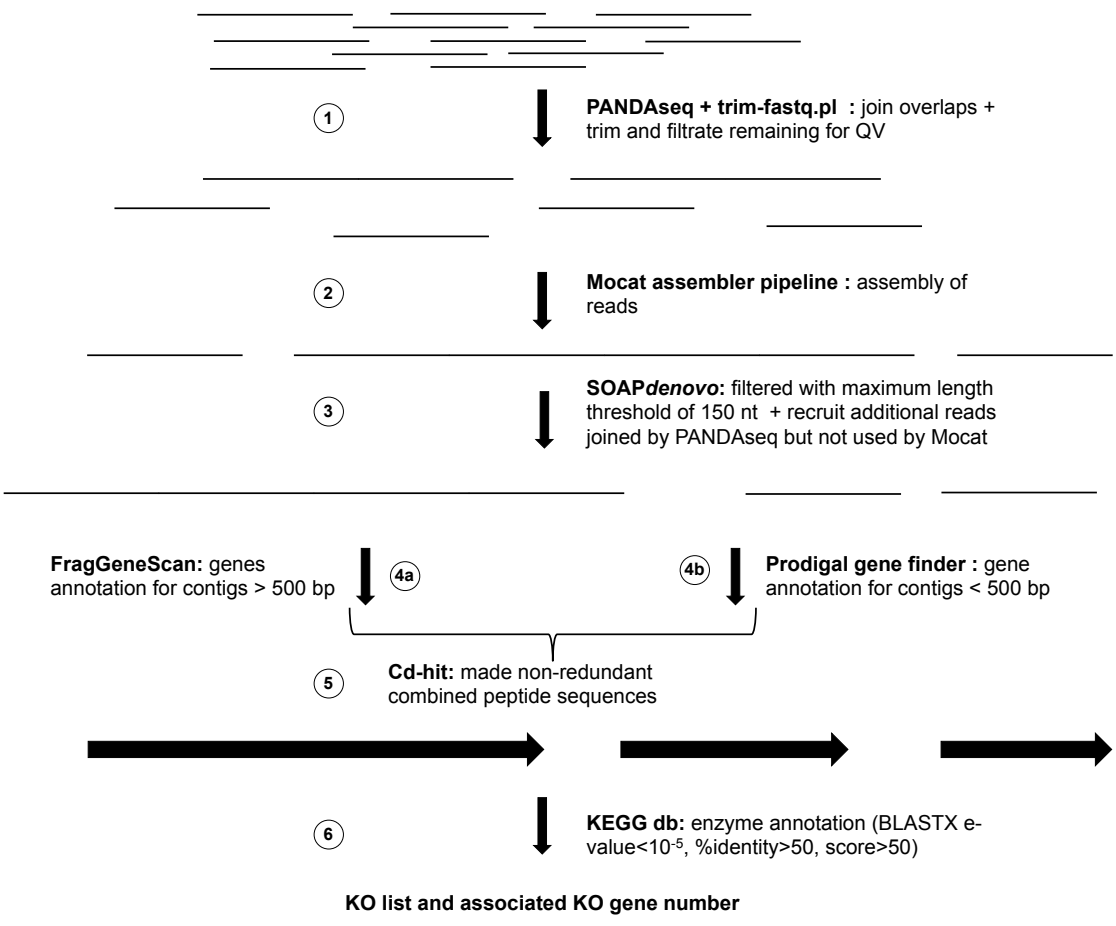

Figure 4.3 Assembly and annotation pipeline.

Gene annotation The non-redundant contigs files were split into two distinct files, one file with contigs of a length below 500 bp and another file with contigs lengths above or equal to 500 bp. The contigs with length below 500 bp were annotated using FragGeneScan (**Figure 4.3, step 4a;** Rho *et al.*, 2010). FragGeneScan has specifically been developed for direct gene identification of short reads obtained from metagenomes. FragGeneScan was run using settings for short sequence reads with sequencing error (-complete 0 –train illumine_5). The contigs with a length equal or above 500 bp were annotated with Prodigal gene finder (v2.60, **Figure 4.3, step 4b**; Hyatt *et al.*, 2010) using the MOCAT gene prediction processing steps. The two resulting amino acids sequences files were then combined into single peptide sequences files and made non-redundant using Cd-hit with sequence identity threshold of 1.0 and a length of description in the cluster file of 5000 amino acids (-c 1.0 –d 5000; **Figure 4.3, step 5**).

The assembly and annotation workflow were repeated by combining the pre-processed reads from the metagenomes and metatranscriptomes for both sampling dates (**Supplementary Table X**) as well as for each sampling date, separately (**Supplementary Table XII**). The Kyoto Encyclopedia of Genes and Genome (KEGG) database was used to annotate enzymes with metabolic reactions. All sequences were mapped onto the KEGG database version 64.0 (BLASTX; e-value<10^{-5}, %identity>50, score>50) using BLAT (Kent, 2002). Sequences were annotated with KOs (KEGG orthology groups; **Figure 4.3, step 6**).

Gene copy and transcript abundances To account for differences in read depth and sampling the number of raw sequence reads from metagenome and metatranscriptome

libraries were balanced by randomly selecting the same number of reads from both libraries using an in-house developed Perl script using the 'shuffle' method from 'List::Util' CPAN package[1]. The size of the smaller library was used to balance the respective datasets (i.e., 14,546,374 reads for the metagenomic and 16,443,761 reads for the metatranscriptomic datasets). The four balanced raw read sequence libraries were then mapped separately to the combined assembly for both sampling dates using SOAP*denovo* (Li *et al.*, 2008) with the following parameter settings: -r 2 -M 4 -l 30 -v 10 -p 8. For each library, the numbers of raw reads matching to an annotated gene were summed using an in-house Perl script. Generally, the expected number of reads mapping to a gene is proportional to both its gene (or transcript) abundance and length (Lee *et al.,* 2010). Therefore, to obtain the abundances of genes and transcripts, the reads from the genomic and transcriptomic libraries were mapped to the annotated genes and were normalized by the effective length of the gene sequence (Lee *et al.,* 2010) leading to the definition of Normalised Gene Abundance (NGA; **Figure 4.4, A**). The NGA_{gene} then defines the gene copy number, whereas $NGA_{transcript}$ the gene transcript abundance. The NGAs of all the genes belonging to the same orthologous gene (KO) were then added thereby providing the KO Abundances (KOA; **Eq. 1, Figure 4.4, B**) for gene copy numbers and transcript abundances respectively. The KOA is calculated as follows:

$$KOA = \sum_{i=1}^{n} (NGA)_i \qquad \textbf{(Eq. 1)}$$

Where KOA denotes the KO abundance, n represents the number of gene belonging to a KO (KO gene number).

[1] http://search.cpan.org/~pevans/Scalar-List-Utils-1.30/lib/List/Util.pm#shuffle_LIST%60

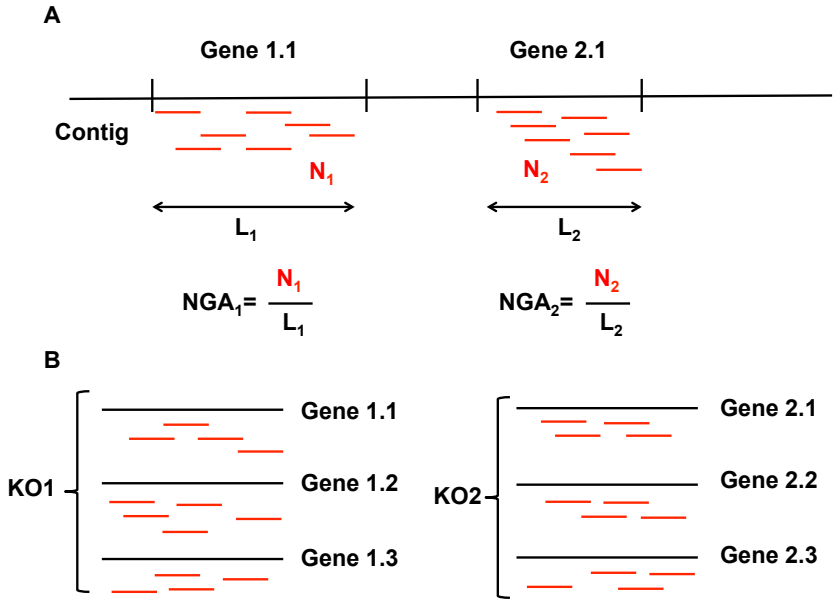

Figure 4.4 Gene abundance measurements. (**A**) Two genes 1.1 and 2.1 are located on within the same contig. Reads mapping to the gene are represented by red lines. Normalised Gene Abundance (NGA) is defined as the number of reads map to the gene (N) divided by the length of the annotated gene (L). (**B**) Representation of the association of genes to particular KOs. In this example, genes 1.1 to 1.3 belong to KO1 (n=3) whereas Genes 2.1 to 2.3 belong to KO2 (n=3).

Using the calculated KOA for both DNA and mRNA (cDNA) reads, relative gene expression (RGE; **Eq. 2**) was determined for each individual KO (**Supplementary Table X**):

$$RGE = Log_2(\frac{\mathrm{KOA_{mRNA}}}{\mathrm{KOA_{DNA}}}) \quad \textbf{(Eq. 2)}$$

Where RGE denotes the KO gene expression level, KOA_{mRNA} denotes the normalised KOA at the transcript level and KOA_{DNA} denotes the normalised KOA of gene copy abundance.

A particular KO is considered to be overexpressed for a $\log_2 \geq 1$ (i.e. more than 2-fold change according to gene copy abundance or gene expression) and underexpressed if this ratio is ≤ -1.

To identify particular KOs associated with a given season, the $\log_2$ ratio of the KO abundances at the genomic and transcriptomic levels were calculated to define season association factors (SAF; **Eqs. 3&4; Supplementary Table X)** according to the following formulae:

$$SAF_{DNA} = Log_2(\frac{KOA_{DNA_{winter}}}{KOA_{DNA_{autumn}}}) \quad \textbf{(Eq. 3)}$$

$$SAF_{mRNA} = Log_2(\frac{KOA_{mRNA_{winter}}}{KOA_{mRNA_{autumn}}}) \quad \textbf{(Eq. 4)}$$

Where SAF_{DNA} denotes the seasonal KO abundance determined at the gene copy number level (gene copies), SAF_{mRNA} denotes seasonal KO abundance determined at the mRNA transcript level. (**Eq.3**) KOA_{DNA} denotes the number of normalised KO gene copy abundances per season and (**Eq. 4**) KOA_{mRNA} the number of normalised KO transcript levels per season.

Finally to define overall gene overrepresentation between seasons (**Eq. 5**, **Supplementary Table X)**, an overall seasonal overrepresentation factor (SOR) was defined as follows:

$$SOR = Log_2(\frac{\frac{KOA_{mRNA_{winter}}}{KOA_{DNA_{winter}}}}{\frac{KOA_{mRNA_{autumn}}}{KOA_{DNA_{autumn}}}}) \quad \textbf{(Eq.5)}$$

Where SOR denotes the seasonal KOs representation. KOA_{DNA} and KOA_{mRNA} are defined as in **Eqs. 3&4**.

As for the relative gene expression values, a KO has a higher seasonal representation if SAF or SOR is ≥ 1 (winter associated) or $\leq$ -1 (autumn associated).

4.2.6. Proteomics

Liquid chromatography Peptides digested from 1 mm gel bands were separated using an Eksigent Nano 2D LC plus system employing splitless nanoflow. Reverse phase high performance liquid chromatography (RP-HPLC) and separation columns were prepared in-house by packing a Kasil fritted capillary [360 µm outer diameter (OD), 75 µm inner diameter (ID)] with a 1 cm bed of ReproSil Pur C18-AQ 3µm 120Å stationary phase (Dr. Maisch GmbH) for the sample trap and desalting column. A Kasil fritted capillary (360 µm O.D., 75 µm I.D.) was packed with a 15 cm bed of the same stationary phase as the separation column and this was connected to a PicoTip emmiter (360 µm OD x 20 µm ID, Tip 10 µm, FS360-20-10-N-20) for nano-electrospray ionisation. For each LC run, the sample was injected for 10 minutes at 2.5 µL/min with loading buffer (2 % v/v acetonitrile and 0.1 % v/v formic acid). The sample was separated by a linear gradient changing from 98 % solvent A (0.1 % v/v formic acid in water) and 2 % solvent B (0.1 % v/v formic acid in acetonitrile) to 40 % A and 60 % B in 60 minutes at 0.3 µl/min.

Mass spectrometry Following LC separation, the peptides were analysed on a Thermo-Fisher LTQ-Velos Orbitrap. MS1 data were collected over the range of 300 – 2,000 *m/z* in the Orbitrap at resolution 30,000. Fourier-transform mass spectrometry (FTMS) preview scan and predictive automatic gain control (pAGC) were enabled. The full scan FTMS target ion volume was $1x10^6$ with a maximum fill time of 500 ms. MS2 data were collected in the LTQ-Velos with a target ion volume of $1x10^4$ and a maximum fill time of 100 ms. The 10 most intense peaks were selected (within a window of 2.0 Da) for higher-energy collisional dissociation (HCD) at 15,000 resolution in the Orbitrap. Dynamic exclusion was enabled in order to exclude an

observed precursor for 180 sec after two observations. The dynamic exclusion list size was set at the maximum 500 and the exclusion width was set at ±5 ppm based on precursor mass. Monoisotopic precursor selection and charge state rejection were enabled to reject precursors with $z = +1$ or unassigned charge state.

Protein identifications For MS analysis, Thermo .RAW files were converted to mzXML format using MSConvert [ProteoWizard (Kessner *et al.*, 2008)] and searched with X!Tandem (Craig *et al.*, 2004) version 2011.12.01.1. Spectra were searched against combined unique assembled contigs and reads bellonging to 150 bp from metagenome and metatranscriptome data sets considering the two dates, common lab protein contaminants, and decoys. Redundancy was removed from these three data sets using BlastClust. The contaminant database was a modified version of the common Repository of Adventitious Proteins (cRAP, www.thegpm.org/crap) with the Sigma Universal Standard Proteins removed and human angiotensin II and [Glu-1] fibrinopeptide B (MS test peptides) added, for a total of 66 entries. Decoys were generated with Mimic (www.kaell.org), which randomly shuffles peptide sequences between tryptic residues, but also retains peptide sequence homology in decoy entries. Search criteria used for X!Tandem included a precursor mass tolerance of 15 ppm and a fragment mass tolerance of 15 ppm for higher-energy collisional dissociation (HCD) spectra. Peptides were assumed to be semi-tryptic (cleavage after K or R except when followed by P). Semi-tryptic peptides with up to 2 missed cleavages were allowed. The search parameters included a static modification of + 57.021464 Da at C for carbamidomethylation by iodoacetamide and a potential modifications of + 15.994915 Da at M for oxidation, -17.026549 Da at N-terminal Q for deamidation, and -18.010565 Da at N-terminal E for loss of water from formation of pyro-Glu. Additionally, -17.026549 Da at the N-terminal carbamidomethylated C for

deamidation from formation of S-carbamoylmethylcysteine and N-terminal acetylation were searched. MS/MS data were analysed using the Trans Proteomic Pipeline (Deutsch *et al.*, 2010) version 4.6 Rev.1. Peptide spectral matches (PSM) generated were analysed with PeptideProphet (Keller *et al.*, 2002) to assign each PSM a probability of being correct. Accurate mass binning was employed to promote PSMs whose theoretical mass closely matched the observed mass of the precursor ion, and to correct for any systematic mass errors. Decoys and the non-parametric model option were used to improve PSM scoring. The PeptideProphet scores were then analysed in iProphet (Shteynberg *et al.*, 2011), which assigns a probability for each unique peptide sequence, based on how often it is observed at different charge states with different modifications and by different search engines (although here, only X!Tandem was used), as well as whether other peptides from the same protein are also observed. Protein identifications were inferred with PeptideProphet. The false discovery rate for a given ProteinProphet probability was calculated using the number of decoy protein inferences at that probability. Only proteins identified at ProteinProphet probabilities corresponding to a false discovery rate (FDR) less than 1.0 % were further considered.

KO annotations and protein frequency The corresponding sequences for the above selected proteins were collected in FASTA format from the non-redundant single peptides/protein sequences database previously generated from the combined metagenome and metatranscriptome assemblies (Section 4.2.5). The resulting proteins sequences were then mapped to the KO library (e-value<10^{-5}, %identity>50, score>50) using BLAT (Kent, 2002). From the resulting list of KOs, the frequency of each KO was defined at the protein level, using an inhouse developed Perl script.

4.2.7. Community-wide metabolic network reconstructions

Community-wide metabolic networks were reconstructed from the entire sample set of KOs identified in the single and combined metagenomic/metatranscriptomic datasets per date. The chosen approach uses a connectivity-centred view of metabolism, in which KOs are represented by nodes connected by undirected edges. Edges represent a common product or substrate shared by two enzymes catalysing, for example, successive reactions, as defined in the KEGG database version 64.0. Each KO was assigned a pair-set of substrate and product metabolites according to the reaction catalysed by the enzyme as defined in KEGG (Greenblum *et al.*, 2012), using KEEG API (application programing interface). The KO to R (reaction) link (http://rest.kegg.jp/link/Ko/reaction) was, alternatively, used to associate each individual KO to the corresponding reaction(s) following the R to RP (reaction pair) link (http://rest.kegg.jp/link/reaction/RP), thereby, associating individual Rs to their corresponding main reaction pair(s). The RPAIR annotation was specifically chosen to ignore unspecific compounds of reactions (water, energy carriers and cofactors), thereby, only taking into account the main compounds of a reaction. In the context of a metabolic network reconstruction of *Escherichia coli*, the use of KEGG RPAIR allows recovery of 93 % of the reference pathways, thereby highlighting its usefulness for comprehensive metabolic network reconstructions (Faust *et al.*, 2009). Finally, the RP-C (compounds) link (http://rest.kegg.jp/list/RP) was used to associate individual RP to corresponding pair(s) of compounds. Then, a search was performed using an in-house Perl program to identify all enzyme pairs in which a product metabolite of one enzyme is a substrate to the other. Undirected network edges were then drawn between each identified enzyme pair, visualised and analysed using the

Cytoscape software, employing a spring embedded layout.

4.2.8. Constraint-based modeling

In order to assess the reliability of the reconstructed metabolic network, a constraint-based community-wide metabolic model was built taking as input all KOs included in the global community-wide metabolic network (reconstructed from the entire sample set of KOs identified in the combined seasons metagenomic/metatranscriptomic datasets) as well as additional transporters which were added manually. The constraint-based model was created using the COBRA Toolbox for MATLAB (Becker *et al.*, 2007). The consistency of the model, which is defined as the proportion of metabolic reactions forming a connected metabolic core able to carry non-zero fluxes, was tested and further refined by the addition of a minimal core set of reaction and transporters using the FASTCORE algorithm (Vlassis *et al.*, 2013). FASTCORE is a generic algorithm for reconstructing context-specific metabolic network models from global genome-wide metabolic networks. The algorithm considers as input a core set of reactions that are known to be active in the context of interest and it searches for a flux consistent sub-network of the global network that contains all reactions from the core set and adds a minimal set of additional reactions (Vlassis *et al.,* 2013).

4.2.9. Topological network analysis

Topological features of each KO within the global community-wide metabolic network reconstruction were computed using the Cytoscape Network-Analyzer plug-

in (Smoot *et al.*, 2011) by considering an undirected network. The different features measured included network modularity (this relates the tendency of a network to be divisible into subgraphs which are densely connected), node degree (reflects the number of edges connected to a node), betweenness centrality (reflects the number of shortest paths from all vertices to all others by passing through a node), neighborhood connectivity (relates the average connectivity of all neighbors of a node) and clustering coefficient (reflects the proportion of all existing edges between a node's neighbours).

4.2.10. Data treatment and statistical analyses

Data treatment and statistical analyses of metabolomic and lipidomic data Changes in abundance of non-polar metabolites among cellular compartments and seasons were assessed using hierarchical clustering performed in Multiple Experiment Viewer (MeV) v4.9 on mean centered metabolite abundances using the Pearson product moment correlation coefficient and average linkage clustering. Changes in measured environmental parameters (air and water temperatures, pH, total suspended solid, nitrate, ammonium and total phosphate; **Table 4.1**) were correlated with variations in extracellular non-polar metabolite abundance patterns using the BIOENV statistical procedure (Clarke & Ainsworth, 1993; Oksanen *et al.*, 2007).

4.3. Results and discussions

4.3.1. Comparative metabolomic and lipidomic analyses

In order to assess compositional differences in lipid content within both temporally resolved autumn and winter LAO-enriched microbial communities, comparative metabolomic and lipidomic analyses of dates, which were showing significant differences in microbial community composition and function, were carried out (Section 2.3.2). The sampled winter LAO-enriched microbial communities were dominated by *Microthrix parvicella* and autumn communities were dominated by organisms closely related to *Alkanindiges* spp. and *Acinetobacter* spp. (Chapter 2, Conclusion and discussions). For the two sampling dates (Section 4.2.1), triplicate extra- and intracellular non-polar metabolites were extracted and analysed using various GC-MS and LC-MS methods (Section 4.2.3). LC-MS was preferred to GC-MS analysis because of its ability to provide reliable quantification of known, pre-selected metabolites (Section 1.2.2.4**)**.

Total ion chromatograms (TICs) obtained after GC-MS analysis, performed by the Metabolomics group at Luxembourg Centre for Systems Biomedicine (LCSB, Luxembourg), were deconvoluted and the generated data matrices were matched to a dedicated in-house metabolite library using the MetaboliteDetector software (Hiller *et al.*, 2009), as previously described (Section 3.2.7). The variation of metabolite profiles (**Supplementary Table VII**) results in a clustering of each individual fraction considered, cellular compartments as well as of the different LAO-enriched microbial communities (**Figure 4.5, A**). The differences between the metabolite profiles of cellular compartments are mainly explained by the first principal component (PC1),

while the dissimilarity between seasonal differences in the exometabolome and LAO-enriched microbial communities is explained by the second principal component (PC2). As previously highlighted (Section 2.3.1.2) the palmitic acid (C16:0) and oleic acid (C18:1) were the most abundant metabolites in both intra- and extracellular compartments in the two sampling dates. In addition, after hierarchical clustering of the abundances of non-polar metabolites (**Supplementary Table VII)**, a clear difference is observed between metabolites detected in the extracellular compartment of the two temporally resolved sampling dates (**Figure 4.5, B**), suggesting that seasonal changes have a more pronounced effect on the metabolites present around the LAO-enriched microbial communities than on the metabolic processes occurring within the LAOs. Changes in physico-chemical factors such as air and water temperatures, pH, total suspended solid, nitrate, ammonium and total phosphate between the seasons (**Figure 4.1, Table 4.1**), were used to define the main factors driving the observed differences. These were determined, using the BIOENV statistical procedure (Section 4.2.3). Significant correlations emerged, despite the use of highly complex metabolomic datasets comprised of hundreds of variables. From the variables tested, 'air temperature' and 'water temperature" (with Spearman's rank correlation coefficients of greater than 0.85) were the only physicochemical factors to show consistent correlation with the non-polar metabolite abundance patterns observed in the extracellular compartments. These results confirm the importance of temperature on the substrates avialability for the LAO-enriched microbial community.

Additional metabolomic analyses were carried out targeting non-polar metabolites to assess potential differences in lipid accumulation of these two seasonally resolved LAO-enriched microbial communities. In order to select the most favourable conditions for lipid accumulation, a comparative lipidomic analysis focused on

triacylglycerols (TAGs) and long chain fatty acids (LCFAs) was carried out on these winter and autumn LAO communities. LC-MS/MS analysis targeted on TAGs, was carried out by the Proteomics group at the Institute for Systems Biology (ISB, Seattle, USA). Among different complex lipid structures, LC-MS/MS analysis of co-extracted intracellular non-polar metabolites resulted in the identification of different TAGs (present in more than 0.1 % of the total ion volume, **Supplementary Table VIII**) including TAGs 34:0 and TAG 66:0 in both LAO communities (**Figure 4.5, C**). The presence of TAG inclusion in *Microthrix parvicella* has recently been hypothesised based on a metabolic network reconstruction (McIlroy *et al.*, 2013). Only a small amount of TAGs were detected within the extracellular compartment (**Supplementary Table VIII**). These results may be explained by a higher dilution rate of complex lipid structures within the extracellular compartment, compared to the accumulation of such metabolites within the microbial biomass. Further selection of TAGs, solely identified in the intracellular compartments, resulted in the detection of dominant TAGs from TAG 44:0 to TAG 54:5 (**Figure 4.5, C)**. A clear difference in TAG composition between autumn and winter LAO communities is observed. The TAG distribution in autumn seems to contain proportionally shorter and fully saturated LCFAs (i.e. TAGs 44:0, 48:0 and 46:0) while, in winter, samples of the LCFAs appear to be longer and unsaturated (i.e. TAGs 52:2, 54:3 and 54:4).

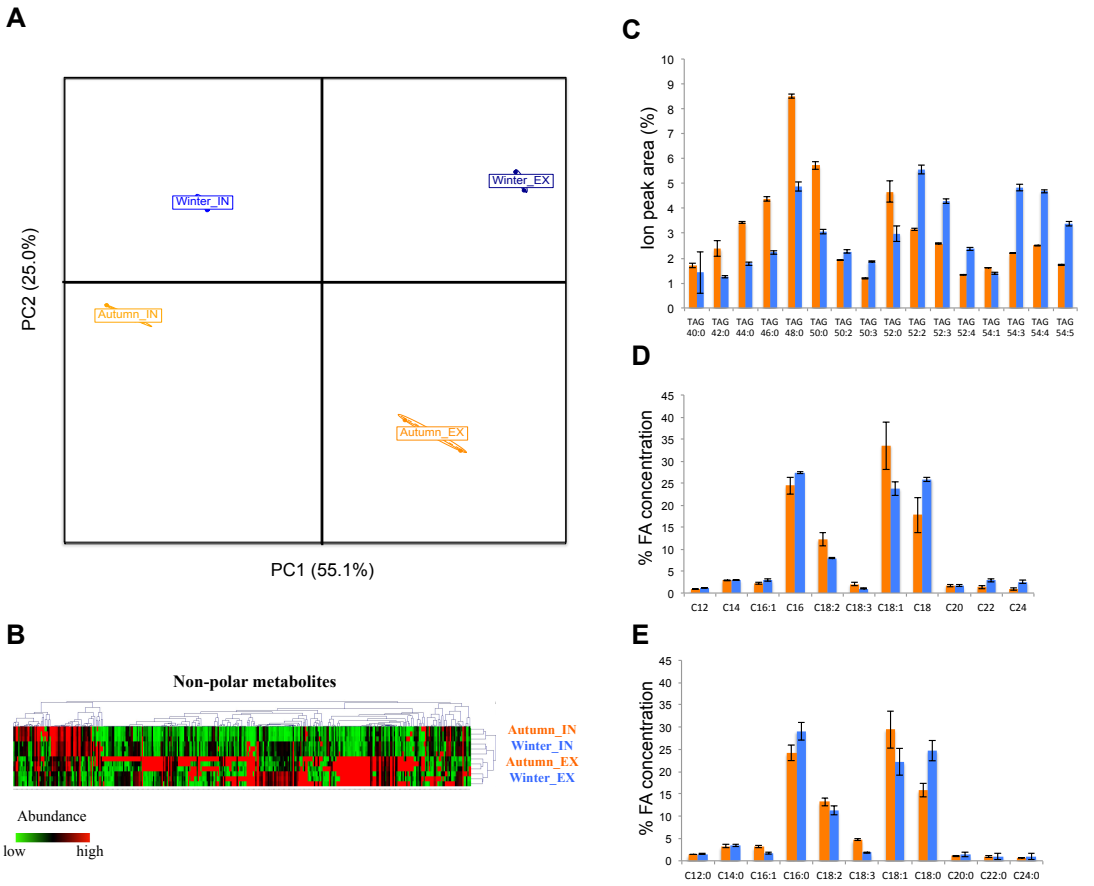

Figure 4.5 Metabolomic and lipidomic analysis of non-polar metabolite fractions extracted from the studied autumn (orange) and winter (light-blue) microbial communities. (**A**) Scatter plot of the two first principal components obtained using principal component analysis (PCA) of the normalised non-polar metabolites data derived from the extra- (EX) and intracellular (IN) compartments respectively. (**B**) Hierarchical clustering of the normalised metabolomics data derived from the extra- (EX) and intracellular (IN) compartment using the Pearson product moment correlation coefficient. (**C**) Bar chart of triacylglycerol (TAG) abundances from the LAO biomass. (**D**,**E**) Bar charts of long-chain fatty acids (LCFA) derived from the intracellular (**D**) and extracellular (**E**) compartments, respectively. LCFAs with less than 1 % of the total LCFA concentration are omitted. The lipid number c:d reflects the number of carbon atoms (c) and number of double bonds (d) in the three esterified fatty acid chains (**C**) or in single LCFA chains (**D**,**E**).

GC-MS/MS analysis targeting LCFAs was carried out by the Analytical Chemistry platform at the Centre de Recherche Public Gabriel Lippmann (CRP-GL, Luxembourg). The absolute quantification of non-polar LCFAs by GC-MS/MS analysis of both temporally resolved LAO communities, (**Supplementary Table IX**) confirmed the presence of distinct LCFAs, mainly dominated by LCFAs from C16 to C18 (**Figure 4.5, D&E)**, within similar proportions between the intra- and extracellular compartments compared to the measurement previously obtained (Section 2.3.1.2**)**. These results strongly suggest that extracellular LCFAs present in the wastewater are taken up directly by LAOs without conversion. The differences in LCFA lengths and numbers of saturations reveal that saturated palmitic acid (C16:0) and stearic acid (C18:0) are the most abundant within the winter LAO community sample, while unsaturated oleic acid (C18:1) and linoleic acid (C18:2) are prevalent in the autumn LAO community. The correspondance between TAGs and LCFAs structures confirms the potential of LAOs to synthesis TAGs using LCFAs and help to differentiate natural occurring TAGs from other lipid body structures present in the cell membrane, for instance. However, single-cell analytical approaches are required to conclusively demonstrate this.

4.3.2. Global and season-specific community-wide metabolic networks

Global and season-specific community-wide metabolic networks were obtained after assembly and annotation of orthologous genes (KO) of a pooling of combined and single season metagenomic and metatranscriptomic datasets (Section 4.2.5). The global community-wide metabolic network comprises 2,104 KOs (nodes) from a total

of 7,270 annotated KOs in the obtained dataset (**Supplementary Table X)** interacting with 62,903 redundant metabolites (edges, **Figure 4.6, A**). The global community-wide metabolic network includes 2,834 unique reactions. The selection of KOs considered was relatively low, but this is due to the low proportion of KOs representing genes coding proteins able to catalyse metabolic reactions, as defined in the KEGG database. In particular, the total metabolic pathways map accounts for 14.31 % of the KOs available in KEGG. The other KOs are related to reactions of transport, signaling, regulation or secretion. The position of particular KOs, involved in the tricarboxylic acid cycle and synthesis of amino acids, as highlighted on the global community-wide metabolic network demonstrates that the metabolic reconstruction preserves the relative position of individual pathways with respect to each other, expect for some genes which have a more off-centre position, due to their connectivity to other pathways (**Figure 4.6, B**).

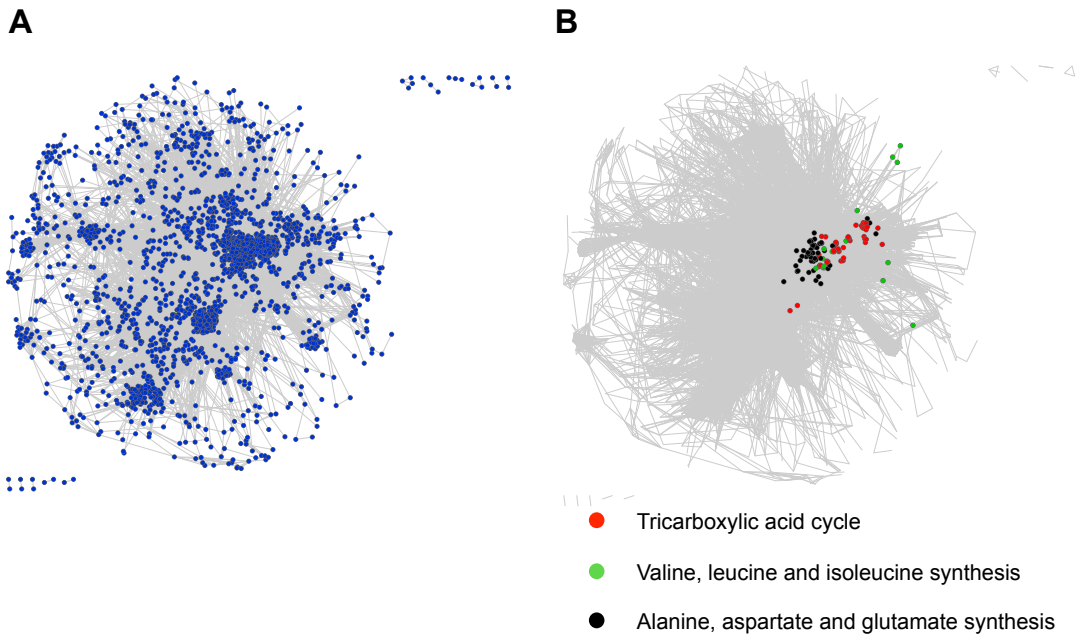

Figure 4.6 Global community-wide metabolic network reconstruction. (**A**) Network constructed from homologous gene information. Nodes (2,104) represent KEGG orthologous genes (KOs) and edges (62,903) connect KOs that are linked by an intermediate metabolite. (**B**) Highlights of the presence and the position of particular KOs involved in the tricarboxylic acid cycle (TCA) and synthesis of amino acids.

In order to assess the reliability of the community-wide metabolic network reconstruction, a constraint-based model reconstruction analysis (COBRA) was performed, considering as input all KOs contained in the previously designed network (Section 4.2.8). The addition of 126 transport reactions from ABC transporters identified in the metagenomic gene set led to an overall consistency of 53.69 % with 56.07 % of the core set reactions present within the reconstructed community-wide metabolic network (**Table 4.3**).

Table 4.3 Constraint-based metabolic network data after model reconstruction using the COBRA toolbox and carrying out model refinement using FASTCORE.

	COBRA Model	FASTCORE model
Core set of reactions	1,589	2,045
Total reactions	2,834	2,834
Additional reactions	0	657
Additional transporters	126	35
Total reactions considered in the model	1,715	2,737
Fraction of core reactions (%)	56.07	72.16
Consistency (%)	53	100

To further test and refine the consistency reached by the constraint-based modelling of all KOs involved within the previously design community-wide metabolic network, the FASTCORE algorithm (Vlassis *et al.,* 2013) was used (Section 4.2.8). This resulted in the addition of 36 transport reactions, as well as 657 metabolic reactions to reach an overall core reaction rate of 72.16 % (**Table 4.3**). By comparison, the global

reconstruction of human metabolism Recon1 (Duarte *et al.*, 2007) and recently published Recon2 (Thiele *et al.*, 2013) reached a core set of reaction of 65.98 % and 78.41 %, respectively. It is noteworthy, that both reconstructions required extensive manual curation and the automated community-wide reconstructions used in this work, therefore, compares very favourably with these.

Interestingly, the pathways with lower level of consistency in the original network are associated with the biodegradation of variable xenobiotic compounds such as polycyclic aromatic hydrocarbons (PAH), benzene derivates, toluene, steroid, naphthalene, styrene and dichlorodiphenyltrichloroethane (DDT; **Supplementary Table XI**). The introduction of xenobiotic compounds into the environment is, largely, due to industrial activity, such as the use of fossil fuels, solvents, surfactants, pesticides, and pharmaceuticals. Xenobiotic compounds are toxic for many organisms in the ecosystem, poorly biodegradable and persist in the environment for extended periods of time (Field *et al.*, 2000). The majority of xenobiotic compounds is slowly or not fully degraded through biological wastewater treatment (Majewsky *et al.*, 2010). The absence of the majority of the reactions involved in the biodegradation of xenobiotic compounds from the obtained genomic and transcriptomic data highlights the deficiency of microbial communities in this WWTP (or at least the enriched LAO communities) to degrade these compounds, or the absence of certain pollutants in the studied wastewater streams. Alternatively, the microbial communities may use other genes which have not yet been characterised and, thus, are not yet represented in KEGG.

The autumn and winter community-wide metabolic networks derived using both the metagenomic and metatranscriptomic data (**Supplementary Table XII**) show similar

structures and share 1,592 common KOs (i.e. 82.9 and 94.5 % of the total annotated KOs in autumn and winter respectively, **Figure 4.7**). However, the autumn metabolic network harbours a larger fraction of known genes, comprising 236 additional KOs (nodes), compared to the winter reconstructed metabolic network.

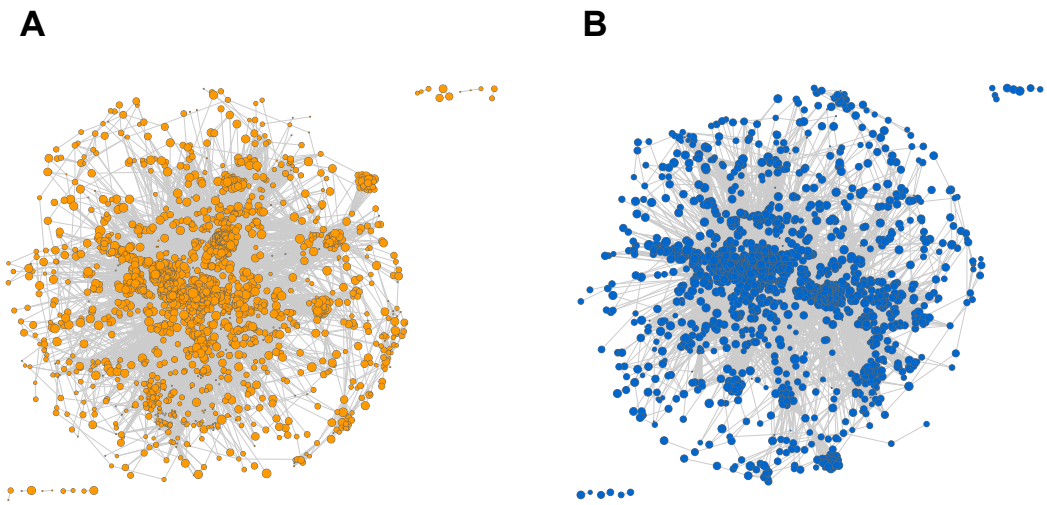

Figure 4.7 Season-specific community-level metabolic networks. Nodes represent orthologous enzyme coding genes (KOs) and edges connect KOs that share common metabolites. The size of nodes are proportional to the $\log_2$ of the number of KO genes number (n). (**A**) Autumn reconstructed metabolic network comprising 1,920 nodes and 54,430 edges. (**B**) Winter reconstructed metabolic network comprising 1,684 nodes and 43,591 edges.

The comparative analysis of the two metagenomic/metatranscriptomic datasets (**Supplementary Table XII**) contained within individual community-wide metabolic networks highlights 328 and 92 unique KOs in autumn and winter LAO communities, respectively. This analysis suggests limited difference in terms of genetic potential and gene expression between the two seasonally resolved LAO communities despite stark differences in community composition (Section 2.3.2.1).

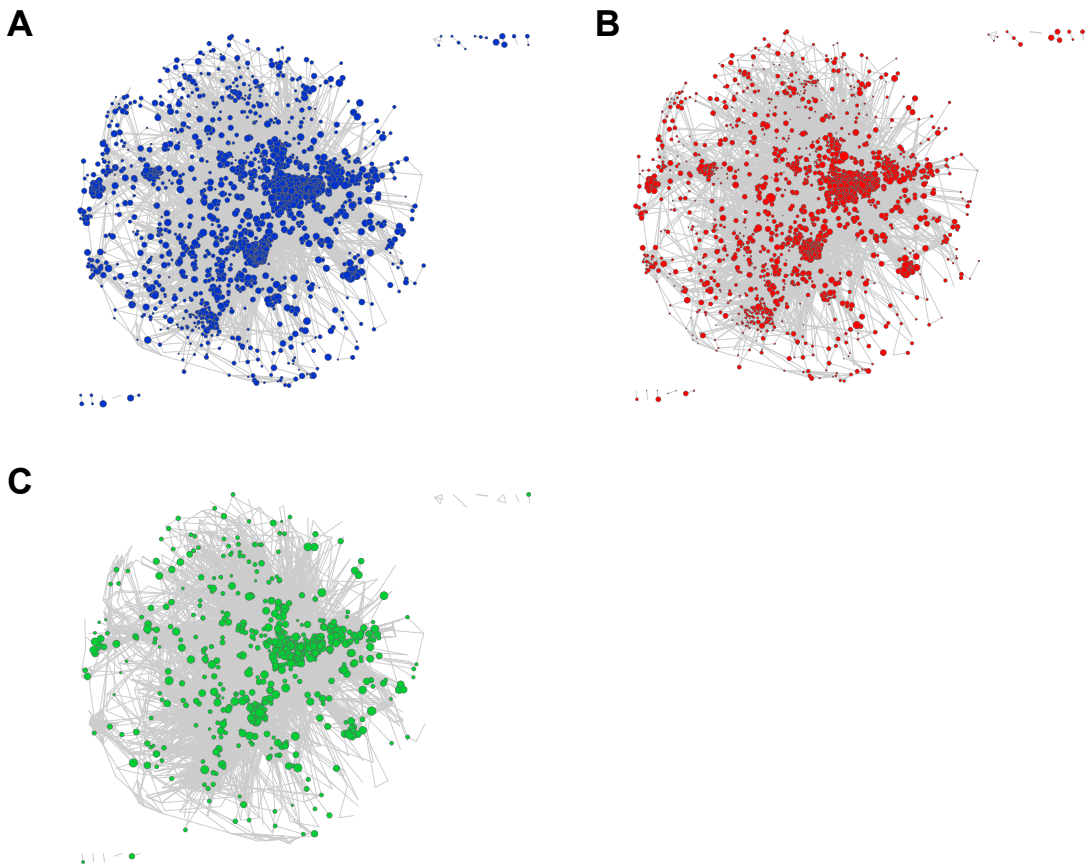

Figure 4.8 Global community-wide metabolic networks at different omic levels. Sizes of nodes are proportional to (**A**) $\log_2$ of normalised gene copy numbers (**Eq. 1**, KOA_{DNA}; blue nodes), (**B**) $\log_2$ of corresponding normalised transcript levels (**Eq. 1**, KOA_{mRNA}; red nodes) and (**C**) $\log_2$ of corresponding normalised protein frequency (green nodes), within the autumn LAO community metabolic network reconstruction.

The nodes within the global community-wide metabolic network were weighted by the $\log_2$ of normalised abundances derived from gene copy numbers, corresponding transcript levels (**Supplementary Table X**) and protein frequency (**Supplementary Table XIII**), respectively (**Figure 4.8**). The mapping of these values onto the represented nodes highlights specific pathways with variable KO abundances within each individual season and each biomolecular level (**Figure 4.9**). The congruency between the metagenomic and metatranscriptomic data is high, at 93.53 % and

93.06 % within the respective autumn and winter community-wide metabolic networks. However, the coverage of nodes from the proteomic data is lower, as 619 (28.83 %) and 589 (27.43 %) nodes within the community-wide metabolic network are covered in autumn and winter LAO community metabolic network reconstructions, respectively. The rather limited congruency reached using the proteomic data is due to current limitations in proteomic technologies which result in only abundant proteins being typically well detectable.

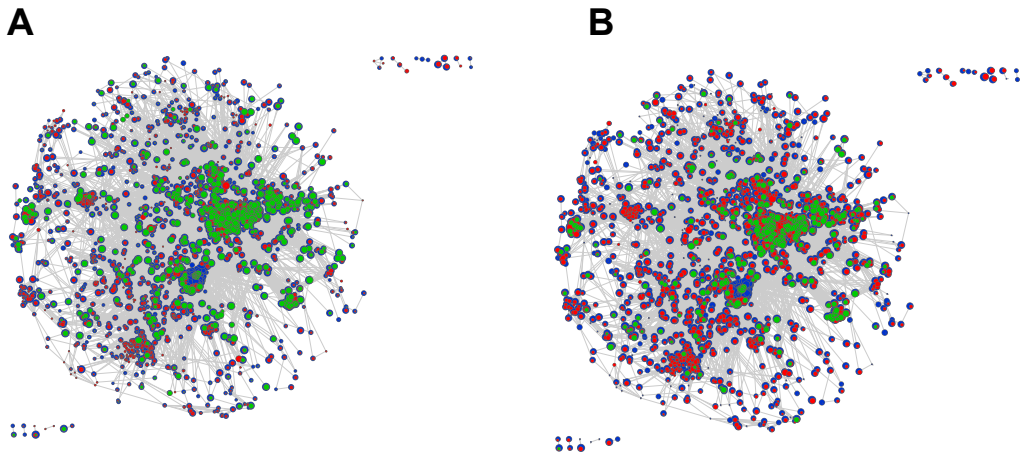

Figure 4.9 Global community-level metabolic networks with integrated multi-omic measurments of the (**A**) autumn LAO community and (**B**) winter LAO community, respectively. Sizes of nodes are proportional to the $\log_2$ of normalised KO gene copy numbers (KOA_{DNA}) in blue, $\log_2$ of corresponding normalised transcript levels (KOA_{mRNA}) in red nodes and $\log_2$ of corresponding normalised protein frequency in green.

4.3.3. Network-wide KO gene number analysis

In order to detect dominant metabolic pathways present within the global community-wide metabolic network reconstruction, the top one hundred most abundant functional

genes included in the global community-wide metabolic network (**Figure 4.10**) were considered. They all fall into five metabolic categories, according to their KEGG annotation. Carbohydrate metabolism (27.40 %) dominates the LAO-enriched microbial community-wide metabolic transformations (**Figure 4.10**). The amino acid (18.24 %), nucleotide (16.84 %), energy (14.40 %) and lipid (12.71 %) metabolic pathways exhibit similar proportions. However, the energy and lipid metabolic reaction categories are highly dominated by nitrogen (75.62 %) and fatty acid-glycerolipid (75.84 %) metabolic transformations, respectively (**Figure 4.10, Supplementary Table X**).

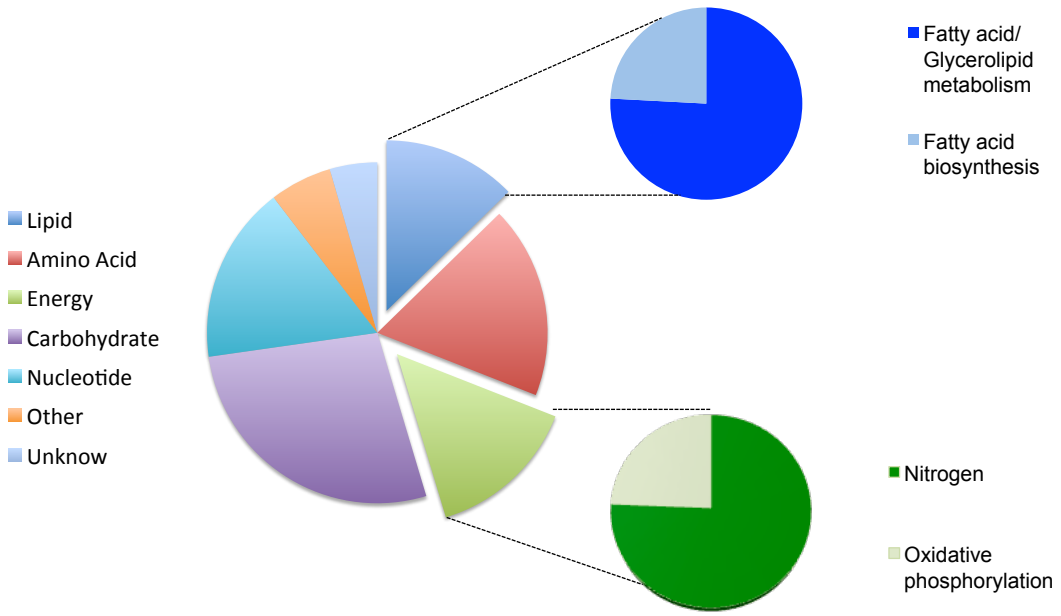

Figure 4.10 Pie chart of the dominant metabolic pathways derived from the gene annotations of the combined sequence assembly for both sampling dates, considering the top one hundred most abundant functional KO gene numbers (n) within the global community-wide metabolic network reconstruction.

The dominance of lipid-associated pathways reinforces the specialisation of the investigated microbial community in lipid metabolism. Nitrogen metabolism accounts for 6.04 % of the total number of genes considered. These results emphasise the importance of specialised bacteria involved in the nitrogen removal in WWTPs.

4.3.4. Identification of overexpressed genes within each season

Relative gene expression (RGE) was determined for each individual KO included in the global community-wide metabolic network, by the determination of the $\log_2$ ratio of the relative abundance of the KOs at the metatranscritptomic (KOA_{mRNA}) and at the metagenomic level (KOA_{DNA}; Section 4.2.5, **Eq.2**; **Supplementary Table X**). Relative gene expression was assessed by comparing gene copy numbers to corresponding transcript levels (Section 4.2.5).

It was found that a very small proportion of KOs were overexpressed, 1.64 and 1.92 % in autumn and winter, compared to a vast majority of KOs underexpressed, 92.30 and 92.69 %, respectively. Importantly, around 40 % of the overexpressed KOs are common between the two seasons. Among them, we found that KOs involved in nitrification (K10944/45/46, **Table 4.4**), especially the subunits of the ammonia monooxygenase enzyme (Amo), had a strong relative expression, as previously described by Yu and colleague in activated sludge (Yu & Zhang, 2012). These particular genes are known to play a key role in the first step of nitrification carried out by aerobic ammonia-oxidising bacteria (AOB), mainly belonging to *Nitrosomonas* and *Nitrosospira* species (Yu & Zhang, 2012). Additionally, within the glycerolipid metabolism, K01046 (**Table 4.4)** which codes for the triacylglycerol lipase enzyme

shows a particularly high overexpression. Furthermore, from the 6,231 genes belonging to this KO, two genes have been successfully matched to the draft genome sequence of *Microthrix parvicella* (strain BIO17-1, Muller *et al.*, 2012). The presence of such enzymes within LAO-enriched microbial communities has recently been found essential for lipid accumulation in a metabolic model reconstruction of *Microthrix parvicella* (McIlroy *et al.*, 2013), but never identified and measured as overexpressed at the community level before. The constant overexpression of the previously mentioned KOs involved within the oxidation of ammonium and hydrolysis of TAGs, nicely emphasises the specificity of the WWTP biomass to degrade mainly inorganic and organic compounds present in wastewater, i.e. ammonia and lipids.

Table 4.4 Overexpressed genes in the autumn and winter LAO-enriched microbial communities.

KO	$Log_2\left(\frac{KOA_{\mathrm{mRNA}}}{KOA_{\mathrm{DNA}}}\right)_{autumn}$	$Log_2\left(\frac{KOA_{\mathrm{mRNA}}}{KOA_{\mathrm{DNA}}}\right)_{winter}$	Metabolic reaction
K10944/45/46	2.10	2.14	Ammonia + Oxygen + Ubiquinol <=> Hydroxylamine + H_2O + Ubiquinone
K01046	4.90	5.68	Triacylglycerol + H_2O <=> Fatty acid + 1,2 diacyl-glycerol

4.3.5. Identification of genes associated with specific seasons

To identify particular KOs associated with a given season (SAF), the $\log_2$ ratio of the KO abundances at the genomic (SAF_{DNA}) and transcriptomic levels (SAF_{mRNA}) were

calculated (Section 4.2.5, **Eq. 3&4**; **Supplementary Table X**), as well as KO overrepresentation levels (SOR) between both seasons (Section 4.2.5, **Eq. 5**; **Supplementary Table X**).

Since the richness of distinct KOs is higher in autumn LAO communities, the comparative analysis of KO abundance at the genome and transcriptome level reveals extensive number of winter-associated KOs. At the genome and transcriptome levels, 62.75 and 64.94 % of the KOs are found to be winter-associated, respectively. At the genome level, these results might be explained by the superior species richness in winter LAO communities (Section 2.3.2.1), as overall gene abundance is higher in winter.

The comparative analysis of KO gene representation levels shows that a comparable proportion of KOs are associated with autumn (25.08 %) or winter (17.04 %) LAO communities. Although a large number of pathways appear to contain seasonally associated KOs, their position within the global community-wide metabolic network is not homogenously distributed and some metabolic pathways show variable seasonal association. For example, it was observed that the phosphotransferase system (PTS) is highly represented among the autumn-associated KOs (**Figure 4.11, A&B)**. This system is, among other function, responsible for the transport of glucose from the extracellular compartment into the cell with the addition of a phosphate group to produce glucose-6-phosphate (**Figure 4.12**, **A**; Götz & Goebel, 2010).

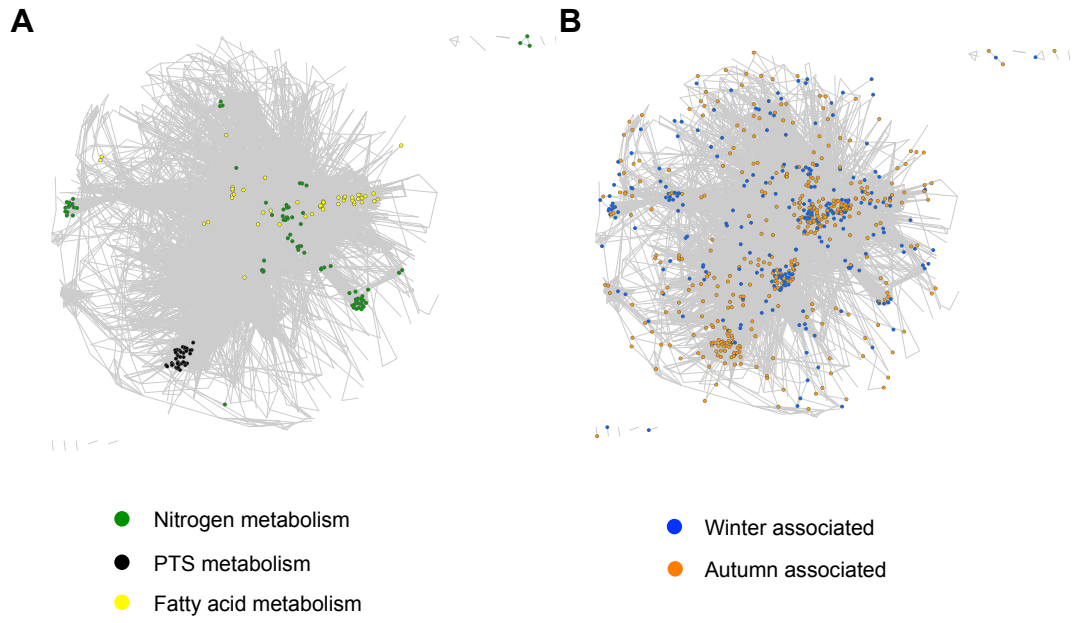

Figure 4.11 Overrepresentation of genes in the global community-wide metabolic network mapped with (**A**) metabolic pathways of interest, (**B**) the seasonally associated genes (**Eq.5**).

Interestingly, within the PTS complex, particular maltose and glucose transporters KOs (K02790/91) are overrepresented in the autumn LAO community. Furthermore, other autumn associated glucose-specific transport KOs (K02777/78/79, **Figure 4.12**, **A**), involved in the transport of glucose from the extracellular compartment to the intracellular compartment and subsequent synthesis of glucose-6-phosphate is corroborated by the polar metabolomics dataset analysed by GC-MS (Section 4.2.3; **Supplementary Table XIV**). Indeed, 10 times more putative glucose-6-phosphate was measured within the sampled autumn LAO community, compared to winter LAO community (**Figure 4.12, B**). These results might suggest a preference of carbohydrates including glucose and maltose as carbon source of the autumn LAO community. Thus, the autumn LAO community is considered to be less dependent on lipids than the winter LAO community and may be enriched in glycogen

accumulating organisms which are known to also be enriched in activated sludge plants with alternating anaerobic/aerobic tanks (Kong *et al.,* 2006).

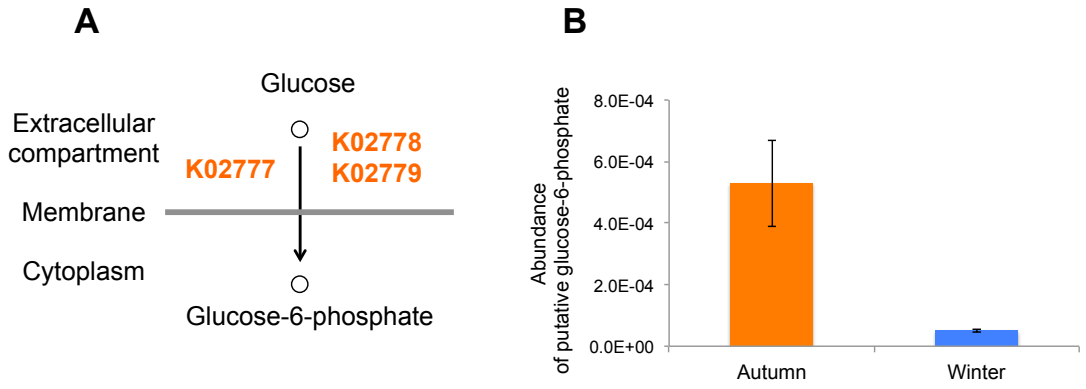

Figure 4.12 Enrichment of glucose-specific phosphatransferase enzyme IIA in the autumn LAO community and corresponding intracellular levels of putative glucose-6-phosphate. (**A**) KOs coding for the glucose-specific IIA component protein which are highly represented in autumn. (**B**) Bar chart of the abundance of putative glucose-6-phosphate in the intracellular compartment of the autumn and winter LAO communities, respectively.

Interestingly, an important variation in the seasonally associated KOs is apparent within the nitrogen and fatty-acid/glycerolipid metabolic networks. Within these two pathways, various KOs are associated with different seasons. More specifically, certain KOs catalysing the same reaction are differentially associated with the two seasons (**Figure 4.13**). In the nitrification reactions of the nitrogen metabolism, KOs constituting the nitrate reductase are, particularly, overrepresented in winter LAO communities. However, a nitrite reductase gene (K00363) is highly represented in autumn LAO community while in winter a ferredoxin-nitrite reductase (K00366) is more prominent. This difference suggests that winter LAO communities have a requirement for reduced ferredoxin as a co-factor to metabolise ammonia. Within the

nitrogen metabolism the nitrification pathway seems more represented in the winter LAO community compared to autumn LAOs (**Figure 4.13, A**). Interestingly, particular ammonium transporters (K03320), as well as nitrogen regulatory proteins (K04751/52), have been detected to be especially overrepresented in autumn LAO communities. These two KOs, recognised for their pivotal role in control of prokaryotic nitrogen metabolism (Arcondéguy *et al.*, 2001), have in particular, been found to be co-expressed under nitrogen limiting conditions (Blauwkamp & Ninfa, 2003). Interestingly, specific archaeal KOs coding for nitrate/nitrite transporters (K15576/77/78/79) have been found to be highly represented in the autumn LAO community. These results suggest that archaea are more involved in the nitrogen metabolism in autumn compared to winter.

Regarding the combined fatty acid and glycerolipid metabolism, we were able to identify, at the gene level, nearly all reactions involving the catabolism of triacylglycerol into fatty acids, β-oxidation of LCFAs and triacylglycerol synthesis from glycerol and fatty acyl-CoA via the Kennedy pathway (**Figure 4.13, B**).

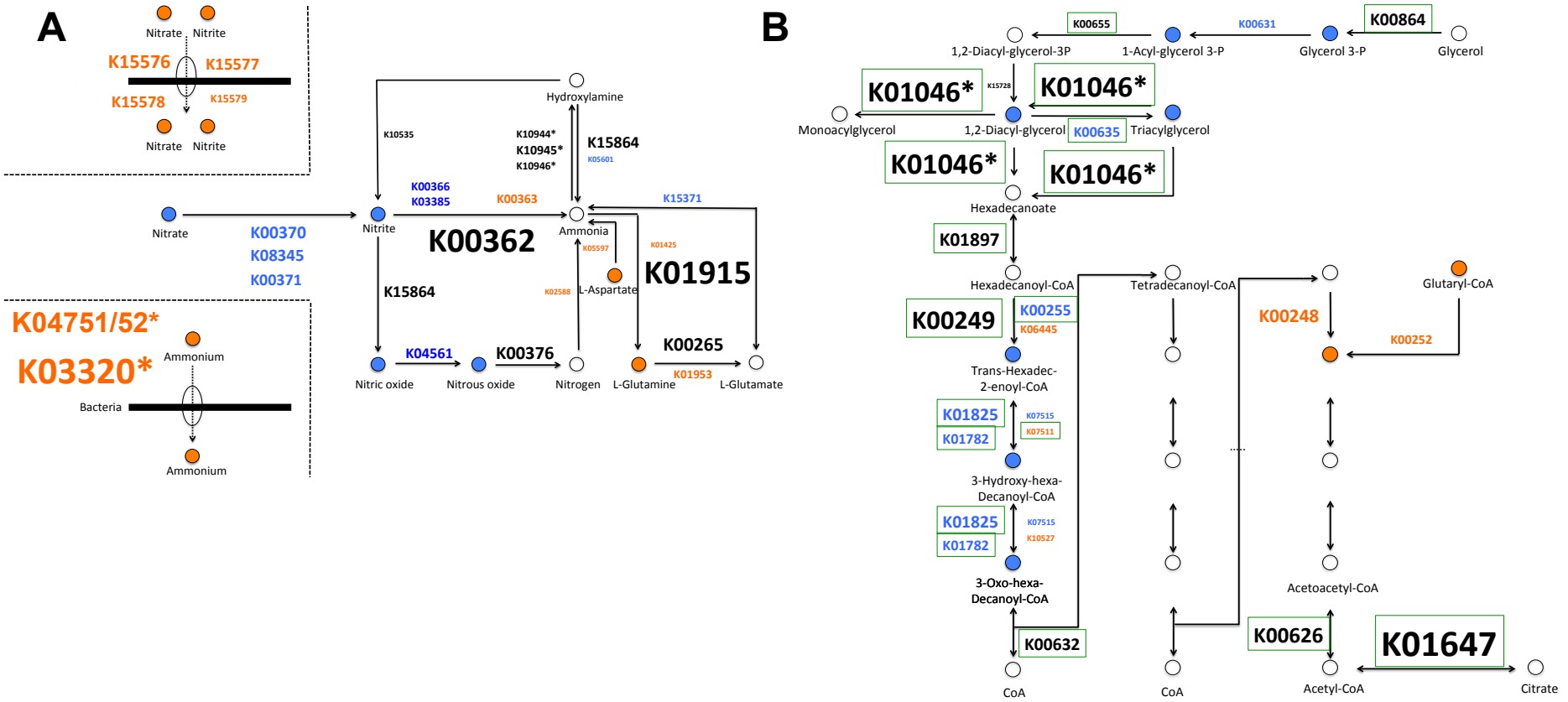
A
Nitrate
Nitrite
K15576
K15577
K15578
K15579
Nitrate
Nitrite
Hydroxylamine
K10535
K10944*
K10945*
K10946*
K15864
K05601
K00366
K03385
K00363
K15371
Nitrate
K00370
K08345
K00371
Nitrite
K00362
Ammonia
K05597
K01425
K02588
L-Aspartate
K01915
K15864
K04751/52*
K03320*
Ammonium
Bacteria
Ammonium
K04561
K00376
K00265
K01953
Nitric oxide
Nitrous oxide
Nitrogen
L-Glutamine
L-Glutamate
B
K00655
K00631
K00864
1,2-Diacyl-glycerol-3P
1-Acyl-glycerol 3-P
Glycerol 3-P
Glycerol
K01046*
K15728
K01046*
Monoacylglycerol
1,2-Diacyl-glycerol
K00635
Triacylglycerol
K01046*
K01046*
Hexadecanoate
K01897
Hexadecanoyl-CoA
Tetradecanoyl-CoA
Glutaryl-CoA
K00249
K00255
K06445
K00248
K00252
Trans-Hexadec-
2-enoyl-CoA
K01825
K01782
K07515
K07511
3-Hydroxy-hexa-
Decanoyl-CoA
K01825
K01782
K07515
K10527
3-Oxo-hexa-
Decanoyl-CoA
Acetoacetyl-CoA
K00626
K01647
K00632
CoA
CoA
Acetyl-CoA
Citrate

Figure 4.1 Abundance and representation of orthologous genes in metabolic pathways according of the two seasons. (**A**) Nitrogen metabolism and (**B**) lipid metabolic pathways including KO gene numbers with the associated enzymatic reactions. The KO gene numbers are proportional to the front size of the respective KO numbers (**Supplementary Table X**). The asterisk highlights particular overexpression of the corresponding KO. The color-coding corresponds to the seasonal representation of KOs, in light-blue KOs overrepresentated in winter and in orange KOs overrepresentated in autumn. The green border, in pane **B**, highlights particular KOs for which genes have been detected in the *Microthrix parvicella* BIO17-1 draft genome (Muller *et al*., 2012).

Remarkably, the majority of the KOs involved in lipid metabolism include genes found to be present within the draft genome sequence of *Microthrix parvicella* (strain BIO17-1, Muller *et al.*, 2012), which confirms its pronounced ability for lipid metabolism and pivotal role within LAO-enriched microbial communities. Furthermore, the biosynthesis of triacylglycerol and β-oxidation appears more highly represented in the winter LAO community, which corroborates the dominance of this particular organism in winter. Interestingly, within the β-oxidation module of the fatty acid metabolism, a particular long-chain-acyl-CoA dehydrogenase (K00255), encoded by the *Microthrix parvicella* genome seems to predominate in the winter LAO community among other acyl-CoA dehydrogenases (K06445), absent from *Microthrix parvicella* genome and more highly represented in the autumn LAO community. This long chain acyl-CoA dehydrogenase may confer a competitive advantage to *Microthrix parvicella* for LCFA catabolism during the winter month.

4.3.6. TAG metabolism

Triacylglycerol lipase genes are among the most abundant KOs within the LAO-enriched microbial communities and, among them, two genes which are encoded by the *Microthrix parvicella* BIO-17.1 genome. Additionally, these genes exhibit a relatively high expression, when compared to gene copy numbers within both autumn and winter LAO communities and are also highly abundant at the protein level. As highlighted previously (Section 1.1.1.2), such lipases are interesting from a biotechnological point of view for the conversion of various lipid-rich feedstocks into biodiesel. Since LAOs have evolved to preferentially use lipids as substractes, their lipases must be highly efficient enzymes possibly with broad substrate specificity. The presence of such enzymes in *Microthrix parvicella* were first demonstrated *in situ* as surface associated extracellular lipases (Nielsen *et al.*, 2002) and they have also been identified in the *Microthrix parvicella* RN1 genome (McIlroy *et al.*, 2013).

Given the recent metabolic model reconstruction of *Microthrix parvicella* based on genomic and metagenomic data (McIlroy *et al.*, 2013) as well as the multi-omic datasets generated as part of this project, a refinement of possible TAG metabolism in *Microthrix parvicella* is proposed (**Figure 4.14)**. In the revised model, TAGs present in wastewater or associated to cell surfaces are hydrolysed into LCFAs and glycerol by triacylglycerol lipase, which supports the low abundance of TAG and high abundance of LCFA detected in the extracellular compartments. Recently, it has been demonstrated in EBPR systems, that *Microthrix parvicella* is able to utilise glycerol only when oleic acid is present as a co-substrate (Kindaichi *et al.*, 2013). The identification of glycerol kinase and glycerol 3-phosphate transport system genes

demonstrate that resulting glycerol, in the presence of LCFAs such as oleic acid, may be converted into glycerol-3-phosphate and then transported into the cell. The apparent hydrophobic nature of the LAO cell surface is then supposed to play a role within the attachment of resulting LCFAs (Nielsen *et al.*, 2002). The similarity of LCFA composition between the extra- and intracellular compartments suggests that LCFAs are transported and activated using long chain fatty acyl-CoA ligase (Sections 2.3.1.2 & 2.3.2.2). The identification of unique long-chain fatty acid transport within the genomic and proteomic dataset seems to confirm LCFA uptake (**Table 4.5**). However, long chain fatty acid transport gene have not be found within *Microthrix parvicella* draft genome and, given the high uptake and storage rate recorded for the LCFAs, more predominant active transport mechanisms are probably involved (McIlroy *et al.*, 2013). However, LCFA uptake mechanisms in Gram-positive bacteria are still unclear (Mcllroy *et al.,* 2013). In some gram-positive bacteria, fatty acids do not require FadL to enter the cell and may traverse the outer membrane by passive diffusion (Vance & Vance, 2012). The pronounced abundance of fatty acid-CoA ligase genes within all omic datasets as well as the *Microthrix parvicella* draft genome suggests that the Microthrix population uses specialised enzymes to activate different LCFAs (Section 2.3.2.2). The *Microthrix parvicella* BIO17-1 genome alone encodes 28 homologs of long chain fatty acid CoA ligases (Muller *et al.,* 2012) of which 10 non-redundant copies were detected in our integrated omic analyses, which in turn suggests that it uses specialised enzymes for the different LCFAs it accumulates. The abundance of these enzymes demonstrates than LCFAs efficiently pass the cell membrane since transport and activation of exogenous LCFAs have been demonstrate to be coupled in *E.coli* (Fujita *et al.*, 2007). The presence of glycerol-3-phosphate acyltransferase, acyl-sn-glycerol-3-phosphate acyltransferase and

diacylglycerol O-acyltransferase genes (*plsB, plsC* and *atf*) and associated proteins involved within the *de novo* synthesis of diacylglycerol-3-phosphate by the sequential acylation of glycerol-3-phosphate (G3P) confirms that the generated fatty acyl-CoAs might be sequentially attached to a glycerol-3-phosphate backbone to result in TAGs synthesis intracellularly with *Microthrix parvicella* (**Figure 4.14**, **Table 4.5**, **Supplementary Table X, XIII & XV)**. Interestingly, the third acylation step catalyzed by the *atf* gene (N°9, **Table 4.5**), is a unique enzymatic reaction of TAG biosynthesis as the two first acylation steps, coded by *plsB* and *plsC*, are also used for the synthesis of membrane phospholipids (Section 1.1.4.2). Thus, diacylglycerol O-acyltransferase (K00635) is considered to be the key gene in TAG biosynthesis (Alvarez *et al.*, 2013; Wältermann *et al.*, 2007) and *Microthrix parvicella* encodes several homologs (**Table 4.5**).

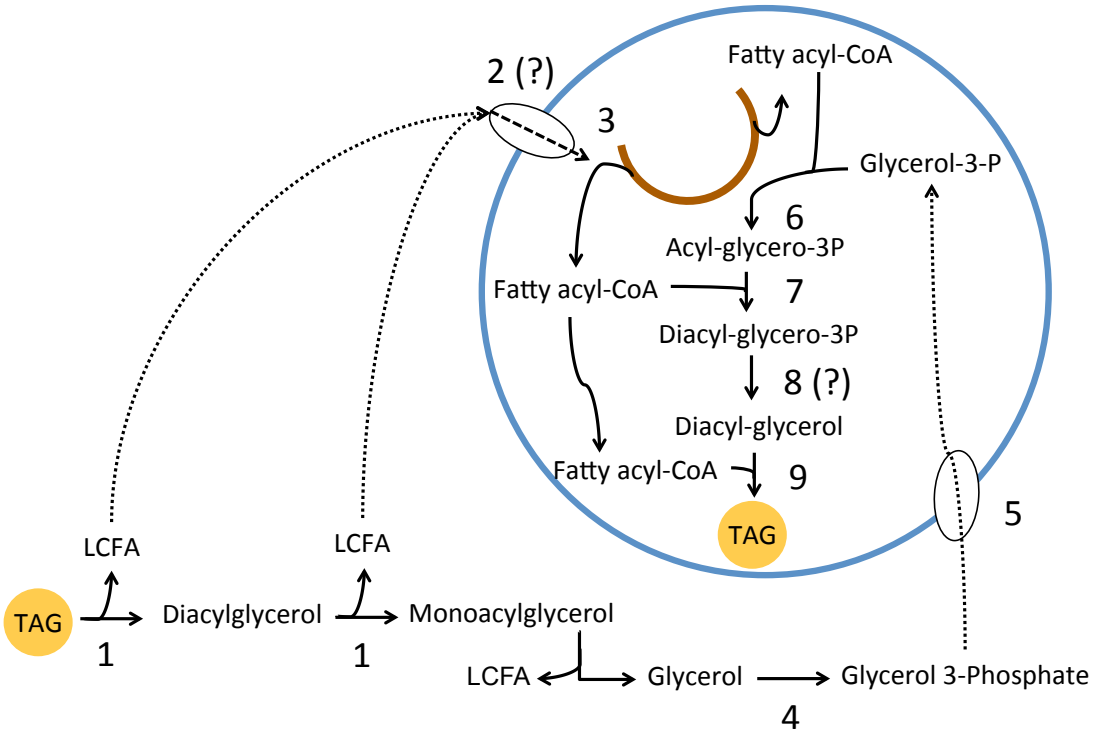

Figure 4.14 Proposed metabolic model for TAG accumulation in *Microthrix parvicella* under anaerobic conditions. Number corresponds to KOs assignment in **Table 4.2**. TAG: triacylglycerol, 3P: 3-phosphate.

Table 4.5 Summary of detected genes and corresponding protein frequencies related to TAG metabolism in the studied LAO communitiy and homologous genes identified in *Microthrix parvicella.* KO gene numbers are deducted from combined metagenomic and metatranscriptomic datasets, protein frequency from protein dataset and gene encoded by *Microthrix parvicella* strain BIO17-1 genome (Muller *et al*., 2012).

N°	KOs	Description (associated gene name)	KO gene number	Protein frequency[1]	Gene encoded by *Microthix parvicella*[2]
1	K01046	Triacylglycerol lipase	6,231	70	✓
2	K06076	Long-chain fatty acid transport protein (*fadL*)	249	1	×
3	K01897	Long-chain acyl-CoA synthetase (*fadD*)	2,178	18	✓
4	K00864	Glycerol kinase (*glpK*)	904	9	✓
5	K05813	Glycerol 3-phosphate transport system	230	14	✓
6	K00631	Glycerol-3-phosphate O-acyltransferase (*plsB*)	225	2	×
7	K00655	1-acyl-sn-glycerol-3-phosphate acyltransferase (*plsC*)	225	0	✓
8	K01080	Phosphatidate phosphatase (*pap*)	0	0	×
9	K00635	Diacylglycerol O-acyltransferase (*atf*)	549	25	✓

[1] Protein frequencies were deduced from KO frequency as described (Section 4.2.6, **Supplementary Table XIII**).

[2] ✓: denote the presence of the KO and ×: denote the absence of the KO (**Supplementary Table XV**)

However, as recently suggested (McIlroy *et al.*, 2013) a gene coding for a putative phosphatidate phosphatase (PAP) is absent from our data and has not been identified within the *Microthrix parvicella* BIO17-1 draft genome (**Figure 4.14**, **Table 4.5**). This may due to the low representation of prokaryote PAP genes, compared to eukaryote PAP genes described in literature and databases. Indeed, eukaryote PAP genes do not possess homologs in bacterial genomes (Kalscheuer, 2010) and despite the functional relevance of this protein in lipid metabolism of oleaginous bacteria, most studies to establish its role have been so far restrict only to eukaryotes (Comba *et al.*, 2013).

Additionally, none of the glycerol-3-phosphate O-acyltransferase genes (*plsB*) annotated in metagenomic/metatranscriptomic combined database and proteomic dataset are found within the *Microthrix parvicella* BIO17-1 genome (**Figure 4.14**, **Table 4.5**). However, a phospholipid/glycerol acyl-transferase has been identified in the *Microthrix parvicella* BIO17-1 genome (Muller *et al.,* 2012), suggesting that this enzyme is present in *Microthrix parvicella* but none of the *plsB* identified in the metagenomic/metatranscriptomic combined database have been identified back to *Microthrix parvicella* BIO17-1. Moreover, the above-mentioned protein present a high similarity [Score: 386, e-value: 1e-134, Identity (%): 100] with putative acyltransferase identified in *Microthrix parvicella* RN1 (accession number: WP_012224404; Mcllroy *et al.,* 2013).

The presence of other community member genes and proteins involved in the β-oxidation pathway also indicate that the LAO-enriched microbial community is able to degrade/assimilated LCFAs. Living at the air-water interface LAO-enriched microbial communities experience aerobic conditions, thus promoting the β-oxidation

pathway to the detriment of Kennedy pathway. These two pathways are known to operate simultaneously, particularly in γ-Proteobacteria (Röttig & Steinbüchel, 2013). To weight the predominance of β-oxidation *versus* the Kennedy pathway, the relative gene expression level of each individual KOs involved in these two pathways in each sampling date was measured. The genes involved in the Kennedy pathway are 13.6 (st.dev +/- 0.15) times more expressed than the genes for β-oxidation, suggesting that LAO-enriched microbial community members may continually accumulate lipids even under the aerobic conditions experienced at the air-water interface.

4.3.7. Linking gene expression and seasonal representation to metabolic network topology

In order to improve understanding of the interactions of previously highlighted KOs within the context of the whole detected set of genes, a detailed network analyses was carried out. Different topological features, such as network modularity, node degree, betweeness centrality and clustering coefficient were determined in the reconstructed global community-wide metabolic network (Section 4.2.9). The measure of the network modularity, based on splitting the metabolic network into densely connected subgraphs, reveals the identity of KOs and associated metabolites connecting representative metabolic pathways. Metabolites allowing the synthesis of amino acids and nucleotides from precursors, specifically carbohydrates, are highlighted using this measure. The measure of node degree emphasises numerous highly connected KOs ("hubs") comprised within the pyruvate metabolism. In particular, the position of pyruvate kinase-related genes at the intersection of key pathways of energy metabolism is well confirmed by the high node degree (305, average network-wide

node degree 59.8) of these genes in the community-wide metabolic network. Betweenness centrality reflects that any nodes can be reached via a small number of reactions, demonstrating the “small world” priority is represented in the community-wide microbial metabolic network. Thus, most nodes are not neighbours of one another, but most nodes can be reached by a small number of pathways (edges).

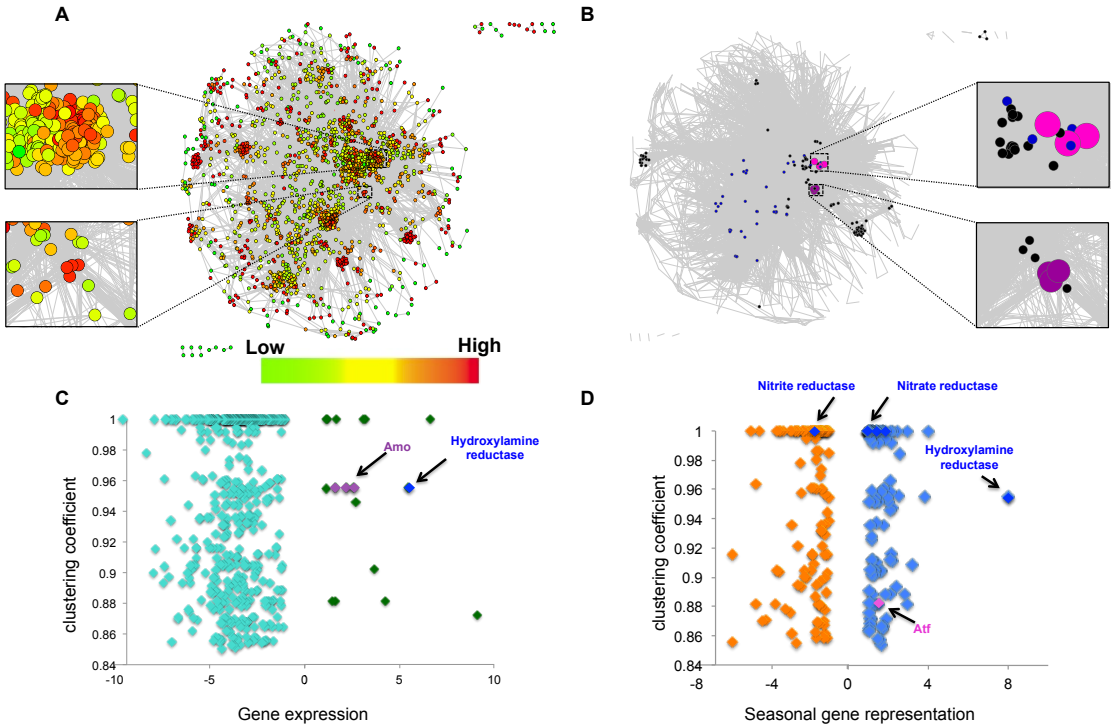

Figure 4.15 Topological analysis of the global community-wide metabolic network with regards to gene expression and seasonal representation. (**A**,**B**) Global community-wide metabolic network with highlighted area corresponding to *plsB*, *plsC* and *atf* genes (top panel) and *amoA, amoB* and *amoC* genes (lower panel). (**A**) Corresponding clustering coefficients of the individual nodes highlighted. (**B**) Specific KOs belonging the glycerolipid (blue nodes) and nitrogen (black nodes) metabolic pathways. Highlighted are ammonia monooxygenase (purple nodes) and TAG biosynthesis genes (pink nodes). (**C**,**D**) Dot plots of KOs exhibiting highest clustering coefficients based on (**C**) their gene expression values (dark green, overexpressed) in the winter LAO community and (**D**) seasonal gene representation (blue, winter overrepresented; orange, autumn overrepresented). Annotation; Amo: ammonia monooxygenase (*amo*), Atf: diacylglycerol O-acyltransferase (*atf*).

The clustering coefficient measures the proportion of existing edges between a node's neighbours. Interestingly, high clustering coefficients (between 0.85 and 1) are associated with some of the overexpressed and season-associated KOs highlighted

previously. Regarding nitrogen metabolism, the majority of the KOs involves within the denitrification and nitrification pathways (**Figure 4.15, A&B**), including the overexpressed KOs coding for the ammonia monooxygenase enzyme (**Figure 4.15, A,B&C**) as well as other seasonal-associated genes KOs, have a clustering coefficient greater than 0.95. Remarkably, the gene coding for the diacylglycerol O-acyltransferase enzyme, known to be a key gene in TAG synthesis (Section 1.1.4.1), has a clustering coefficient of 0.87 (**Figure 4.15, C&D**).

Given the aforementioned patterns of highly clustered genes exhibiting marked differences in relative gene expression, it is proposed to introduce the concept of "keystone genes", in analogy to the keystone species identifiable through pairwise associations in inferred species interaction networks (Faust & Raes, 2012). Reconstructed networks (see **Figure 4.16, A** for an example) allow the highlighting of hubs which correspond to species which are associated with many others. The removal of such species would have a large impact on the community structure (Faust & Raes, 2012). Consequently, a keystone species is a species that has a disproportionately large effect on its environment, relative to its abundance (Paine, 1995). In gut microbiota, for instance, specialist primary degraders such as *Ruminococcus bromii* are considered as a keystone species because of their ability to initiate the degradation of particulate substrates, such as dietary fiber and resistant starch allowing release of energy from recalcitrant substrates (Ze *et al.*, 2012). Furthermore, keystone species typically exhibit high levels of activity compared to other community members despite their comparatively low numbers (de Visser *et al.*, 2013). Similarly, the removal of a "keystone gene" from a community-wide metabolic would have a high impact on the overall community function and structure. In analogy to the keystone species concept, a keystone gene is here defined as a gene

which has an overall high effect on eco-system function, by exhibiting a relatively high gene expression and clustering coefficient. The hypothetical metabolic network (**Figure 4.16, B**) highlights genes having a particular high clustering coefficient compared to other nodes within the metabolic network. The removal of such nodes would result in loss of linkage within a metabolic network and this would greatly impact the overall structure of the community-wide metabolic network.

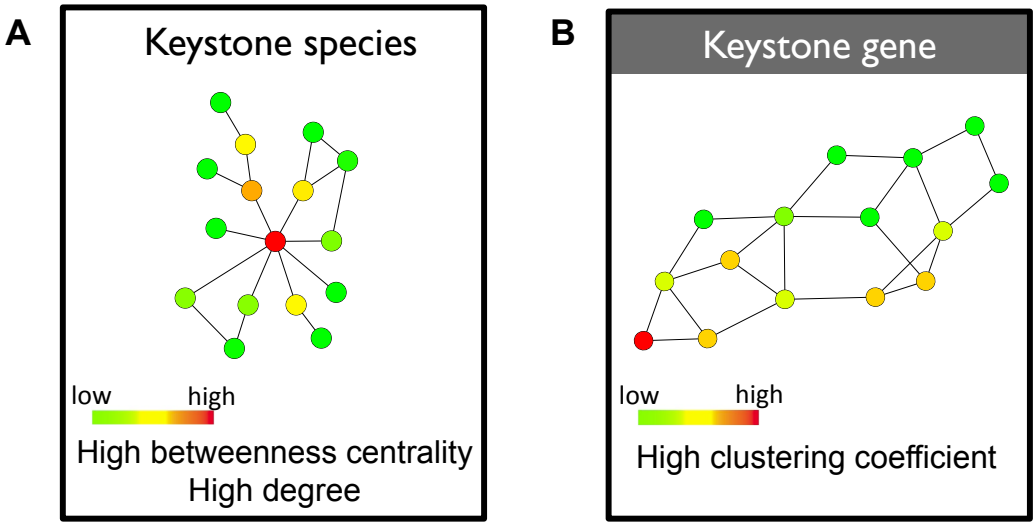

Figure 4.16 Concept of keystone nodes in microbial community-wide networks. (**A**) Keystone species within a network inference where nodes represent taxa and edges represent relationships between them. (**B**) Keystone genes within a community-wide metabolic network. In this network, nodes represent enzyme-coding genes and edges correspond to metabolites.

4.4. Conclusion and perspectives

4.4.1. The fate of fats in LAO communities

The excessive abundance and overexpression of genes related to triacylglycerol lipase confirms the importance of LAOs to export such lipases to degrade TAGs, potentially bound to the cell surface. The lack of a predominant active transport gene or proteins involved in the transport of resulting LCFAs encoded by the *Microthrix parvicella* genome suggests the possible existence of unknown transport protein involved in LCFAs uptake (Hajri & Abumrad, 2002). The prevalence of the activity of long-chain acyl-CoA synthetase genes as well as the similarity of structure between LCFAs found in extra- and intracellular compartments supports the idea that LCFAs are taken up and catabolysed by LAOs. Prominent TAG biosynthesis by LAO-enriched communities (even under the aerobic conditions experienced at the surface of the anoxic treatment tanks) is supported by the presence of the main genes, transcripts and proteins involved in the Kennedy pathway as well as their higher abundance in terms of expression compared to β-oxidation genes.

4.4.2. Keystone genes within community-wide metabolic networks

The identification of potential "keystone genes" in microbial communities was possible through detailed analysis of topological features and relative gene expression differences exhibited by enzyme coding genes within the reconstructed community-wide metabolic networks. This analysis highlighted, primarily, genes involved in

nitrogen and glycerolipid metabolism as occupying such keystone functions in LAO-enriched microbial communities.

CHAPTER 5: GENERAL DISCUSSION

Biological wastewater treatment employing the activated sludge process arguably represents the most widely used biotechnological process in the world (Seviour, 2010). As demonstrated previously (Raunkjær *et al.*, 1994), the lipid fraction present in domestic wastewater is comprised, primarily, of TAGs which, themselves, are composed of LCFAs, particularly stearic, oleic, linoleic and palmitic acids. Overall, lipids can represent up to 31 % of the total biodegradable organic fraction (Raunkjær *et al.,* 1994); thereby, representing a potentially important resource, even more so, because these lipids may be rather easily converted into biofuel. Consequently, lipids in wastewater represent an interesting feedstock for bioenergy production, especially, as they would not encumber the same problems associated with other feedstocks, such as those which compete with food crops and land available to agriculture (Mondala *et al.*, 2009). However, in order for the chemical energy with the lipid fraction of wastewater to be exploitable, a detailed understanding of lipid transformations, especially within lipid accumulating organisms (LAOs), which naturally occur within activated sludge biomass, is necessary. Obtaining unprecedent insights into LAO-enriched microbial communities was the primary objective of the present doctoral research project.

Within anaerobic/aerobic sludge cycling, specialised filamentous LAOs, e.g. *Microthrix parvicella*, are hypothesised to export highly efficient extracellular lipases for TAG hydrolysis, take up resulting LCFAs and store them as neutral lipids primarily during anaerobic treatment phases (Section 1.1.3). Particularly, the uptake and storage capacity of oleic acid by *Microthrix parvicella* under anoxic conditions has previously been shown to be high, relatively stable and more pronounced when compared to other floc-forming microorganisms (Kindaichi *et al.*, 2013; Noutsopoulos *et al.*, 2012). Additionally, while accumulation of TAGs in intracellular

lipid-bodies is a property of only a few prokaryotes (Wältermann & Steinbüchel, 2005), *Microthrix parvicella* appear to have the capacity to store TAGs (Rossetti *et al.*, 2005), as well as LAO communities enriched in this species (Chapter 4). These findings are backed up by genome sequences of isolate *Microthrix parvicella* strains (McIlroy *et al.*, 2013; Muller *et al.,* 2012) which were found to carry all the genes necessary for TAG lipid storage. Although the isolated genomes, as well information from metagenomics (McIlroy *et al.*, 2013), have provided ample evidence for a pronounced capacity of *Microthrix* parvicella to accumulate lipids, the exact pathways involved in lipid uptake and storage within LAO communities *in situ* (which is relevant for devising biological lipid-reclamation strategies from wastewater) can only be fully elucidated using integration of different analytical techniques, as carried out in the present work (Chapter 4).

The application of microscopic and molecular analytical tools to the study of LAO-enriched communities from the biological WWTP located at Schifflange (Luxembourg), resulted in preliminary physiological and phylogenetic insights into the studied communities (Chapter 2). These initial insights resulted in the selection of LAO-enriched communities found at the air-water interface of the anoxic treatment tank of this WWTP to carry out the following methodological developments (Chapter 3) and analyses of the present study. Following biomolecular isolation and high-throughput analyses, the integration of resulting multi-omic data enabled the refinement of a metabolic model for lipid accumulation within the studied LAO-enriched communities (Chapter 4).

Proteobacteria (primarily members of the α, β and γ subgroups) are the most abundant phyla in all LAO-enriched microbial community samples analysed as part of the

present doctoral research project. This is similar to the results of microbial communities in soil (Roesch *et al.*, 2007), sewage (McLellan *et al.*, 2010), the deep seawater (Qian *et al.*, 2010) and activated sludge (Zhang *et al.*, 2011). The bacterial diversity of LAO-enriched community is distinct from the bacterial diversity found in activated sludge (Chapter 2). At the phylum level, LAO communities exhibit high levels of *Actinobacteria* (Chapter 2). The dominance of *Actinobacteria* is mainly due to the prominence of *Microthrix parvicella* (Chapter 2). The dominance of such filamentous microorganisms, characteristic of foaming and bulking issues encountered at the Schifflange WWTP as well as equivalent WWTPs worldwide is, therefore, of pronounced importance to elucidate details of lipid accumulation in LAO-enriched communities.

With relatively short generation times and rapid growth under favourable conditions, microbial communities are among the fastest components of an ecosystem to respond to changing environmental conditions (Prosser *et al.*, 2007). Seasonal variation in bacterial community structure has been, widely, reported in natural habitats such forest soil (Kuffner *et al.*, 2012), lakes (Newton *et al.*, 2011), the Baltic Sea (Andersson *et al.*, 2009), air (Bowers, 2012) and marine coastal regions (Gilbert *et al.*, 2011). In this latter study, the authors suggest that seasonal changes in environmental variables have more impact than trophic interactions on the structure of the community. A seasonal microbial community shift has been reported in a saline sewage treatment plant in Hong Kong (Yan *et al.*, 2011) and in that study, 12 monthly samples fell into four different clusters in accordance with the four seasons from which samples were obtained. As demonstrated in the present work, in LAO-enriched microbial communities a seasonal shift is associated with changes in abundance of three abundant organisms namely *Alkanindiges* spp., Acinetobacter *spp.* and

Microthrix parvicella (Chapter 2). As demonstrated (Chapter 2) the presence of *Microthrix parviella* is strongly correlated with season and associated temperature variations. *Microthrix parvicella* is a low temperature tolerant bacterium and considerable growth has been observed at 7 °C in full-scale plants (Rossetti *et al.*, 2005). The dominance of *Microthrix parvicella* in winter is likely due to his advantage for growth in cold seasons compared to its other potential competitors, i.e. *Alkanindiges* spp. and *Acinetobacter* spp. (Chapter 2).

Despite high abundances in the known LAO *Microthrix parvicella* in sampled winter and autumn communities (Chapter 2), clear differences are apparent in the relative amounts of intracellular triacylglycerides and corresponding long chain fatty acids in the representative samples for both seasons (Section 4.3.1). These differences likely correspond to the distinct microbial community compositions observed for both seasons (Section 2.3.2.1). Furthermore, apart from an identified association between orthologous gene sequences for the gene diacylglycerol O-acyltransferase and the sampled winter LAO-enriched microbial community, other orthologous genes involved in lipid metabolism did not exhibit similar seasonal associations. These results suggest that *Microthix parvicella* is not the only lipid accumulating population within the community and that other members are also able to assimilate LCFAs and accumulate TAGs. Particularly, the dominant *Alkanindiges* spp. may also be an LAO since members of the Moraxellaceae family, i.e. *Acinetobacter* spp., are known to accumulate TAGs and WEs. Thus, the dominant *Microthrix* spp. and *Alkanindiges* spp. organisms with the LAO communities of the anoxic tank of the Schifflange biological wastewater treatment plant might have similar gene inventories, but may differ by some important genes, e.g. wax ester synthase/acyl coenzyme A (acyl-CoA):diacyglycerol acyltransferase (WS/DGAT) found in *Acinetobacter* spp.

(Stöveken *et al.*, 2005). However, detailed future investigations are required to elucidate if the dominant *Alkanindiges* spp. and *Acinetobacter* spp. are indeed LAOs.

The results of the present doctoral research support previous findings and suggest that TAGs in wastewater are processed by lipid accumulating organisms, by harnessing their aptitude of hydrolysing TAGs to LCFAs and efficiently assimilating and activating these before being reconverted to TAGs intracellularly (Section 4.3.6). More specifically, TAGs present in wastewater are hydrolysed to LCFAs and glycerol by triacylglycerol lipases (Section 4.3.6). Based on the data generated as part of this doctoral research project, the resulting glycerol is likely converted into glycerol-3-phosphate and then transported into LAOs (Section 4.3.6). Resulting LCFAs are likely first associated to the cell surface of LAOs by their hydrophobicity before being assimilated. The precise mechanism for LCFAs uptake by LAOs is presently unknown (Mcllroy *et al.,* 2013), but upon entry into the cell, they are likely activated to LCFA-CoA by diverse set of long chain fatty acid CoA ligases. The *Microthrix parvicella* BIO17-1 genome alone encodes 28 homologs of this gene (Muller *et al.,* 2012) of which 10 were detected in our integrated omic analyses (section 4.3.6), which in turn suggests that it uses specialised enzymes for the different LCFAs it accumulates. At the surface of the anoxic phase, LAO-enriched microbial communities experience aerobic conditions and a proportion of the newly synthesised fatty acyl-CoA is most likely catabolised through β-oxidation pathway. However, based on a comparison of overall gene expression levels derived from the multi-omic data (Chapter 4), the majority of the fatty acyl-CoAs are likely used to (re-)synthetise TAGs through the Kennedy pathway using the assimilated glycerol-3-phosphate. The main species responsible for TAG cycling from wastewater within the studied LAO-enriched communities based on the present work are likely *Microthrix parvicella,*

Alkanindiges spp. and *Acinetobacter* spp. (Chapter 2&4). However, future work is clearly needed to elucidate their relative roles in conferring the community-wide TAG accumulation phenotype.

In 1987, in his essay on the state of microbial ecology, Thomas Brock argued that measures of diversity were pointless, because the dynamics of the microbial world mean that communities do not have a characteristic diversity, but change as the environment changes (Curtis & Sloan, 2004). A few years later, the biologist EO Wilson reinforced the futility of the study of microbial diversity, notwithstanding its function, when he observed that microbial diversity is 'beyond practical calculation' (Wilson, 1999). Collectively, these two points raise two questions; how can one measure microbial diversity and what does microbial diversity mean in the context of community function (Curtis *et al.*, 2004)?

Two decades on from Brock and Wilson, microbial ecology has undergone a dramatic revolution, which is driven by the advent of high-throughput molecular tools (Section 1.2.2). These are now allowing deep measurements of microbial diversity (as carried out in the present work through deep 16S rRNA gene sequencing, Chapter 2) the study of structure-function relationships within communities and their environment (as carried out in the present work through a comparative integrated omic study, Chapter 4). Particularly, the study of the dynamic processes and interactions, between microorganisms within ecosystems, can now be traced at the genomic, transcriptomic, proteomic and metabolomic levels. When linked to given environmental variables, these approaches allow linkages to be established between population-level genetic information and overall ecological function *in situ* (Morales & Holben, 2011). A combined genomic and metagenomic approach has recently been applied to facilitate

the reconstruction of a metabolic model for *Microthrix parvicella* (McIlroy *et al.,* 2013). This study, along with the previous genome sequence (Muller *et al.,* 2012), provides the foundation for subsequent metagenomic, metatranscriptomic, metaproteomic and metabolomic studies in order to clarify the details of the lipids transformations in LAOs as carried out in the present work.

Besides the identification of the main actors and processes relating to the studied LAO-enriched communities, a major achievement of this PhD project has been the setting up of appropriate wet- (Chapter 3) and dry- (Chapter 4) lab methodologies, allowing unprecedented insights into microbial community structure and function within the framework of an eco-systems biology approach. The developed biomolecular isolation framework (Chapter 3) lays the foundation for standardised, reproducible and simultaneous high-resolution measurements of multiple features from single unique microbial community samples which in turn, as demonstrated in Chapter 4, allows meaningful resolution of community- and population-level roles, spanning genetic potential to final phenotype. As carried out for the present doctoral research project (Chapter 4), integration of systematic multi-omic measurements within a community-level metabolic network establishes a unique framework for comparative Eco-Systems Biology. The resulting networks and models allow description of the function of entire microbial communities by simultaneously resolving, genetic potential, gene expression, protein frequency, as well as the associated systems-level network topological features, in order to understand the dynamics and evolution of microbial ecosystems and predict possible control strategies through the identification of for example keystone genes.

The developed methodology goes far beyond its application to the microbial ecology field and has the potential to be applied universally in different domains of Systems Biology. In *Foundation of Systems Biology* (Kitano, 2001), Kitano defined the ideal systematic measurement as the simultaneous measurement of multiple features of a single sample. More than ten years later, such systematic molecular measurements are now possible, thanks to the developed methodologies. The application of such systematic measurments to temporally resolved samples, e.g. LAO-enriched microbial communities, along with appropriate bioinformatics and statistical analyses, will allow the discrovery of previously unknown traits of specific microbial community members in accordance with a discovery-driven planning approach (Muller *et al.,* 2013). In the context of the present doctoral research project, such an approach may prove helpful to provide pointers to the function of the highly abundant protein of unknow function identified within the LAO-enriched microbial community, encoded by *Microthrix parvicella* (Chapter 2&3).

Integrated omics of 51 individual LAO-enriched community samples taken weekly from the Schifflange anoxic tank over an entire year are currently being carried out. Integrated analysis of the generated genomic, transcriptomic, proteomic and metabolomic temporally resolved datasets will provide additional datasets to further confirm the findings generated in the current project and will allow an unprecedented dynamic analysis of the LAO-enriched microbial communities at a multi-omics level. Importantly, through the generation of dynamic systematic measurements, causal relationships between community-wide phenotypes, specific taxa and genes will be resolvable.

The overall work presented in the present doctoral thesis ranged from the initial characterisation of lipid accumulating microbial communities in a particular WWTP in Luxembourg (Chapter 2) to the establishment of proper systematic molecular measurement techniques (Chapter 3) and to the identification of specific keystone genes in community-wide metabolic networks by integrated omics (Chapter 4), thereby allowing significant insight into the major actors and processes in lipid processing and storage in the studied LAO-enriched microbial communities. However, in an overall Eco-Systems Biology approach this only represents the first step resulting in hypotheses which will need to be tested through *in vitro* and field (*in situ*) experiments in order to ultimately devise and validate concrete microbial community-wide control strategies (**Figure 5.1**).

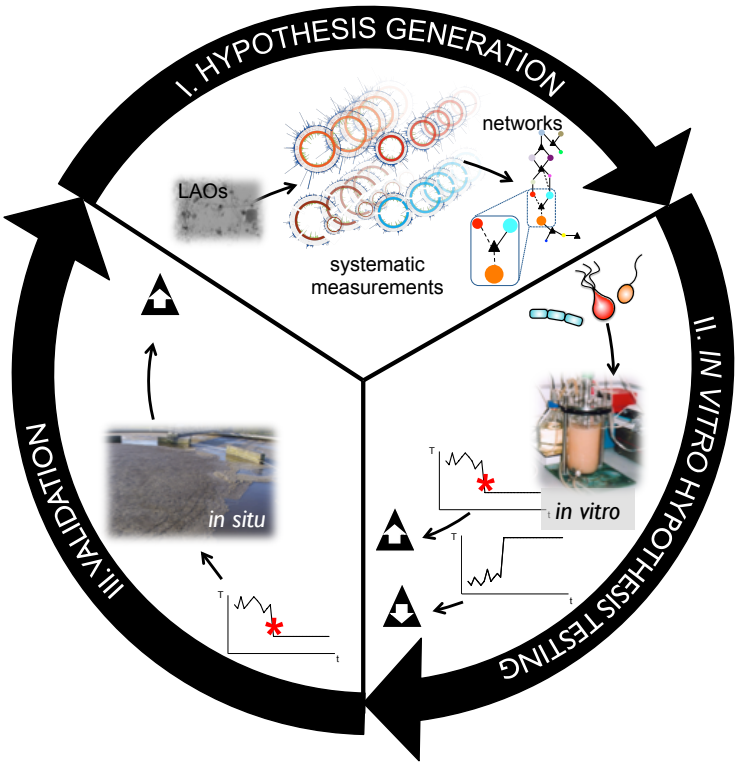

Figure 5.1 Iterative experimental workflow for Eco-Systems Biology as highlighted by the study of LAO-enriched microbial communities. I. Hypothesis generation: Community-wide network reconstructions established based on systematically generated integrated omic data. II. *In vitro* hypothesis testing: Bioreactor-based experiments allow the confirmation or the rejection of hypotheses derived from the high-resolution molecular data. III. Validation: To causally link for example changes in microbial community structure resulting from changes in physico-chemical parameters (as illustrated in this example by a specific water temperature) with associated biomolecular patterns (associated with a pronounced community-wide lipid accumulation) *in situ* validations will be indispensable. *denotes a setpoint in water temperature, the triangle represents TAGs, charts represent the variation of the temperature (T) with the time (t). Adapted from Fritz *et al.*, 2013.

Consequently, in order to test the findings generated in the current project and future datasets, future research plans need to include *in vitro* experiments involving the main LAO members from different pre-selected sludge samples from the Schifflange WWTP to test the system-wide beviours resolvable using high-resolution molecular tools *in situ*. For such experiments, I would for example propose to run three different bioreactors, one specifically enriched with *Alkanindiges* spp. and *Acinetobacter* spp. using a sludge sample from autumn, one for *Microthrix parvicella* using a sludge sample from winter and a last one comprising a co-culture of both using a sludge sampled at the transition point between summer and winter. These setups, along with additional temporally-resolved high-resolution biomolecular characterisations, will allow the specific characterisation of the lipid accumulation of different LAO members and highlight possible co-occurrence, co-exclusion or niche specialisations of the main LAOs members. The sludge sample used for the inoculation of each individual bioreactor should be pre-selected using the results obtained in the current project as well as the detailed time series experiment. Such bioreactor experiments could also be used to validate the physico-chemical parameters, e.g. water temperature, allowing maximisation of lipid uptake by LAO-enriched microbial communities. However, in order to meaningfully design such experiments, at least the detailed time course data will be indispensable. The generated knowledge may eventually be used to design new sustainable biological wastewater treatment processes based on excess LAO biomass production and subsequent conversion of this to biofuel.

In this PhD thesis project, it is highlighted that LAO-enriched microbial communities can be outstanding model systems for microbial ecology research. The communities are easily accessible and temporally-resolved samples can be easily obtained along

with extensive metadata. Most importantly, specific bacterial groups dominate the communities with a relatively low evenness (i.e. certain species dominate the overall community) and the system exhibits a clear seasonal ecological succession. Microbiologists and engineers have long used theory and ideas from ecosystem science and ecology to inform their study and design of activated sludge systems, though some would argue that the results could be of a higher standard (McMahon *et al.*, 2007; Prosser *et al.*, 2007). Researchers can now borrow heavily from the field of systems biology, in order to create a more holistic understanding of such microbial communities that is well grounded in knowledge about eco-systems-wide behaviour at the molecular level. Such detailed knowledge will allow models to be built for accurately predicting for example changes in microbial community structure and resulting alterations of metabolic transformations relevant to biological wastewater treatment processes based on physico-chemical parameters. Furthermore, such models may allow formulation of control strategies for microbial communities. Control may be for example achievable through alteration of the relative levels of specific keystone genes/enzymes within the communities as those described in the present work.

REFERENCES

Al-Halbouni D, Dott W, Hollender J. (2009). Occurrence and composition of extracellular lipids and polysaccharides in a full-scale membrane bioreactor. Water Research **43:** 97-106.

Alvarez H, Steinbüchel A. (2002). Triacylglycerols in prokaryotic microorganisms. Applied Microbiology and Biotechnology **60:** 367-376.

Alvarez HM, Mayer F, Fabritius D, Steinbüchel A. (1996). Formation of intracytoplasmic lipid inclusions by *Rhodococcus opacus* strain PD630. Arch Microbiol **165:** 377-386.

Alvarez HM, Kalscheuer R, Steinbüchel A. (1997). Accumulation of storage lipids in species of *Rhodococcus* and *Nocardia* and effect of inhibitors and polyethylene glycol. Lipid / Fett **99:** 239-246.

Alvarez HM, Silva RA, Herrero M, Hernández MA, Villalba MS. (2013). Metabolism of triacylglycerols in *Rhodococcus* species: insights from physiology and molecular genetics. Journal of Molecular Biochemistry **2**.

Amann R, Ludwig W. (2000). Ribosomal RNA-targeted nucleic acid probes for studies in microbial ecology. FEMS Microbiology Reviews **24:** 555-565.

Andersson AF, Riemann L, Bertilsson S. (2009). Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. The ISME Journal **4:** 171-181.

Andreasen K, Nielsen PH. (1998). *In situ* characterization of substrate uptake by *Microthrix parvicella* using microautoradiography. Water Science & Technology **37:** 19-26.

Arcondéguy T, Jack R, Merrick M. (2001). PII signal transduction proteins, pivotal players in microbial nitrogen control. Microbiology and Molecular Biology Reviews **65:** 80-105.

Armengaud J. (2013). Microbiology and proteomics, getting the best of both worlds! Environmental Microbiology **15:** 12-23.

Azam F, Worden AZ. (2004). Microbes, Molecules, and Marine Ecosystems. Science **303:** 1622-1624.

Azócar L, Ciudad G, Heipieper H, Navia R. (2010). Biotechnological processes for biodiesel production using alternative oils. Applied Microbiology and Biotechnology **88:** 621-636.

Barksdale L, Kim K-S. (1977). *Mycobacterium*. Bacteriological reviews **41:** 217.

Barnard JL. (1975). Biological nutrient removal without the addition of chemicals. Water Research **9:** 485-490.

Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ. (2007). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat Protoc **2:** 727-738.

Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP *et al*. (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science **289:** 1902 - 1906.

Benndorf D, Balcke GU, Harms H, von Bergen M. (2007). Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. The ISME Journal **1:** 224-234.

Björnsson L, Hugenholtz P, Tyson GW, Blackall LL. (2002). Filamentous *Chloroflexi* (green non-sulfur bacteria) are abundant in wastewater treatment processes with biological nutrient removal. Microbiology **148:** 2309-2318.

Blackall L, Crocetti G, Saunders A, Bond P. (2002). A review and update of the microbiology of enhanced biological phosphorus removal in wastewater treatment plants. Antonie van Leeuwenhoek **81:** 681-691.

Blackall LL, Seviour EM, Cunningham MA, Seviour RJ, Hugenholtz P. (1995). "*Microthrix parvicella*" is a Novel, Deep Branching Member of the Actinomycetes Subphylum. Systematic and Applied Microbiology **17:** 513-518.

Blackall LL, Stratton H, Bradford D, Del Dot T, Sjorup C, Seviour EM *et al*. (1996). "*Candidatus* Microthrix parvicella," a Filamentous Bacterium from Activated Sludge Sewage Treatment Plants. International Journal of Systematic Bacteriology **46:** 344-346.

Blauwkamp TA, Ninfa AJ. (2003). Antagonism of PII signalling by the AmtB protein of Escherichia coli. Molecular Microbiology **48:** 1017-1028.

Bogan BW, Sullivan WR, Kayser KJ, Derr K, Aldrich HC, Paterek JR. (2003). *Alkanindiges illinoisensis* gen. nov., sp. nov., an obligately hydrocarbonoclastic, aerobic squalane-degrading bacterium isolated from oilfield soils. International journal of systematic and evolutionary microbiology **53:** 1389-1395.

Bond PL, Erhart R, Wagner M, Keller J, Blackall LL. (1999). Identification of some of the major groups of bacteria in efficient and nonefficient biological phosphorus removal activated sludge systems. Applied and Environmental Microbiology **65:** 4077-4084.

Bordenave S, Goni-Urriza MS, Caumette R, Duran R (2007). Effects of heavy fuel oil on the bacterial community structure of a pristine microbial mat. Applied and Environmental Microbiology **19**: 6089-6097.

Borenstein E. (2012). Computational systems biology and *in silico* modeling of the human microbiome. Briefings in Bioinformatics **13:** 769-780.

Bowers RM. (2012). The spatiotemporal dynamics of microbes in the near-surface atmosphere. ProQuest Dissertations And Theses; Thesis (Ph.D.)--University of Colorado at Boulder, 2012.

Bradford D, Hugenholtz P, Seviour EM, Cunningham MA, Stratton H, Seviour RJ *et al*. (1996). 16S rRNA analysis of isolates obtained from Gram-negative, filamentous bacteria micromanipulated from activated sludge. Systematic and Applied Microbiology **19:** 334-343.

Brdjanovic D, Slamet A, Van Loosdrecht MCM, Hooijmans CM, Alaerts GJ, Heijnen JJ. (1998). Impact of excessive aeration on biological phosphorus removal from wastewater. Water Research **32:** 200-208.

Bulow SE, Francis CA, Jackson GA, Ward BB. (2008). Sediment denitrifier community composition and *nirS* gene expression investigated with functional gene microarrays. Environmental Microbiology **10:** 3057-3069.

Bundy JG, Davey MP, Viant MR. (2009). Environmental metabolomics: a critical review and future perspectives. Metabolomics **5:** 3-21.

Bura R, Cheung M, Liao B, Finlayson J, Lee BC, Droppo IG *et al*. (1998). Composition of extracellular polymeric substances in the activated sludge floc matrix. Water Science and Technology **37:** 325-333.

Caporaso J, Kuczynski J, Stombaugh J, Bittinger K, Bushman F, Costello E *et al*. (2010a). QIIME allows analysis of high-throughput community sequencing data. Nat Methods **7:** 335 - 336.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ *et al*. (2010b). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences **15**:108-115.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N *et al*. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal **6:** 1621-1624.

Carr E, Kämpfer P, Patel BK, Gürtler V, Seviour RJ. (2003). Seven novel species of *Acinetobacter* isolated from activated sludge. International journal of systematic and evolutionary microbiology .**4**: 953-963.

Chaurand P, Luetzenkirchen F, Spengler B. (1999). Peptide and protein identification by matrix-assisted laser desorption ionization (MALDI) and MALDI-post-source decay time-of-flight mass spectrometry. Journal of the American Society for Mass Spectrometry **10:** 91-103.

Chey S, Claus C, Liebert UG. (2011). Improved method for simultaneous isolation of proteins and nucleic acids. Analytical Biochemistry **411:** 164-166.

Chindelevitch L, Stanley S, Hung D, Regev A, Berger B. (2012). MetaMerge: scaling up genome-scale metabolic reconstructions with application to Mycobacterium tuberculosis. Genome Biol **13:** r6.

Chipasa KB, Mędrzycka K. (2006). Behavior of lipids in biological wastewater treatment processes. Journal of Industrial Microbiology and Biotechnology **33:** 635-645.

Cho I, Blaser MJ. (2012). The human microbiome: at the interface of health and disease. Nature Reviews Genetics **13:** 260-270.

Chomczynski P. (1993). A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. Biotechniques **15:** 532-536.

Clarke K, Ainsworth M. (1993). A method of linking multivariate community structure to environmental variables. Marine Ecology-Progress Series **92:** 205-205.

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

Comba S, Menendez-Bravo S, Arabolaza A, Gramajo H. (2013). Identification and physiological characterization of phosphatidic acid phosphatase enzymes involved in triacylglycerol biosynthesis in *Streptomyces coelicolor*. Microbial cell factories **12:** 9.

Conley DJ, Paerl HW, Howarth RW, Boesch DF, Seitzinger SP, Havens KE *et al*. (2009). Controlling eutrophication: nitrogen and phosphorus. Science **323:** 1014-1015.

Cottret L, Milreu PV, Acuña V, Marchetti-Spaccamela A, Stougie L, Charles H *et al*. (2010). Graph-based analysis of the metabolic exchanges between two co-resident intracellular symbionts, *Baumannia cicadellinicola* and *Sulcia muelleri*, with their insect host, *Homalodisca coagulata*. PLoS computational biology **6:** e1000904.

Craig R, Cortens JP, Beavis RC. (2004). Open source system for analyzing, validating, and storing protein identification data. Journal of Proteome Research **3:** 1234-1242.

Curtis TP, Sloan WT. (2004). Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. Current Opinion in Microbiology **7:** 221-226.

Daims H, Nielsen JL, Nielsen PH, Schleifer K-H, Wagner M. (2001). *In situ* characterization of nitrospira-like nitrite-oxidizing bacteria active in wastewater treatment plants. Applied and Environmental Microbiology **67:** 5273-5284.

Daims H, Lücker S, Wagner M. (2006a). daime, a novel image analysis program for microbial ecology and biofilm research. Environmental Microbiology **8:** 200-213.

Daims H, Taylor MW, Wagner M. (2006b). Wastewater treatment: a model system for microbial ecology. Trends in Biotechnology **24:** 483-489.

de Visser S, Thébault E, de Ruiter PC (2013). Ecosystem engineers, keystone species. *Ecological Systems*. Springer. pp 59-68.

Dean AP, Sigee DC, Estrada B, Pittman JK. (2010). Using FTIR spectroscopy for rapid determination of lipid accumulation in response to nitrogen limitation in freshwater microalgae. Bioresource Technology **101:** 4499-4507.

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N-U *et al*. (2006). Community genomics among stratified microbial assemblages in the Ocean's interior. Science **311:** 496 - 503.

Denef VJ, Mueller RS, Banfield JF. (2010). AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. ISME J **4:** 599-610.

Dennis P, Edwards EA, Liss SN, Fulthorpe R. (2003). Monitoring gene expression in mixed microbial communities by using DNA microarrays. Applied and Environmental Microbiology **69:** 769-778.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al*. (2006). Greengenes, a chimera-checked 16s rRNA gene database and workbench compatible with arb. Appl Environ Microbiol **72:** 5069-5072.

Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H, Tasman N *et al*. (2010). A guided tour of the Trans-Proteomic Pipeline. Proteomics **10:** 1150-1159.

Drinan JE, Spellman FR. (2012). Water and wastewater treatment: A guide for the nonengineering professional. Crc Press.

Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD *et al*. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc Natl Acad Sci USA **104:** 1777-1782.

Dueholm TE, Andreasen KH, Nielsen PH. (2001). Transformation of lipids in activated sludge. Water Science & Technology **43:** 165-172.

Dufreche S, Hernandez R, French T, Sparks D, Zappi M, Alley E. (2007). Extraction of lipids from municipal wastewater plant microorganisms for production of biodiesel. Journal of the American Oil Chemists' Society **84:** 181-187.

Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N *et al*. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nature protocols **6:** 1060-1083.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27:** 2194-2200.

Erhart R, Bradford D, Seviour RJ, Amann R, Blackall LL. (1997). Development and use of fluorescent in situ hybridization probes for the detection and identification of "*Microthrix parvicella*" in activated sludge. Systematic and Applied Microbiology **20:** 310-318.

Excoffier L, Smouse PE, Quattro JM. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131:** 479-491.

Fahy E, Sud M, Cotter D, Subramaniam S. (2007). LIPID MAPS online tools for lipid research. Nucl Acids Res **35:** W606-612.

Faust K, Croes D, van Helden J. (2009). Metabolic pathfinding using RPAIR annotation. Journal of Molecular Biology **388:** 390-414.

Faust K, Raes J. (2012). Microbial interactions: from networks to models. Nature Reviews Microbiology **10:** 538-550.

Feofilova EP, Sergeeva YE, Ivashechkin AA. (2010). Biodiesel-fuel: Content, production, producers, contemporary biotechnology (Review). Applied Biochemistry and Microbiology **46:** 369-378.

Fiehn O. (2002). Metabolomics–the link between genotypes and phenotypes. Plant molecular biology **48:** 155-171.

Field J, Cervantes F, Van der Zee F, Lettinga G. (2000). Role of quinones in the biodegradation of priority pollutants: a review. Water Science and Technology**:** 215-222.

Fischer CR, Wilmes P, Bowen BP, Northen TR, Banfield JF. (2011). Deuterium-exchange metabolomics reveals N-methyl lyso phosphatidylethanolamines as abundant lipids in acidophilic mixed microbial communities. Metabolomics.

Fitzsimons NA, Akkermans ADL, de Vos WM, Vaughan EE. (2003). Bacterial gene expression detected in human faeces by reverse transcription-PCR. Journal of Microbiological Methods **55:** 133-140.

Fjerbaek L, Christensen KV, Norddahl B. (2009). A review of the current state of biodiesel production using enzymatic transesterification. Biotechnology and Bioengineering **102:** 1298-1315.

Fleige S, Pfaffl MW. (2006). RNA integrity and the effect on the real-time qRT-PCR performance. Molecular Aspects of Medicine **27:** 126-139.

Florens L, Carozza MJ, Swanson SK, Fournier M, Coleman MK, Workman JL *et al*. (2006). Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. Methods **40:** 303-311.

Freilich S, Kreimer A, Meilijson I, Gophna U, Sharan R, Ruppin E. (2010). The large-scale organization of the bacterial network of ecological co-occurrence interactions. Nucleic Acids Research **38:** 3857-3868.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al*. (2008). Microbial community gene expression in ocean surface waters. Proceedings of the National Academy of Sciences **105:** 3805-3810.

Fritz JV, Desai MS, Shah P, Schneider JG, Wilmes P. (2013). From meta-omics to causality: experimental models for human microbiome research. Microbiome **1:** 14.

Fujita Y, Matsuoka H, Hirooka K. (2007). Regulation of fatty acid metabolism in bacteria. Molecular Microbiology **66:** 829-839.

García-Martín H, Ivanova N, Kunin V, Warnecke F, Barry KW, McHardy AC *et al*. (2006). Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nature Biotechnology **24:** 1263-1269.

Garcia DE, Baidoo EE, Benke PI, Pingitore F, Tang YJ, Villa S *et al*. (2008). Separation and mass spectrometry in microbial metabolomics. Current Opinion in Microbiology **11:** 233-239.

Gagiulo CE, Stuhlsatz-Krouper SM, Schaffer JE (1999). Localization of adipocyte long-chain fatty acyl-CoA synthetase at the plasma membrane. Journal of Lipid Research **40:** 881-892.

Geesey G, Kloeke FvO, Droppo I, Leppard G, Liss S, Milligan T. (2004). Extracellular enzymes associated with microbial flocs from activated sludge of wastewater treatment systems. Flocculation in natural and engineered environmental systems**:** 295-315.

Gidez LI. (1984). The lore of lipids. Journal of Lipid Research **25:** 1430-1436.

Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P *et al*. (2008). Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoS ONE **3:** e3042.

Gilbert JA, Steele JA, Caporaso JG, Steinbrück L, Reeder J, Temperton B *et al*. (2011). Defining seasonal marine microbial community dynamics. The ISME Journal **6:** 298-308.

Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ, Samuel BS *et al*. (2006). Metagenomic analysis of the human distal gut microbiome. Science **312:** 1355 - 1359.

Gosalbes M, Durban A, Pignatelli M, Abellan J, Jimenez-Hernandez N, Perez-Cobas A *et al*. (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. PLoS ONE **6:** e17447.

Götz A, Goebel W. (2010). Glucose and glucose-6-phosphat as carbon sources in extra- and intracellular growth of enteroinvasive *Eschirichia coli* and *Salmonella enterica*. Microbiology **156:** 1176-1187.

Greenblum S, Turnbaugh PJ, Borenstein E. (2012). Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. Proceedings of the National Academy of Sciences **109:** 594-599.

Greenblum S, Chiu H-C, Levy R, Carr R, Borenstein E. (2013). Towards a predictive systems-level model of the human microbiome: progress, challenges, and opportunities. Current Opinion in Biotechnology **24**: 810-820.

Hajri T, Abumrad NA. (2002) Fatty acid transport across membranes: relevance to nutrition and metabolic pathology. Annual reviw of nutrition **22:** 383-415.

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chemistry & Biology **5:** R245-R249.

Handelsman J. (2004). Metagenomics: application of genomics to uncultured microorganisms. Microbiology and Molecular Biology Reviews **68:** 669-685.

Hanson KG, Anuranjini N, Madhavi K, Anjana JD. (1997) News & Notes: Bioremediation of Crude Oil Contamination with Acinetobacter sp. A3. Current microbiology **3**: 191-193.

Hartmann M, Howes CG, Abarenkov K, Mohn WW, Nilsson RH. (2010). V-Xtractor: an open-source, high-throughput software tool to identify and extract hypervariable regions of small subunit (16S/18S) ribosomal RNA gene sequences. Journal of Microbiological Methods **83:** 250-253.

He S, Wurtzel O, Singh K, Froula JL, Yilmaz S, Tringe SG *et al*. (2010). Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. Nat Meth **7:** 807-812.

Head I, Saunders J, Pickup R. (1998). Microbial evolution, diversity, and ecology: a decade of ribosomal RNA analysis of uncultivated microorganisms. Microbial Ecology **35:** 1-21.

Helbling DE, Ackermann M, Fenner K, Kohler H-PE, Johnson DR. (2012). The activity level of a microbial community function can be predicted from its metatranscriptome. ISME J **6:** 902-904.

Herlemann DPR, Labrenz M, Jurgens K, Bertilsson S, Waniek JJ, Andersson AF. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J **5:** 1571-9.

Hesselsoe M, Nielsen JL, Roslev P, Nielsen PH. (2005). Isotope labeling and microautoradiography of active heterotrophic bacteria on the basis of assimilation of $^{14}CO_2$. Appl Environ Microbiol **71:** 646-655.

Hettich RL, Pan C, Chourey K, Giannone RJ. (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:**4203-4214.

Hiller K, Hangebrauk J, JaÃàger C, Spura J, Schreiber K, Schomburg D. (2009). MetaboliteDetector: Comprehensive Analysis Tool for Targeted and Nontargeted GC/MS Based Metabolome Analysis. Analytical Chemistry **81:** 3429-3439.

Huang Y, Niu B, Gao Y, Fu L, Li W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics **26:** 680-682.

Hug T, Gujer W, Siegrist H. (2006). Modelling seasonal dynamics of "*Microthrix parvicella*". Water science and technology: a journal of the International Association on Water Pollution Research **54:** 189.

Hugenholtz P. (2002). Exploring prokaryotic diversity in the genomic era. Genome Biology **3:** reviews0003.0001 - reviews0003.0008.

Hummon AB, Lim SR, Difilippantonio MJ, Ried T. (2007). Isolation and solubilization of proteins after TRIZOL extraction of RNA and DNA from patient material following prolonged storage. Biotechniques **42:** 467-472.

Hyatt D, Chen G-L, LoCascio P, Land M, Larimer F, Hauser L. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:** 119.

Jahn CE, Charkowski AO, Willis DK. (2008). Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. Journal of Microbiological Methods **75:** 318-324.

Jardé E, Mansuy L, Faure P. (2005). Organic markers in the lipidic fraction of sewage sludges. Water Research **39:** 1215-1232.

Jones PA, Schuler AJ. (2010). Seasonal variability of biomass density and activated sludge settleability in full-scale wastewater treatment systems. Chemical Engineering Journal **164:** 16-22.

Joyce AR, Palsson BO. (2006). The model organism as a system: integrating 'omics' data sets. Nat Rev Mol Cell Biol **7:** 198-210.

Junier P, Molina V, Dorador C, Hadas O, Kim O-S, Junier T *et al*. (2010). Phylogenetic and functional marker genes to study ammonia-oxidizing microorganisms (AOM) in the environment. Applied Microbiology and Biotechnology **85:** 425-440.

Juretschko S, Timmermann G, Schmid M, Schleifer K-H, Pommerening-Röser A, Koops H-P *et al*. (1998). Combined molecular and conventional analyses of nitrifying bacterium diversity in activated sludge: *Nitrosococcus mobilis* and *Nitrospira*-like bacteria as dominant populations. Applied and Environmental Microbiology **64:** 3042-3051.

Kalscheuer R, Steinbüchel A. (2003). A Novel bifunctional wax ester synthase/acyl-coa:diacylglycerol acyltransferase mediates wax ester and triacylglycerol biosynthesis in *Acinetobacter calcoaceticus* ADP1. Journal of Biological Chemistry **278:** 8075-8082.

Kalscheuer R, Stöveken T, Malkus U, Reichelt R, Golyshin PN, Sabirova JS *et al*. (2007). Analysis of storage lipid accumulation in *Alcanivorax borkumensis*: evidence for alternative triacylglycerol biosynthesis routes in bacteria. Journal of Bacteriology **189:** 918-928.

Kalscheuer R (2010). Genetics of wax ester and triacylglycerol biosynthesis in bacteria. Handbook of Hydrocarbon and Lipid Microbiology. Springer. pp 527-535.

Kartal B, de Almeida NM, Maalcke WJ, Op den Camp HJM, Jetten MSM, Keltjens JT. (2013). How to make a living from anaerobic ammonium oxidation. FEMS Microbiology Reviews **37:** 428-461.

Keller A, Nesvizhskii AI, Kolker E, Aebersold R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Analytical Chemistry **74:** 5383-5392.

Kent WJ. (2002). BLAT—the BLAST-like alignment tool. Genome Research **12:** 656-664.

Kessner D, Chambers M, Burke R, Agus D, Mallick P. (2008). ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics **24:** 2534-2536.

Kim H, Hao O, McAvoy T. (2001). SBR system for phosphorus removal: ASM2 and simplified linear model. Journal of Environmental Engineering **127:** 98-104.

Kim SD, Cho J, Kim IS, Vanderford BJ, Snyder SA. (2007). Occurrence and removal of pharmaceuticals and endocrine disruptors in South Korean surface, drinking, and waste waters. Water Research **41:** 1013-1021.

Kindaichi T, Nierychlo M, Kragelund C, Nielsen JL, Nielsen PH. (2013). High and stable substrate specificities of microorganisms in enhanced biological phosphorus removal plants. Environmental Microbiology **15:** 1821-31.

Kitano H (2001). Systems Biology: Toward System-level Understanding of Biological Systems. In: Kitano H (ed). *Foundations of Systems Biology*. The MIT Press: Cambridge, Massachusetts, USA. pp 1-36.

Kitano H. (2002). Systems Biology: A Brief Overview. Science **295:** 1662-1664.

Klein AN, Frigon D, Raskin L. (2007). Populations related to *Alkanindiges*, a novel genus containing obligate alkane degraders, are implicated in biological foaming in activated sludge systems. Environmental Microbiology **9:** 1898-1912.

Knight R, Jansson J, Field D, Fierer N, Desai N, Fuhrman JA *et al*. (2012). Unlocking the potential of metagenomics through replicated experimental design. Nature Biotechnology **30:** 513-520.

Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A *et al*. (2011). PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS ONE **6:** e15925.

Kong Y, Xia Y, Nielsen JL, Nielsen PH. (2006). Ecophysiology of a group of uncultured *Gammaproteobacterial* glycogen-accumulating organisms in full-scale enhanced biological phosphorus removal wastewater treatment plants. Environmental Microbiology **8:** 479-489.

Kozarewa I, Turner DJ (2011). 96-plex molecular barcoding for the illumina genome analyzer. *High-Throughput Next Generation Sequencing*. Springer. pp 279-298.

Kragelund C, Nielsen JL, Thomsen TR, Nielsen PH. (2005). Ecophysiology of the filamentous *Alphaproteobacterium Meganema perideroedes* in activated sludge. FEMS Microbiology Ecology **54:** 111-112.

Kragelund C, Remesova Z, Nielsen JL, Thomsen TR, Eales K, Seviour R *et al*. (2007). Ecophysiology of mycolic acid-containing Actinobacteria (*Mycolata*) in activated sludge foams. FEMS Microbiology Ecology **61:** 174-184.

Kragelund C, Levantesi C, Borger A, Thelen K, Eikelboom D, Tandoi V *et al*. (2008). Identity, abundance and ecophysiology of filamentous bacteria belonging to the *Bacteroidetes* present in activated sludge plants. Microbiology **154:** 886-894.

Kraume M, Bracklow U, Vocks M, Drews A. (2005). Nutrients removal in MBRs for municipal wastewater treatment. Water Science & Technology **51:** 391-402.

Kristensen GH, Jørgensen PE, Nielsen PH. (1994). Settling characteristics of activated sludge in Danish treatment plants with biological nutrient removal. Water Science & Technology **29:** 157-165.

Kuffner M, Hai B, Rattei T, Melodelima C, Schloter M, Zechmeister-Boltenstern S *et al*. (2012). Effects of season and experimental warming on the bacterial community in a temperate mountain forest soil assessed by 16S rRNA gene pyrosequencing. FEMS Microbiology Ecology **82:** 551-562.

Kultima JR, Sunagawa S, Li J, Chen W, Chen H, Mende DR *et al*. (2012). MOCAT: a metagenomics assembly and gene prediction toolkit. PLoS ONE **7:** e47656.

Kwon EE, Kim S, Jeon YJ, Yi H. (2012). Biodiesel production from sewage sludge: new paradigm for mining energy from municipal hazardous material. Environmental science & technology **46:** 10222-10228.

Lacko N, Bux F, Kasan H. (1999). Survey of filamentous bacteria in activated sludge plants in KwaZulu-Natal. WATER SA-PRETORIA- **25:** 63-68.

Larsen P, Collart F, Field D, Meyer F, Keegan K, Henry C *et al*. (2011). Predicted Relative Metabolomic Turnover (PRMT): determining metabolic turnover from a coastal marine metagenomic dataset. Microb Informatics Exp **1:** 1-11.

Lee S, Seo CH, Lim B, Yang JO, Oh J, Kim M *et al*. (2011). Accurate quantification of transcriptome from RNA-Seq data by effective length normalization. Nucleic Acids Research **39:** e9-e9.

Leimena MM, Ramiro-Garcia J, Davids M, van den Bogert B, Smidt H, Smid EJ *et al*. (2013). A comprehensive metatranscriptome analysis pipeline and its validation using human small intestine microbiota datasets. BMC Genomics **14:** 530.

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

Levantesi C, Rossetti S, Thelen K, Kragelund C, Krooneman J, Eikelboom D *et al*. (2006a). Phylogeny, physiology and distribution of '*Candidatus* Microthrix calida', a new *Microthrix* species isolated from industrial activated sludge wastewater treatment plants. Environmental Microbiology **8:** 1552-1563.

Li M, Wang B, Zhang M, Rantalainen M, Wang S, Zhou H *et al*. (2008). Symbiotic gut microbes modulate human metabolic phenotypes. Proceedings of the National Academy of Sciences **105:** 2117-2122.

Li R, Li Y, Kristiansen K, Wang J. (2008). SOAP: short oligonucleotide alignment program. Bioinformatics **24:** 713-714.

Libhaber M, Orozco-Jaramillo Á. (2012). Sustainable treatment and reuse of municipal wastewater: for decision makers and practicing engineers. Iwa publishing.

Lin LI. (1989). A concordance correlation coefficient to evaluate reproducibility. Biometrics **45:** 255-268.

Liu CM, Soldanova K, Nordstrom L, Dwan MG, Moss OL, Contente-Cuomo TL *et al.* (2013): Medical therapy reduces microbiota diversity and evenness in surgically recalcitrant chronic rhinosinusitis. *International forum of allergy & rhinology*, **on line access**.

Liu Z, Klatt CG, Wood JM, Rusch DB, Ludwig M, Wittekindt N *et al*. (2011). Metatranscriptomic analyses of chlorophototrophs of a hot-spring microbial mat. The ISME Journal **5:** 1279-1290.

Logan BE, Rabaey K. (2012). Conversion of wastes into bioelectricity and chemicals by using microbial electrochemical technologies. Science **337:** 686-690.

Lolo M, Pedreira S, Vázquez B, Franco C, Cepeda A, Fente C. (2007). Cryogenic grinding pre-treatment improves extraction efficiency of fluoroquinolones for HPLC-MS/MS determination in animal tissue. Analytical and Bioanalytical Chemistry **387:** 1933-1937.

Maixner F, Noguera DR, Anneser B, Stoecker K, Wegl G, Wagner M *et al*. (2006). Nitrite concentration influences the population structure of *Nitrospira*-like bacteria. Environmental Microbiology **8:** 1487-1495.

Majewsky M, Gallé T, Zwank L, Fischer K. (2010). Influence of microbial activity on polar xenobiotic degradation in activated sludge systems. Water science and technology: a journal of the International Association on Water Pollution Research **62:** 701.

Manchester KL. (1995). Value of A260/A280 ratios for measurement of purity of nucleic acids. Biotechniques **19:** 208-210.

Manilla-Pérez E, Lange AB, Hetzler S, Steinbüchel A. (2010). Occurrence, production, and export of lipophilic compounds by hydrocarbonoclastic marine bacteria and their potential use to produce bulk chemicals from hydrocarbons. Applied Microbiology and Biotechnology **86:** 1693-1706.

Mara D, Horan NJ. (2003) Handbook of water and wastewater microbiology. Academic press.

Markowitz VM, Chen I-MA, Chu K, Szeto E, Palaniappan K, Grechkin Y *et al*. (2012). IMG/M: the integrated metagenome data management and comparative analysis system. Nucleic Acids Research **40:** D123-D129.

Martens-Habbena W, Berube PM, Urakawa H, de La Torre JR, Stahl DA. (2009). Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. Nature **461:** 976-979.

Martins AMP, Pagilla K, Heijnen JJ, van Loosdrecht MCM. (2004). Filamentous bulking sludge—a critical review. Water Research **38:** 793-817.

Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. (2012). PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics **13:** 31.

Mashego MR, Rumbold K, De Mey M, Vandamme E, Soetaert W, Heijnen JJ. (2007). Microbial metabolomics: past, present and future methodologies. Biotechnol Lett **29:** 1-16.

McIlroy SJ, Kristiansen R, Albertsen M, Karst SM, Rossetti S, Nielsen JL *et al*. (2013). Metabolic model for the filamentous '*Candidatus* Microthrix parvicella'based on genomic and metagenomic analyses. The ISME Journal **7:** 1161-72.

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

McMahon KD, Martin HG, Hugenholtz P. (2007). Integrating ecology into biotechnology. Current Opinion in Biotechnology **18:** 287-292.

Metcalf, Eddy. (2003). *Wastewater engineering: treatment and reuse*, 4th edn. McGraw-Hill Higher Education: New York.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M *et al*. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics **9:** 386.

Miana P, Grando L, Caravello G, Fabris M. (2002). Microthrix parvicella foaming at the Fusina WWTP. Water science and technology: a journal of the International Association on Water Pollution Research **46:** 499.

Michael Beman J, Arrigo KR, Matson PA. (2005). Agricultural runoff fuels large phytoplankton blooms in vulnerable areas of the ocean. Nature **434:** 211-214.

Mielczarek AT, Kragelund C, Eriksen PS, Nielsen PH. (2012). Population dynamics of filamentous bacteria in Danish wastewater treatment plants with nutrient removal. Water Research **46:** 3781-3795.

Mondala A, Liang K, Toghiani H, Hernandez R, French T. (2009). Biodiesel production by in situ transesterification of municipal primary and secondary sludges. Bioresource Technology **100:** 1203-1210.

Morales SE, Holben WE. (2011). Linking bacterial identities and ecosystem processes: can 'omic'analyses be more than the sum of their parts? FEMS Microbiology Ecology **75:** 2-16.

Morgenroth E, Kommedal R, Harremoës P. (2002). Processes and modeling of hydrolysis of particulate organic matter in aerobic wastewater treatment – a review. Water Science & Technology **45:** 25-40.

Morse S, Shaw G, Larner S. (2006). Concurrent mRNA and protein extraction from the same experimental sample using a commercially available column-based RNA preparation kit Biotechniques **40:** 54–58.

Mulder A, Graaf A, Robertson L, Kuenen J. (1995). Anaerobic ammonium oxidation discovered in a denitrifying fluidized bed reactor. FEMS Microbiology Ecology **16:** 177-184.

Muller EE, Glaab E, May P, Vlassis N, Wilmes P. (2013). Condensing the omics fog of microbial communities. Trends in Microbiology **21:** 325-33.

Muller EEL, Pinel N, Gillece JD, Schupp JM, Price LB, Engelthaler DM *et al*. (2012). Genome sequence of "*Candidatus* Microthrix parvicella" BIO17-1, a long-chain-fatty-acid-accumulating filamentous actinobacterium from a biological wastewater treatment plant. Journal of Bacteriology **194:** 6670-6671.

Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. (2011). A guide to the natural history of freshwater lake bacteria. Microbiology and Molecular Biology Reviews **75:** 14-49.

Ng C, DeMaere MZ, Williams TJ, Lauro FM, Raftery M, Gibson JA *et al*. (2010). Metaproteogenomic analysis of a dominant green sulfur bacterium from Ace Lake, Antarctica. The ISME Journal **4:** 1002-1019.

Nielsen JL, Vidal M. (2010). Systems biology of microorganisms. Current Opinion in Microbiology **13:** 335-336.

Nielsen JL, Christensen D, Kloppenborg M, Nielsen PH. (2003). Quantification of cell-specific substrate uptake by probe-defined bacteria under in situ conditions by microautoradiography and fluorescence *in situ* hybridization. Environmental Microbiology **5:** 202-211.

Nielsen PH, Muro MAd, Nielsen JL. (2000). Studies on the *in situ* physiology of *Thiothrix* spp. present in activated sludge. Environmental Microbiology **2:** 389-398.

Nielsen PH, Roslev P, Dueholm TE, Nielsen JL. (2002). *Microthrix parvicella*, a specialized lipid consumer in anaerobic–aerobic activated sludge plants. Water Science & Technology **46:** 73-80.

Nielsen PH, Kragelund C, Seviour RJ, Nielsen JL. (2009). Identity and ecophysiology of filamentous bacteria in activated sludge. FEMS Microbiology Reviews **33:** 969-998.

Nielsen PH, Mielczarek AT, Kragelund C, Nielsen JL, Saunders AM, Kong Y *et al*. (2010). A conceptual ecosystem model of microbial communities in enhanced biological phosphorus removal plants. Water Research **44:** 5070-5088.

Noutsopoulos C, Mamais D, Andreadakis A. (2012). A hypothesis on *Microthrix parvicella* proliferation in biological nutrient removal activated sludge systems with selector tanks. FEMS Microbiology Ecology **80:** 380-389.

Oberhardt MA, Palsson BO, Papin JA. (2009). Applications of genome-scale metabolic reconstructions. Mol Syst Biol **5**.

Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ *et al*. (2007). The vegan package. Community ecology package Disponível em: http://www.R-project.org/Acesso/em **10:** 2008.

Olukoshi ER, Packter NM. (1994). Importance of stored triacylglycerols in Streptomyces: possible carbon source for antibiotics. Microbiology **140:** 931-943.

Ostle AG, Holt JG. (1982). Nile blue A as a fluorescent stain for poly-beta-hydroxybutyrate. Appl Environ Microbiol **44:** 238-241.

Pace NR. (1997). A molecular view of microbial diversity and the biosphere. Science **276:** 734-740.

Paine RT (1995). A conversation on refining the concept of keystone species. JSTOR **9:** 962-964.

Parawira W. (2010). Biodiesel production from Jatropha curcas: a review. Scientific Research and Essays **5:** 1796-1808.

Paredes D, Kuschk P, Mbwette T, Stange F, Müller R, Köser H. (2007). New aspects of microbial nitrogen transformations in the context of wastewater treatment–a review. Engineering in Life Sciences **7:** 13-25.

Park H-D, Wells GF, Bae H, Criddle CS, Francis CA. (2006). Occurrence of ammonia-oxidizing archaea in wastewater treatment plant bioreactors. Applied and Environmental Microbiology **72:** 5643-5647.

Patel A, Noble RT, Steele JA, Schwalbach MS, Hewson I, Fuhrman JA. (2007). Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. Nat Protocols **2:** 269-276.

Paul E, Liu Y. (2012). Biological sludge minimization and biomaterials/bioenergy recovery technologies. Wiley.

Prosser JI, Bohannan BJ, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP *et al*. (2007). The role of ecological theory in microbial ecology. Nature Reviews Microbiology **5:** 384-392.

Qian P-Y, Wang Y, Lee OO, Lau SC, Yang J, Lafi FF *et al*. (2010). Vertical stratification of microbial communities in the Red Sea revealed by 16S rDNA pyrosequencing. The ISME Journal **5:** 507-518.

Quéméneur M, Marty Y. (1994). Fatty acids and sterols in domestic wastewaters. Water Research **28:** 1217-1226.

Quince C, Lanzén A, Curtis TP, Davenport RJ, Hall N, Head IM *et al*. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. Nature Methods **6:** 639-641.

Radax R, Rattei T, Lanzen A, Bayer C, Rapp HT, Urich T *et al*. (2012). Metatranscriptomics of the marine sponge Geodia barretti: tackling phylogeny and function of its microbial community. Environmental Microbiology **14:** 1308-1324.

Radpour R, Sikora M, Grussenmeyer T, Kohler C, Barekati Z, Holzgreve W *et al*. (2009). Simultaneous isolation of DNA, RNA, and proteins for genetic, epigenetic, transcriptomic, and proteomic analysis. Journal of Proteome Research **8:** 5264-5274.

Raes J, Bork P. (2008). Molecular eco-systems biology: towards an understanding of community function. Nat Rev Micro **6:** 693-699.

Ram RJ, VerBerkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC, II *et al*. (2005). Community proteomics of a natural microbial biofilm. Science **308:** 1915-1920.

Ramette A. (2007). Multivariate analyses in microbial ecology. FEMS Microbiology Ecology **62:** 142-160.

Rashid U, Anwar F, Moser BR, Ashraf S. (2008). Production of sunflower oil methyl esters by optimized alkali-catalyzed methanolysis. Biomass and bioenergy **32:** 1202-1205.

Raunkjær K, Hvitved-Jacobsen T, Nielsen PH. (1994). Measurement of pools of protein, carbohydrate and lipid in domestic wastewater. Water Science & Technology **28:** 251-262.

Reaves ML, Rabinowitz JD. (2011). Metabolomics in systems microbiology. Current Opinion in Biotechnology **22:** 17-25.

Rho M, Tang H, Ye Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Research **38:** e191-e191.

Rittmann BE, McCarty PL. (2001). *Environmental biotechnology*. McGraw-Hill New York.

Rittmann BE. (2006). Microbial ecology to manage processes in environmental biotechnology. Trends in Biotechnology **24:** 261-266.

Rittmann BE. (2008). Opportunities for renewable bioenergy using microorganisms. Biotechnology and Bioengineering **100:** 203-212.

Roesch LFW, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD *et al*. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J **1:** 283-290.

Röling WFM, Ferrer M, Golyshin PN. (2010). Systems approaches to microbial communities and their functioning. Current Opinion in Biotechnology **21:** 532-538.

Rossetti S, Christensson C, Blackall LL, Tandoi V. (1997). Phenotypic and phylogenetic description of an Italian isolate of "*Microthrix parvicella*". Journal of Applied Microbiology **82:** 405-410.

Rossetti S, Tomei MC, Nielsen PH, Tandoi V. (2005). "*Microthrix parvicella*", a filamentous bacterium causing bulking and foaming in activated sludge systems: a review of current knowledge. FEMS Microbiology Reviews **29:** 49-64.

Röttig A, Steinbüchel A. (2013). Random mutagenesis of *atfA* and screening for *Acinetobacter baylyi* mutants with an altered lipid accumulation. Eur J Lipid Sci Technol **115:** 394-404.

Roume H, Muller EE, Cordes T, Renaut J, Hiller K, Wilmes P. (2013). A biomolecular isolation framework for eco-systems biology. The ISME Journal **7:** 110-121.

Salem S, Berends D, Heijnen J, Van Loosdrecht M. (2003). Bio-augmentation by nitrification with return sludge. Water Research **37:** 1794-1804.

Santala S, Efimova E, Kivinen V, Larjo A, Aho T, Karp M *et al.* (2011). Improved Triacylglycerol Production in *Acinetobacter baylyi* ADP 1 by Metabolic Engineering. Microbial cell factories **10**: 36-36.

Sauer U, Heinemann M, Zamboni N. (2007). Genetics: getting closer to the whole picture. Science **316:** 550-551.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al*. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology **75:** 7537-7541.

Schmid M, Twachtmann U, Klein M, Strous M, Juretschko S, Jetten M *et al*. (2000). Molecular evidence for genus level diversity of bacteria capable of catalyzing anaerobic ammonium oxidation. Systematic and Applied Microbiology **23:** 93-106.

Schmidt I, Sliekers O, Schmid M, Bock E, Fuerst J, Kuenen JG et al. (2003). New concepts of microbial treatment processes for the nitrogen removal in wastewater. FEMS Microbiology Reviews **4:** 481-492.

Schuyler RG, Tamburini JR, Tamburini SJ. (2011). Using the Oxidation Index (OXI) as an activated sludge process control tool. Proceedings of the Water Environment Federation **2011:** 5405-5421.

Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS, Huttenhower C. (2013). Computational meta'omics for microbial community studies. Molecular Systems Biology **9**.

Seviour EM, Williams C, DeGrey B, Soddell JA, Seviour RJ, Lindrea KC. (1994). Studies on filamentous bacteria from australian activated sludge plants. Water Research **28:** 2335-2342.

Seviour RJ, Nielsen PH. (2010). *Microbial ecology of activated sludge*. IWA Publishing.

Seviour RJ, Pethica LM, McClure S. (1984). A simple modified procedure for preparing microbial cells for scanning electron microscopy. Journal of Microbiological Methods **3:** 1-5.

Seviour RJ, Maszenan AM, Soddell JA, Tandoi V, Patel BKC, Kong Y *et al*. (2000). Microbiology of the 'G-bacteria' in activated sludge. Environmental Microbiology **2:** 581-593.

Seviour RJ, Mino T, Onuki M. (2003). The microbiology of biological phosphorus removal in activated sludge systems. FEMS Microbiology Reviews **27:** 99-127.

Seviour RJ, Kragelund C, Kong Y, Eales K, Nielsen JL, Nielsen PH. (2008). Ecophysiology of the *Actinobacteria* in activated sludge systems. Antonie van Leeuwenhoek **94:** 21-33.

Seviour RJ. (2010). *Microbial ecology of activated sludge*. International Water Assn.

Shi S, Valle-Rodríguez JO, Siewers V, Nielsen J. (2011). Prospects for microbial biodiesel production. Biotechnology Journal **6:** 277-285.

Shteynberg D, Deutsch EW, Lam H, Eng JK, Sun Z, Tasman N *et al*. (2011). iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Molecular & Cellular Proteomics **10**.

Siddiquee MN, Kazemian H, Rohani S. (2011). Biodiesel production from the lipid of wastewater sludge using an acidic heterogeneous catalyst. Chemical Engineering & Technology **34:** 1983-1988.

Singh SP, Singh D. (2010). Biodiesel production through the use of different sources and characterization of oils and their esters as the substitute of diesel: a review. Renewable and Sustainable Energy Reviews **14:** 200-216.

Slijkhuis H. (1983). *Microthrix parvicella*, a filamentous bacterium isolated from activated sludge: cultivation in a chemically defined medium. Appl Environ Microbiol **46:** 832-839.

Slijkhuis H, Van Groenestijn JW, Kylstra DJ. (1984). *Microthrix parvicella*, a filamentous bacterium from activated sludge: growth on Tween 80 as carbon and energy source. Journal of General Microbiology **130:** 2035-2042.

Slijkhuis H, Deinema MH. (1988). Effect of environmental conditions on the occurrence of *Microthrix parvicella* in activated sludge. Water Research **22:** 825-828.

Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T. (2011). Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics **27:** 431-432.

Soddell JA, Seviour RJ. (1994). Incidence and morphological variability of *Nocardia pinensis* in australian activated sludge plants. Water Research **28:** 2343-2351.

Sonune A, Ghate R. (2004). Developments in wastewater treatment methods. Desalination **167:** 55-63.

Sorokin DY, Lücker S, Vejmelkova D, Kostrikina NA, Kleerebezem R, Rijpstra WIC *et al*. (2012). Nitrification expanded: discovery, physiology and genomics of a nitrite-oxidizing bacterium from the phylum *Chloroflexi*. The ISME Journal **6:** 2245-2256.

Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, Barofsky DF *et al*. (2009). Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. ISME J **3:** 93-105.

Srinath EG, Sastry CA, Pillai SC. (1959). Rapid removal of phosphorus from sewage by activated sludge. Experientia **15:** 339-340.

Staley JT, Konopka A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annual Reviews in Microbiology **39:** 321-346.

Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, Leigh JA *et al*. (2007). Metabolic modeling of a mutualistic microbial community. Mol Syst Biol **3**.

Stöveken T, Kalscheuer R, Malkus U, Reichelt R, Steinbüchel A. (2005). The Wax Ester Synthase/Acyl Coenzyme A:Diacylglycerol Acyltransferase from *Acinetobacter* sp. Strain ADP1: Characterization of a Novel Type of Acyltransferase. Journal of Bacteriology **187:** 1369-1376.

Strous M, Fuerst JA, Kramer EHM, Logemann S, Muyzer G, van de Pas-Schoonen KT *et al*. (1999). Missing lithotroph identified as new planctomycete. Nature **400:** 446-449.

Subramaniam R, Dufreche S, Zappi M, Bajpai R. (2010). Microbial lipids from renewable resources: production and characterization. J Ind Microbiol Biotechnol **37:** 1271-1287.

Tandoi V, Rossetti S, Blackall LL, Majone M. (1998). Some physiological properties of an Italian isolate of "*Microthrix parvicella*". Water Science and Technology **37:** 1-8.

Tang J. (2011). Microbial metabolomics. Current Genomics **12:** 391.

Tchobanoglous G, Burton FL. (2003). Wastewater engineering. Management **7:** 1-4.

Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK *et al*. (2013). A community-driven global reconstruction of human metabolism. Nature Biotechnology **31:** 419-425.

Thomas T, Gilbert J, Meyer F. (2012). Metagenomics-a guide from sampling to data analysis. Microb Inform Exp **2**.

Tolosa JM, Schjenken1 JE, Civiti TD, Clifton VL, Smith R. (2007). Column-based method to simultaneously extract DNA, RNA, and proteins from the same sample. Biotechniques **43:** 799–804.

Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW *et al*. (2005). Comparative metagenomics of microbial communities. Science **308:** 554-557.

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM *et al*. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature **428:** 37-43.

Van Veen WL. (1973). Bacteriology of activated-sludge, in particular filamentous bacteria. Antonie Van Leeuwenhoek Journal of Microbiology **39:** 189-205.

Vance JE, Vance DE. (2008). *Biochemistry of lipids, lipoproteins and membranes*. Access Online via Elsevier.

Velagapudi VR, Hezaveh R, Reigstad CS, Gopalacharyulu P, Yetukuri L, Islam S *et al*. (2010). The gut microbiota modulates host energy and lipid metabolism in mice. Journal of Lipid Research **51:** 1101-1112.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al*. (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea. Science **304:** 66-74.

VerBerkmoes NC, Denef VJ, Hettich RL, Banfield JF. (2009a). Systems Biology: Functional analysis of natural microbial consortia using community proteomics. Nat Rev Micro **7:** 196-205.

Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J *et al*. (2009b). Shotgun metaproteomics of the human distal gut microbiota. ISME J **3:** 179-189.

Viant MR, Sommer U. (2013). Mass spectrometry based environmental metabolomics: a primer and review. Metabolomics **9:** 144-158.

Vlassis N, Pacheco MP, Sauter T. (2013). Fast reconstruction of compact context-specific metabolic network models. arXiv preprint arXiv:13047992.

Wagner M, Amann R, Kämpfer P, Assmus B, Hartmann A, Hutzler P *et al*. (1994). Identification and *in situ* detection of Gram-negative filamentous bacteria in activated sludge. Systematic and Applied Microbiology **17:** 405-417.

Wagner M, Smidt H, Loy A, Zhou J. (2007). Unravelling microbial communities with DNA-microarrays: challenges and future directions. Microbial Ecology **53:** 498-506.

Wältermann M, Steinbüchel A. (2005). Neutral lipid bodies in prokaryotes: recent insights into structure, formation, and relationship to eukaryotic lipid depots. J Bacteriol **187:** 3607-3619.

Wältermann M, Stöveken T, Steinbüchel A. (2007). Key enzymes for biosynthesis of neutral lipid storage compounds in prokaryotes: Properties, function and occurrence of wax ester synthases/acyl-CoA:diacylglycerol acyltransferases. Biochimie **89:** 230-242.

Wang Z, Gerstein M, Snyder M. (2008). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet **10:** 57-63.

Wanner J, Ruzicková I, Jetmarová P, Krhutková O, Paraniaková J. (1998). A national survey of activated sludge separation problems in the Czech Republic: filaments, floc characteristics and activated sludge metabolic properties. Water Science and Technology **37:** 271-279.

Warnecke F, Hess M. (2009). A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. Journal of Biotechnology **142:** 91-95.

Weckwerth W, Wenzel K, Fiehn O. (2004). Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. Proteomics **4:** 78-83.

Wilmes P, Bond PL. (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. Environ Microbiol **6:** 911-920.

Wilmes P, Wexler M, Bond PL. (2008). Metaproteomics provides functional insight into activated sludge wastewater treatment. PLoS ONE **3:** e1778.

Wilmes P, Bond PL. (2009a). Microbial community proteomics: elucidating the catalysts and metabolic mechanisms that drive the Earth's biogeochemical cycles. Current Opinion in Microbiology **12:** 310-7.

Wilmes P, Simmons SL, Denef VJ, Banfield JF. (2009b). The dynamic genetic repertoire of microbial communities. FEMS Microbiology Reviews **33:** 109-132.

Wilmes P, Bowen BP, Thomas BC, Mueller RS, Denef VJ, VerBerkmoes NC *et al*. (2010). Metabolome-proteome differentiation coupled to microbial divergence. mBio **1:** e00246-00210.

Wilson EO. (1999). *The diversity of life*. WW Norton & Company.

Windey K, De Bo I, Verstraete W. (2005). Oxygen-limited autotrophic nitrification–denitrification (OLAND) in a rotating biological contactor treating high-salinity wastewater. Water Research **39:** 4512-4520.

Wintermute EH, Silver PA. (2010). Emergent cooperation in microbial metabolism. Molecular Systems Biology **6**.

Witting M, Lucio M, Tziotis D, Schmitt-Kopplin P (2012). Ultrahigh resolution mass spectrometry based non-targeted microbial metabolomics. *Genetics Meets Metabolomics*. Springer. pp 57-71.

Woese CR, Kandler O, Wheelis ML. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences **87:** 4576-4579.

Xia Y, Kong Y, Thomsen TR, Halkjær Nielsen P. (2008). Identification and ecophysiological characterization of epiphytic protein-hydrolyzing *Saprospiraceae* ("*Candidatus* Epiflobacter" spp.) in activated sludge. Applied and Environmental Microbiology **74:** 2229-2238.

Xie B, Dai XC, Xu YT. (2007). Cause and pre-alarm control of bulking and foaming by *Microthrix parvicella*—A case study in triple oxidation ditch at a wastewater treatment plant. Journal of Hazardous Materials **143:** 184-191.

Xiong J, Yang Q, Kang J, Sun Y, Zhang T, Margaret G *et al*. (2011). Simultaneous isolation of DNA, RNA, and protein from *Medicago truncatula* L. Electrophoresis **32:** 321-330.

Yan Q, Zhang X, Zhang T, Fang HH. (2011). Seasonal microbial community shift in a saline sewage treatment plant. Frontiers of Environmental Science & Engineering in China **5:** 40-47.

Yu K, Zhang T. (2012). Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge. PLoS ONE **7:** e38183.

Ze X, Duncan SH, Louis P, Flint HJ. (2012). *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. The ISME Journal **6:** 1535-1543.

Zhang T, Shao M-F, Ye L. (2011). 454 Pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. The ISME Journal **6:** 1137-1147.

Zhang W, Li F, Nie L. (2010). Integrating multiple 'omics' analysis for microbial biology: application and methodologies. Microbiology **156:** 287-301.

Zhu G, Peng Y, Li B, Guo J, Yang Q, Wang S (2008). Biological removal of nitrogen from wastewater. In: Whitacre D (ed). *Reviews of Environmental Contamination and Toxicology*. Springer New York. pp 159-195.

Zomorrodi AR, Maranas CD. (2012). OptCom: A multi-level optimization framework for the metabolic modeling and analysis of microbial communities. PLoS Comput Biol **8:** e1002363.

Zweytick D, Athenstaedt K, Daum G. (2000). Intracellular lipid particles of eukaryotic cells. Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes **1469:** 101-120.

PUBLICATIONS

The ISME Journal (2013) 7, 110–121
© 2013 International Society for Microbial Ecology All rights reserved 1751-7362/13
www.nature.com/ismej

Open

ORIGINAL ARTICLE

A biomolecular isolation framework for eco-systems biology

Hugo Roume[1,2], Emilie EL Muller[1], Thekla Cordes[1], Jenny Renaut[2], Karsten Hiller[1] and Paul Wilmes[1,2]

[1]*Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg and* [2]*Department of Environment and Agro-Biotechnologies, Public Research Centre - Gabriel Lippmann, Belvaux, Luxembourg*

Mixed microbial communities are complex, dynamic and heterogeneous. It is therefore essential that biomolecular fractions obtained for high-throughput omic analyses are representative of single samples to facilitate meaningful data integration, analysis and modeling. We have developed a new methodological framework for the reproducible isolation of high-quality genomic DNA, large and small RNA, proteins, and polar and non-polar metabolites from single unique mixed microbial community samples. The methodology is based around reproducible cryogenic sample preservation and cell lysis. Metabolites are extracted first using organic solvents, followed by the sequential isolation of nucleic acids and proteins using chromatographic spin columns. The methodology was validated by comparison to traditional dedicated and simultaneous biomolecular isolation methods. To prove the broad applicability of the methodology, we applied it to microbial consortia of biotechnological, environmental and biomedical research interest. The developed methodological framework lays the foundation for standardized molecular eco-systematic studies on a range of different microbial communities in the future.

The ISME Journal (2013) **7,** 110–121; doi:10.1038/ismej.2012.72; published online 5 July 2012
Subject Category: integrated genomics and post–genomics approaches in microbial ecology
Keywords: biomolecules; eco-systems biology; extraction; isolation; metabolomics; sample heterogeneity

Introduction

Natural microbial communities play fundamental roles in the Earths biogeochemical cycles as well as in human health and disease, and provide essential services to mankind. They represent highly complex, dynamic and heterogeneous systems (Denef *et al.*, 2010). The advent of high-resolution molecular biology methodologies, including genomics (Tyson *et al.*, 2004), transcriptomics (Frias-Lopez *et al.*, 2008), proteomics (Ram *et al.*, 2005) and metabolomics (Li *et al.*, 2008), is facilitating unprecedented insights into the structure and function of microbial consortia *in situ*. Beyond isolated biomolecular characterization, integrated omics combined with relevant statistical analyses offer the ability to unravel fundamental ecological and evolutionary relationships, which are indiscernible from isolated omic data sets (Wilmes *et al.*, 2010). Furthermore, space- and time-resolved integrated omics have the potential to uncover associations between distinct biomolecules, which allows for discovery of previously unknown biochemical traits of specific microbial community members (Fischer *et al.*, 2011). Such linkages will, however, only be discernible if representative biomolecular fractions are obtained. Thus, within the emerging field of molecular eco-systems biology (Raes and Bork, 2008), molecule-level systematic characterizations of microbial consortia will only fulfill their full potential if biomolecular fractions (DNA, RNA, proteins and small molecules) are obtained from single unique samples (Kitano, 2001). Only then will the subsequent integration of high-resolution omic data enable true systems-level views of community-wide, population-wide and individual-level processes. To our knowledge, no methodology currently exists for the isolation of all concomitant small molecules (metabolites) and biomacromolecules (DNA, RNA and proteins) from biological systems.

A well-established method for the isolation of concomitant biomacromolecules only is based around the simultaneous addition of a monophasic mixture of phenol and guanidine isothiocyanate, commercially available as TRIzol and TRI Reagent (TR), and chloroform to biological samples to obtain

Correspondence: P Wilmes, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, L-4362 Esch-sur-Alzette, Luxembourg.
E-mail: paul.wilmes@uni.lu
Received 8 February 2012; revised 25 May 2012; accepted 29 May 2012; published online 5 July 2012

an aqueous phase containing primarily RNA, an interphase containing DNA and an organic phase containing proteins (Chomczynski, 1993; Hummon *et al.*, 2007; Chey *et al.*, 2011; Xiong *et al.*, 2011). Individual biomacromolecular fractions are purified from the respective phases and may be subjected to specialized downstream analyses. Adaptation of the standard TR protocol for the additional extraction of small molecules was carried out on plant material by Weckwerth *et al.* (2004). For this, a solvent mixture of methanol, chloroform and water is first used for the fractionation of small molecules into polar and non-polar metabolites, and the precipitation of biological macromolecules. Polar and non-polar metabolites are retrieved from the aqueous and organic phases, respectively. RNA and proteins are isolated from the remaining pellet following extraction in dedicated buffers and phenol, respectively. However, no genomic DNA fraction was obtained using this method, a need that is particularly important in microbial communities that exhibit extensive genetic heterogeneity (Wilmes *et al.*, 2009) and for which genomic context is thus essential for meaningful interpretation of functional omic data.

Owing to the hazardous chemicals involved and the methods being unsuited for routine high-throughput laboratory use, chromatographic spin column-based procedures have been introduced for the isolation of concomitant biomacromolecules (Morse *et al.*, 2006; Tolosa *et al.*, 2007; Radpour *et al.*, 2009). These methods rely on the pH- and salt-concentration-dependent adsorption of nucleic acids and proteins to solid phases such as silica or glass. The solid phases are washed and the biomacromolecules of interest are sequentially eluted. Such methods are now available as commercial kits from Qiagen (AllPrep DNA/RNA/Protein Mini kit, QA) and Norgen Biotek (All-in-One Purification kit for large RNA, small/microRNA, total proteins and genomic DNA; NA). The latter has the advantage of offering the ability to deplete the total RNA fraction of small RNA (<200 nt), which can be analyzed separately. Chromatographic spin column-based biomacromolecular isolation has yet to be combined with the extraction of concomitant small molecules.

Here, we combine sample processing, sample cryopreservation, cell disruption by cryomilling, small molecule extraction and biomacromolecular isolation based on chromatographic spin columns to result in a universal methodological framework that allows the standardized isolation of extracellular and intracellular polar and non-polar metabolites, genomic DNA, RNA (divided into large and small RNA fractions) and proteins from single microbial community samples (Figure 1). We validated the performance of the methodology by comparison to widely used exclusive and simultaneous biomolecular isolation methods. Furthermore, we proved its general applicability to diverse mixed microbial communities of biotechnological, environmental and biomedical research interest, that is, lipid-accumulating organisms (LAOs), that may be exploited for the reclamation of energy-rich lipids from wastewater, river water filtrate and human feces. The framework may also find broader appeal for integrated omics on other heterogeneous and/or precious biological samples.

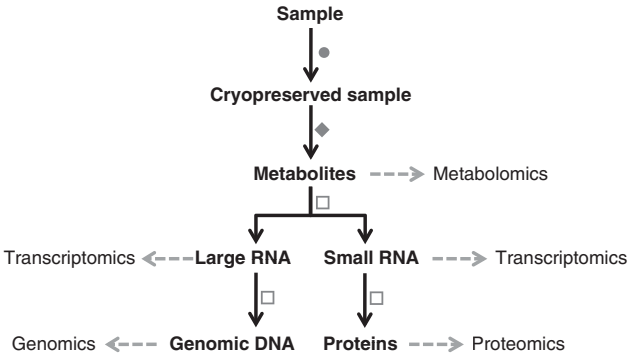

Figure 1 The developed biomolecular isolation framework highlighting the NA-based methodological workflow. Legend: (●) Sample processing and preservation: immediate snap-freezing by immersion in liquid nitrogen for LAO-enriched mixed microbial communities; concentration by tangential flow filtration followed by high-speed centrifugation, then snap-freezing of the resulting cell pellet for freshwater mixed microbial communities; homogenization with RNAlater followed by centrifugation steps before snap-freezing for fresh human fecal samples (Materials and methods). (◆) Cryomilling and metabolite extraction: cryomilling of cell pellets and solvent extraction of the intracellular polar and non-polar metabolite fractions (extracellular fractions were only prepared for the LAO-enriched microbial communities; Materials and methods). (□) Physicochemical biomacromolecular isolation: use of sequential physicochemical separation based around chromatographic spin columns following bead-beating in the modified NA-lysis buffer, resulting in the isolation of high-purity biomacromolecular fractions (Materials and methods).

Materials and methods

For more details see Supplementary Materials and methods.

Sampling and sample processing

LAO-enriched mixed microbial community. Floating LAO biomass was sampled from the air–water interface of the anoxic activated sludge tank at the Schifflange wastewater treatment plant (Esch-sur-Alzette, Luxembourg; 49°30′48.29″N; 6°1′4.53″E). For each sampling date, four different 'islets' (I1-I4, defined herein as biological replicates; Supplementary Figure 1) were sampled using a levy cane of 500 ml. Samples were collected in 50 ml sterile Falcon tubes and then immediately snap-frozen by immersion in liquid nitrogen and stored at -80 °C.

For the determination of sample heterogeneity by metabolomics, four replicates of 200 mg of LAO biomass (technical replicates) were obtained

from four different islets (biological replicates) for four different dates, leading to an overall sample set of 64. The different sampling dates were chosen to reflect different LAO microbial community compositions. Samples were taken on 4 October 2010 (date 1, D1; wastewater temperature 20.7 °C), 25 October 2010 (date 2, D2; 18.9 °C), 25 January 2011 (date 3, D3; 14.5 °C) and 23 February 2011 (date 4, D4; 13.9 °C). For the development and assessment of the biomolecular extraction protocols, a single representative islet sample taken on 13 December 2010 was used. For the comparative metabolomic analysis of the three microbiomes, three technical replicates from D4 islet 4 were used. For each extraction, technical replicates of 200 mg were subsampled from the collected islet using a sterile metal spatula while at all times guaranteeing that the samples remained in the frozen state.

Following weighing out, 200 mg of LAO-enriched microbial community samples were briefly thawed on ice followed by centrifugation at 18 000 *g* for 10 min at 4 °C to separate the supernatant (~150 µl; extracellular fraction) from the biomass (intracellular fraction). The intracellular fraction was then immediately refrozen in liquid nitrogen before homogenization by cryomilling (Figure 1). In contrast, the extracellular fraction immediately underwent metabolite extraction.

Freshwater mixed microbial community. Forty liters of river water were collected at a depth of around 1 m from the Alzette river (Schifflange, Luxembourg; 49°30′28.04″N; 6°0′11.48″E). Cells were concentrated by tangential flow filtration and centrifugation (Supplementary Materials and methods). Resulting cell pellets were snap-frozen and stored at −80 °C until the cryomilling step (Figure 1).

Human fecal samples. Three fresh human fecal samples, 300 mg each, were collected from a young healthy individual and placed immediately on ice. Samples were treated with RNAlater and cell pellets were obtained following centrifugation (Supplementary Materials and methods). Pellets were stored at −80 °C until the cryomilling step (Figure 1).

Cryomilling. Each of the three different microbial community samples were homogenized by cryomilling for 2 min at 30 Hz using two 5 mm and five 2 mm stainless steel milling balls (Retsch, Haan, Germany) in a Mixer Mill MM 400 (Retsch; Figure 1).

Metabolite extractions

Extracellular metabolite extractions were only carried out on supernatant from the LAO-enriched microbial communities. For the river water filtrate and human fecal samples, supernatants were not obtainable because of the need for concentrating the river water sample by tangential flow filtration and the very limited liquid content in the human fecal samples, respectively.

Briefly, small molecules were cold (4 °C) solvent extracted by bead-beating (2 min at 20 Hz in a Retsch Mixer Mill MM400) the samples in defined mixtures of polar (methanol and water) and non-polar solvents (chloroform) using the same stainless steel balls as used previously for sample cryomilling. For a detailed description of the respective metabolite extraction protocols, see Supplementary Materials and methods. Following centrifugation at 14 000 *g* for 10 min at 4 °C, metabolite extractions resulted in an upper phase comprising polar metabolites, an interphase pellet comprising genomic DNA, large and small RNA, proteins and non-lysed cells, and a lower phase containing non-polar metabolites. Defined volumes of both polar and non-polar metabolites extracts were dried in specific gas chromatography (GC) glass vials before metabolomic analyses (Supplementary Materials and methods).

Sample processing and biomacromolecular isolations

After removal of the respective metabolite fractions, the interphase pellet (along with the steel milling balls) was kept on ice for the subsequent total RNA (enriched in large RNA), genomic DNA, small RNA and protein sequential isolations and purifications using the different methods as specified below. As the described metabolite extraction is a major modification of the typical extraction workflow, the interphase pellet was lysed in the respective lysis buffers by bead-beating at 30 Hz for 30 s at 4 °C (Retsch Mixer Mill MM 400) with stainless steel balls (the same as previously used for sample cryomilling and metabolite extraction steps).

The Norgen Biotek All-in-One Purification kit-based biomacromolecular isolation method (NA, Labomics S.A., Nivelles, Belgium; Figure 1) was applied to the interphase pellet according to the manufacturer's instructions with a few important modifications (Supplementary Materials and methods).

The Qiagen AllPrep DNA/RNA/Protein Mini kit-based method (QA, Qiagen, Venlo, The Netherlands) was applied to the interphase pellet according to the manufacturer's instructions (Supplementary Materials and methods).

The TRI Reagent-based method (TR, Sigma-Aldrich, Bornem, Belgium) was directly applied to the interphase pellet according to the manufacturer's instructions with a few important modifications (Supplementary Materials and methods).

Reference methods

To qualitatively and quantitatively assess the biomolecular fractions obtained through the sequential and simultaneous biomolecular isolation protocols, widely used dedicated biomolecular extraction and purification methods were used as reference

methods. In each case, the reference methods were applied to 200 mg of LAO-enriched biomass (a single islet sampled on 13 December 2010).

DNA extraction was performed using the PowerSoil DNA isolation kit (MO BIO Laboratories, Brussels, Belgium) according to the manufacturer's instructions (Supplementary Materials and methods). RNA extraction was performed with an RNeasy Mini kit (Qiagen) including the optional DNase treatment to eliminate contaminating genomic DNA according to the manufacturer's instructions. Protein extractions were carried out using a metaproteomic extraction method for activated sludge (Wilmes and Bond, 2004; Supplementary Materials and methods).

Metabolomics

Identical procedures were used for the analysis of both intracellular and extracellular metabolite fractions.

For the determination of sample heterogeneity by comparative metabolomics, gas chromatography (GC) coupled to mass spectroscopy (MS) measurements were performed on intracellular polar extracts from the LAO-enriched microbial communities obtained as specified above. A pool, from which analytical replicates were obtained, was also prepared by combining 100 µl of each of the 64 polar extracts. Aliquots of 40 µl of both intracellular polar extracts and of the pool were dried under vacuum and analyzed following resuspension and derivatization (Supplementary Materials and methods). Each pool aliquot represented an analytical replicate.

For the metabolomic analyses of the representative LAO-enriched microbial community samples, 50 µl of the polar and non-polar phase extracts of the extracellular and intracellular compartments were dried down and analyzed following resuspension and derivatization (Supplementary Materials and methods).

For the comparative metabolomic analysis of the different microbial communities, 50, 10 and 5 µl of the intracellular polar and non-polar phase extracts of the LAO-enriched microbial community, river water filtrate and diluted human fecal samples were dried and analyzed, respectively. To prevent overloading of the GC column, a 1 in 10 dilution in a 1:1 (v:v) methanol:water mixture was previously carried out on the raw polar phase extracts derived from the human fecal samples.

Biomacromolecular quality and quantity assessment

Isolated genomic DNA was separated by electrophoresis on agarose gels (Supplementary Materials and methods). PCR amplifications were carried out for each of the triplicate DNA fractions obtained using the NA-based method from the different microbial community samples, without any additional DNA purification steps (Supplementary Materials and methods).

For RNA quality assessment and quantification, an Agilent 2100 Bioanalyser (Agilent Technologies, Diegem, Belgium) was used. The Agilent RNA 6000 Nano kit and Agilent Small RNA kit for prokaryotes were used. To compare different RNA fraction traces obtained using the Agilent 2100 Bioanalyser, fluorescent units were normalized (Supplementary Materials and methods).

Genomic DNA and RNA fractions were quantified and assessed using a NanoDrop Spectrophotometer 1000 (Thermo Scientific, Erembodegem, Belgium).

Protein extracts were separated using 1D sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE; Bio-Rad Laboratories, Nazareth, Belgium; Supplementary Materials and methods). The protein fractions were quantified using a 2-D Quant kit (GE Healthcare, Diegen, Belgium).

Peptide tandem mass spectra were obtained on a Matrix-Assisted Laser Desorption Ionization-Time of Flight mass spectrometer (MALDI-ToF/ToF, 4800 Proteomics Analyzer, Applied Biosystems, Nieuwerkerk a/d IJssel, The Netherlands) after tryptic digestion of a dominant protein band excised from a SDS-PAGE gel on which an NA-based LAO-enriched microbial community protein extract had been separated.

Determination of intact cells versus cells with a compromised cell membrane

To evaluate the efficiency of cell lysis and variation between biomolecular extraction protocols on representative LAO-enriched microbial community samples, the percentage of damaged cells was determined by epifluorescent microscopic quantification after staining using the Live/Dead BacLight Bacterial Viability kit (Invitrogen, Merelbeke, Belgium; Supplementary Materials and methods).

Data treatment and statistical analyses

The data sets resulting from the individual analyses were processed and analyzed using appropriate statistical methods (Supplementary Materials and methods).

Results

Analysis of sample heterogeneity by comparative metabolomics

The metabolome represents the final output that results from the cellular interactions of the genome, transcriptome, proteome and environment and, thus, is the most sensitive indicator of cellular activity and sample-to-sample variation. To assess the extent of heterogeneity within microbial communities, we carried out an initial metabolomics experiment on spatially and temporally resolved LAO-enriched microbial communities from the air–water interface of a biological wastewater treatment plant (Supplementary Figure 1). These

LAO-enriched microbial communities allow spatially and temporally defined samples to be taken, exhibit medial species richness (α diversity of around 900 species; unpublished information) and are typically dominated by organisms belonging to the genus *Microthrix* spp. (Levantesi *et al*., 2006).

For the four different sampling dates, we extracted intracellular polar metabolites from four different biological replicates, that is, different islets (Supplementary Figure 1), and per islet we extracted four technical replicates (different subsamples derived from the same islet). In addition, an analytical reference sample for quality control and normalization, consisting of a pool of all polar metabolite extracts (Dunn *et al*., 2011), was prepared using the remaining polar extracts from all the samples (Materials and methods).

The samples were analyzed by GC coupled to mass spectrometry (GC–MS) and the peak intensities of detected metabolite features integrated using specialized software (Supplementary Materials and methods; Supplementary Table 1). The data obtained for 16 analytical pool replicates highlights the high reproducibility of the metabolomics method (mean concordance correlation coefficient (Lin, 1989) $=0.993\pm0.006$; mean CCC ± s.d.). Owing to the apparent unevenness of total integrated peak intensities in the metabolomics data (Supplementary Table 1) that results from sample heterogeneity, we chose a unit vector normalization strategy analogous to that commonly used for shotgun proteomics experiments (Florens *et al*., 2006). This involves dividing each metabolite peak intensity by the total peak intensity of the sample (Wilmes *et al*., 2010). To further exclude experimental variation, the metabolomics data set was filtered according to the presence in all analytical pool replicates. Finally, to account for instrument drift, which was observed for a very small fraction of metabolites, we normalized the data using the pool replicate data (Dunn *et al*., 2011) and finally re-normalized the data to the total intensity (Supplementary Materials and methods).

Using the normalized metabolomics data, extensive sample-to-sample variation is apparent for both biological (islets) and technical replicates. Although most samples from specific dates are distinguishable by their metabolomic profiles (Figure 2a and Supplementary Figure 2), extensive overlap between biological and technical replicates is apparent (Figure 2b), with numerous samples clustering outside of their respective replicate groups (Figure 2c). These results indicate that sample heterogeneity is an important consideration for integrated omic studies of natural microbial consortia.

Conceptualization of a biomolecular isolation framework

In the context of dealing with extensive sample-to-sample and endogenous sample heterogeneity, understanding of these complex systems' structure and dynamics requires integrated analyses of their parts, which in our case demands the isolation of concomitant metabolites, nucleic acids and proteins from single unique microbial community samples before the individual fractions are able to be subjected to specialized metabolomic, genomic, transcriptomic and proteomic analyses (Figure 1). Major considerations to allow comprehensive and reproducible extraction and purification of biomolecules from mixed microbial communities are: (i) standardized and representative sample preservation, (ii) indiscriminate cell lysis and (iii) retrieval of high-quality biomolecular fractions. These constraints were taken into account when devising the biomolecular isolation framework and resulted in a methodology, which is based on (i) immediate snap-freezing of mixed microbial community samples in liquid nitrogen following sampling (or after sample-specific cold preprocessing) and preservation of the biomass at a minimum of $-80\,^{\circ}$C, (ii) mechanical (cryogenic) and chemical cell lysis, and (iii) reliance on robust methods for biomolecular extraction and purification. The resultant methodological framework offers a fully integrated workflow that can easily be adjusted for specific samples and which allows flexibility for use of sequential or dedicated biomolecular extraction and purification protocols. We developed and validated the methodology on representative LAO-enriched microbial community sample (Materials and methods).

Cell lysis efficiency

Conservation of cell integrity before sample processing and representative cell lysis before biomolecular extraction are essential considerations for the methodology to result in reproducible and representative biomolecular fractions. Following snap-freezing in liquid nitrogen, sample preservation at $-80\,^{\circ}$C and thawing on ice, almost all cells are left intact (Figures 3a and d). The conceptualized methodological framework relies on an initial cryo-milling homogenization and lysis step followed by polar and non-polar metabolite extractions. These steps result in indiscriminate and comprehensive cell disruption and lysis (Figures 3b and d). Following the further addition of dedicated lysis buffers, the vast majority of cells ($90.20\pm6.46\%$; NA-based method mean ± s.d.) are lysed (Figures 3c and d). Importantly, as highlighted in Figure 3d, the lysis efficiencies of the methods allowing sequential biomolecular isolations (NA and QA) compare favorably to widely used reference methods for exclusive biomacromolecular isolation, that is, no statistically significant differences are apparent (Kruskal–Wallis, $P=0.189$, $n=10$), and they significantly outperform the standard TR-based simultaneous biomolecular isolation method (Kruskal–Wallis, $P=0.002$, $n=10$). These results validate

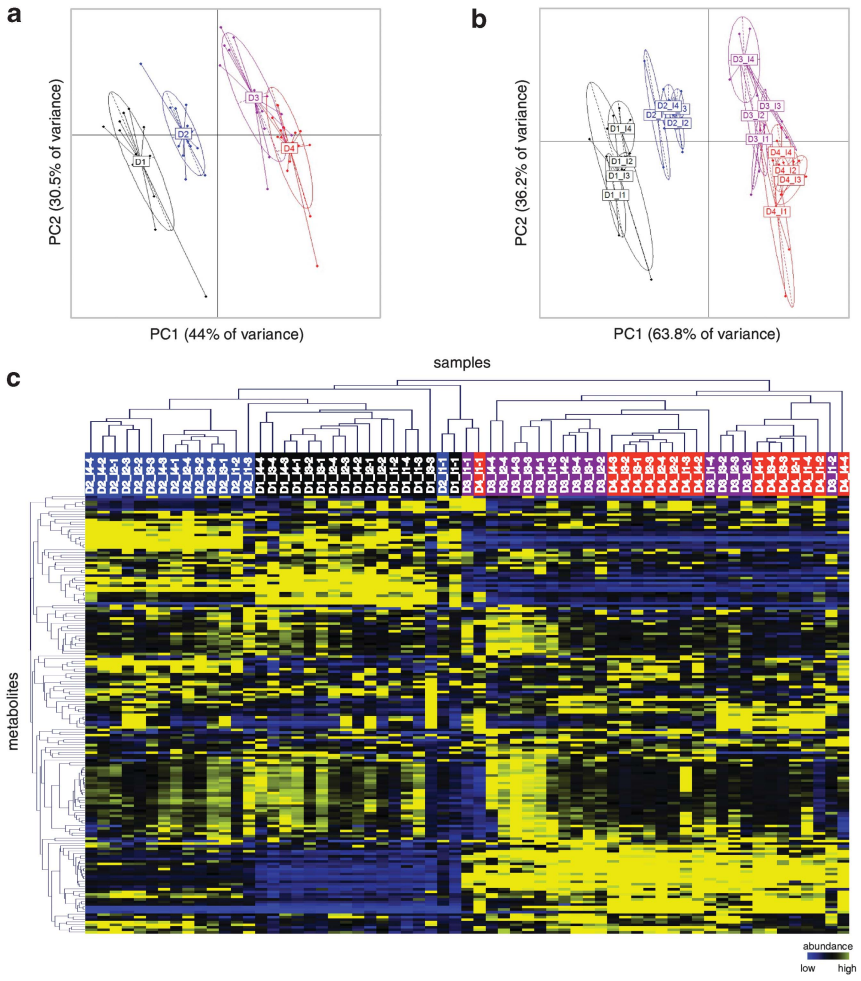

Figure 2 Metabolome heterogeneity within LAO-enriched microbial community samples. (**a**) Scatter plot of the two first principal components obtained using principal component analysis (PCA) of the normalized metabolomics data derived from the four biological replicates (islets; I1-I4), for each of the four distinct sampling dates (D1-D4). Each microbial community metabolome is indicated by a dot and color-coded according to sampling date. (**b**) Between-class PCA of the individual technical replicates for each biological replicate (islets; I1-I4). (**a** and **b**) The center of gravity for each date/islet cluster is marked by a rectangle and the colored ellipse covers 67% of the samples belonging to the cluster. (**c**) Hierarchical clustering of the normalized metabolomics data using the Pearson product moment correlation coefficient.

the chosen approach for sample preservation, homogenization and cell lysis.

Quality and quantity of obtained biomolecular fractions

In addition to the need for efficient cell lysis, the most important consideration is the requirement for obtaining representative high-quality biomolecular fractions. As metabolites are extracted first, the methodology results in identical results for metabolomic analyses if subsequent biomacromolecular isolations are carried out or not. Metabolomic analysis of the representative LAO-enriched microbial community sample resulted in clear and reproducible total ion chromatograms (Supplementary Figure 3). These allowed the robust detection and semi-quantification of 267 polar and 242 non-polar intracellular metabolites as well as 268 polar and 176 non-polar extracellular metabolites (Supplementary Table 2). Using a spectral reference library, this enabled the combined identification of 62 polar and 30 non-polar metabolites. Consequently, the vast majority (>85%) of detected metabolites were not identified in line with earlier metabolomics results obtained on mixed microbial communities (Wilmes *et al.*, 2010).

The nucleic acid fractions obtained using the different extraction protocols were firstly analyzed by a NanoDrop spectrophotometer, the absorbance ratios at 260/280 nm in particular reflecting purity of the respective fractions obtained (Supplementary Table 3). The mean absorbance ratios at 260/280 nm for the DNA fractions are between 1.9 and 2.1 for the methods used, generally accepted as 'pure' (Manchester, 1995). The only exception is the DNA fraction obtained using the simultaneous biomolecular isolation using the TR-based method for

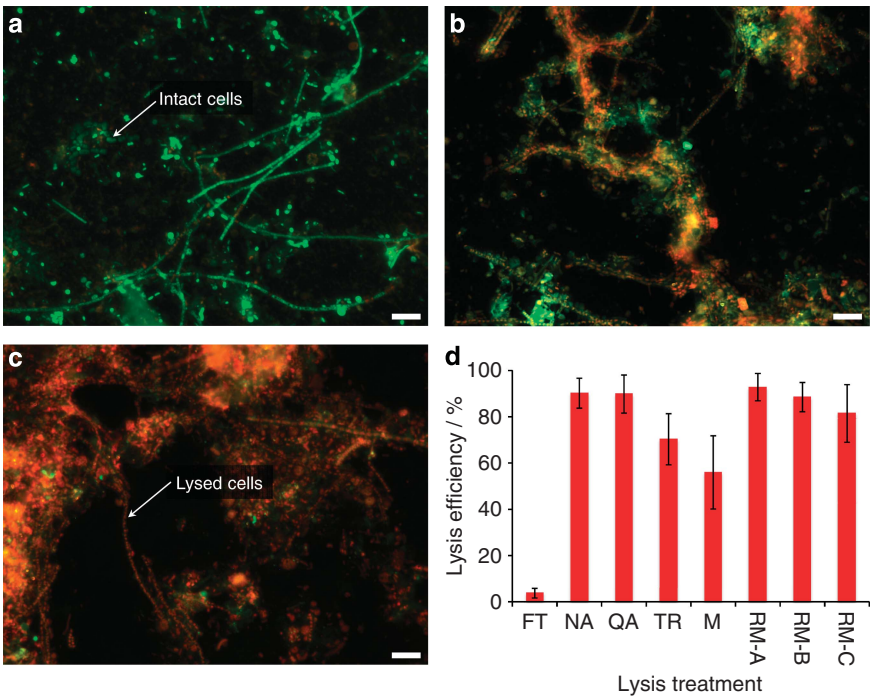

Figure 3 Efficiencies of different cell lysis methods. (**a–c**) Representative epifluorescence micrographs of microbial cells from a representative LAO-enriched microbial community sample stained with the Live/Dead stain (intact cells highlighted in green, disrupted cells with a compromised cell membrane in red). Scale bar is equivalent to 10 μm. (**a**) Sample having undergone a single freeze-thaw cycle. (**b**) Sample having undergone polar and non-polar metabolite extractions. (**c**) Sample having undergone the additional mechanical and chemical lysis step using the NA kit's modified lysis buffer. (**d**) Bar chart highlighting the percentages of cells disrupted by the different treatments ($n=10$; error bars represent s.d.). *x*-axis legend: FT, sample having undergone a single freeze-thaw cycle, reflecting panel **a**; NA, sample having been subjected to cell lysis in the modified NA lysis buffer, reflecting panel **c**; QA, QA lysis buffer; TR, TR-based lysis; M, sample after polar and non-polar metabolite extractions reflecting panel **b**; RM-A, sample having been subjected to the dedicated DNA extraction reference method; RM-B, dedicated RNA extraction method; RM-C, dedicated protein extraction method.

which a poor mean ratio of 1.5 was measured, highlighting the superior performance of the chromatographic spin column-based methods.

The absorbance ratios at 260/280 nm of all RNA fractions are between 1.8 and 2.1 (Supplementary Table 3), indicating that overall high-quality RNA was extracted using the different protocols. The Agilent Bioanalyzer 2100 electropherograms (Figure 4a) show distinct peaks between 100 and 4500 nt representing the total RNA fractions obtained using the different methods. The 23S/16S rRNA ratios vary depending on the RNA extraction method used from 0.7 to 1.5 (Supplementary Table 3). However, integrity of the isolated RNA was mainly assessed using the RNA integrity number (RIN) score, which is now commonly accepted as a better RNA quality indicator (Fleige and Pfaffl, 2006; Jahn *et al.*, 2008). The RIN scores obtained with the NA-based method (7.03 ± 1.20) are similar to those obtained using the Qiagen RNeasy Mini kit-based reference method (6.60 ± 0.88; Kruskal–Wallis, $P=0.806$, $n=5$) and the sequential biomolecular extraction based on the TR-based method (7.44 ± 0.28; Kruskal–Wallis, $P=0.352$, $n=5$) but lower than those obtained for the QA-based method (9.68 ± 0.05; Kruskal–Wallis, $P=0.011$, $n=5$), which had a very high and consistent score. The RNA fraction obtained using the TR-based method is particularly enriched in small RNAs, indicative of extensive RNA degradation resulting in a non-representative RNA fraction (Figure 4a). Overall, the RNA fractions derived using the simultaneous and sequential biomolecular isolation methods are of sufficient quality for downstream ribosomal RNA removal, reverse transcription and high-throughput cDNA sequencing (He *et al.*, 2010).

The NA-based method allows for the additional subfractionation of the total RNA extracted into a small RNA fraction (Figure 4b). As expected, the major components of the small RNA fractions are transfer RNAs with a mean size of around 60 nt and other small RNAs of larger size, represented by multiple peaks around 120 nt, for example, 5S rRNA. A 'miRNA'-like region is observed as a broad peak ranging from 10 to 40 nt (Figure 4b). The 'miRNA' region relative to small RNA content ('miRNA'/small RNA ratio) is $26.43 \pm 1.97\%$ for the NA-based method. Owing to the important regulatory roles fulfilled by small bacterial regulatory RNAs, the small RNA fraction, only provided by the NA-based method, adds an important additional level of information to the integrative molecular analyses that can be carried out using the reported methodology.

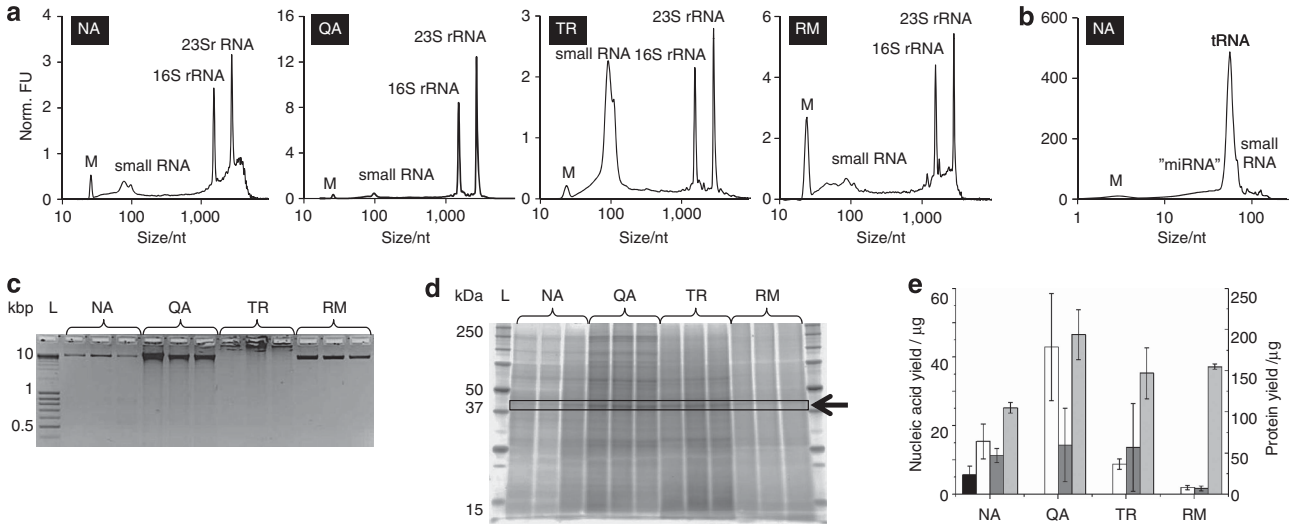

Figure 4 Quality and quantity of biomacromolecular fractions isolated from the representative LAO-enriched microbial community sample using either the NA-, QA- and TR-based method (following prior metabolite extractions) or using the reference methods (no metabolite extractions were carried out before the respective extractions). (**a** and **b**) Representative Agilent Bioanalyzer 2100 electropherograms of the total RNA and small RNA fractions, respectively. (**c**) Agarose gel image highlighting representative genomic DNA fractions obtained (Mean amount ($n=3$) loaded in µg ± s.d., from right to left; NA: 0.35 ± 0.17; QA: 0.35 ± 0.08; TR: 0.83 ± 0.85; RM: 0.08 ± 0.04) and (**d**) SDS-PAGE image of representative protein fractions (Mean amount ($n=3$) loaded in µg ± s.d., from right to left; NA, first elution: 3.20 ± 0.19; QA: 5.44 ± 1.06; TR: 3.88 ± 0.30; RM: 4.62 ± 0.09). The arrow and the box represent the dominant gel band which was submitted to tryptic digestion and MALDI-ToF/ToF analysis. (**e**) Biomacromolecular yield obtained for the ■ small RNA, □ total and large RNA, ■ DNA and ■ protein (first elution) fractions ($n=5$, error bars represent s.d.). FU, fluorescent unit; L, ladder; M, marker; NA, NA-based method; Norm., normalized; nt, nucleotides; RM, reference methods; TR, TR-based method; QA, QA-based method.

The size, the quality (degraded versus intact) and semi-quantitative amount of DNA extracted were determined by gel electrophoresis (Figure 4c). Rather expectably, the dedicated reference method results in the best DNA quality extract (Figure 4c, lane RM). The genomic DNA fraction isolated using the sequential biomolecular extraction protocol based on NA provides the most similar results to the reference method. Sequential biomolecular isolation based on QA also results in DNA extracts exhibiting intense and large bands. Importantly, the TR-based extraction method results in poor quality DNA extracts (Figure 4c), in concordance with the low-absorbance ratios measured at 260/280 nm (Supplementary Table 3) and, thus, is rather ill suited for comprehensive biomolecular isolations.

Gel electrophoresis of protein fractions (SDS-PAGE) provides a visual representation of the community metaproteomes derived from the representative LAO-enriched microbial community samples (Figure 4d). Importantly, in terms of band diversity and clarity, the efficiency of protein extraction is superior for the sequential and simultaneous isolation protocols (Figure 4d, lanes NA, QA and TR) compared to those obtained using the reference method (Figure 4d, lane RM). This is most likely due to the removal of 'contaminant' biomolecular fractions (for example, small molecules, nucleic acids) from the protein fraction during the sequential or simultaneous isolation methods. In contrast, some of these biomolecules may be retained in the extracts obtained with the reference method. In summary, the developed methodology results in higher quality protein extracts than the dedicated protein extraction method.

To further assess the compatibility of the protein fractions obtained with the NA-based method with subsequent downstream analyses, a protein band of ~45 kDa was excided from an SDS-PAGE gel (Figure 4d) and analyzed by MALDI-ToF/ToF following tryptic in-gel digestion. A *de novo* peptide sequence was derived from a clear tandem mass spectrum (Supplementary Figure 4). The peptide (*de novo* sequence: VESITAPVVVTEDQTQR) has been putatively identified as a possible cell surface protein (*Corynebacterium glucuronolyticum* spp.; GI: 227541404, e-value = 3.7, score = 31.6, BLASTP against NCBInr). Overall, these results highlight the ability to carry out proteomic analyses using either gel- or liquid chromatography-based separation followed by mass spectrometry on the obtained protein fractions.

In terms of yields, larger quantities of nucleic acids were obtained using the chromatographic spin column-based methods (NA and QA) compared with the reference methods (Kruskal–Wallis, $P=0.009$, $n=5$; Figures 4c and e and Supplementary Table 3). An analogous finding was previously highlighted in a comparative analysis of RNA yields obtained using a dedicated RNA extraction method and the simultaneous biomacromolecular extraction method based on TR (Chomczynski, 1993). In terms of protein quantity (Figures 4d

and e), the overall yield is highest for the QA-based method, followed by the exclusive isolation method. However, extensive variation in terms of the quantities of obtained biomolecular fractions is apparent for the QA-based isolation methodology. In contrast, the NA-based method results in the most consistent results for the tested sequential/simultaneous extraction methods.

Broad applicability of the methodological framework

The broad applicability of the developed methodology was demonstrated by applying the NA-based method to two additional mixed microbial community samples of environmental and biomedical research interest, that is, river water filtrate and human feces. Some minor sample-specific pretreatments, for example, RNAlater treatment of the fecal samples, were necessary to guarantee sufficient recovery of biomolecules. Analyses of the respective biomacromolecular fractions' quality and quantity resulted in no apparent differences compared with those obtained from the LAO-enriched microbial community (Kruskal–Wallis, $0.054 < P < 0.729$, $n = 5$, Figure 5, Supplementary Table 4 and Supplementary Table 5). To ascertain the quality of

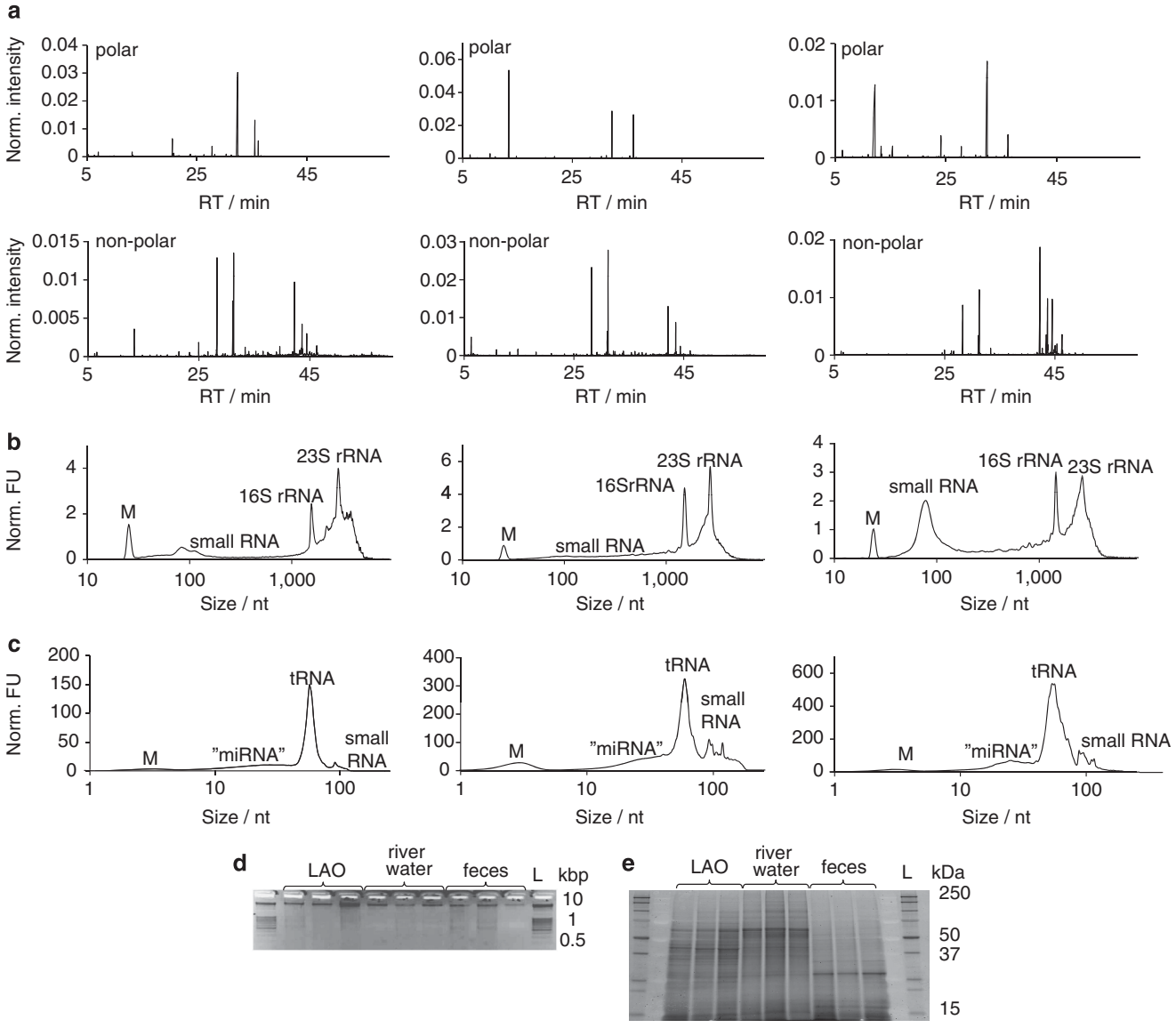

Figure 5 Application of the developed biomolecular isolation methodology to a LAO-enriched microbial community ($n = 3$), river water filtrate ($n = 3$) and human feces ($n = 3$). (**a–c**) Left-hand panels represent LAO-enriched microbial communities, middle panels represent river water filtrate and right-hand panels represent human feces. (**a**) Representative GC–MS total ion chromatograms of polar and non-polar metabolite fractions. Representative Agilent Bioanalyzer 2100 electropherograms of the (**b**) total RNA fractions and (**c**) small RNA fractions. (**d**) Agarose gel electrophoresis image of the genomic DNA fractions (Mean amount loaded in μg ± s.d., right to left; LAO: 0.63 ± 0.28; river water: 0.35 ± 0.03; feces: 0.61 ± 0.26) for each of the three technical replicates considered. (**e**) SDS-PAGE image of protein fractions, first elution (Mean amount loaded in μg ± s.d., right to left; LAO: 1.19 ± 0.40; river water: 1.70 ± 0.63; feces: 1.19 ± 1.21) for each of the three technical replicates considered. L, ladder; RT, retention time.

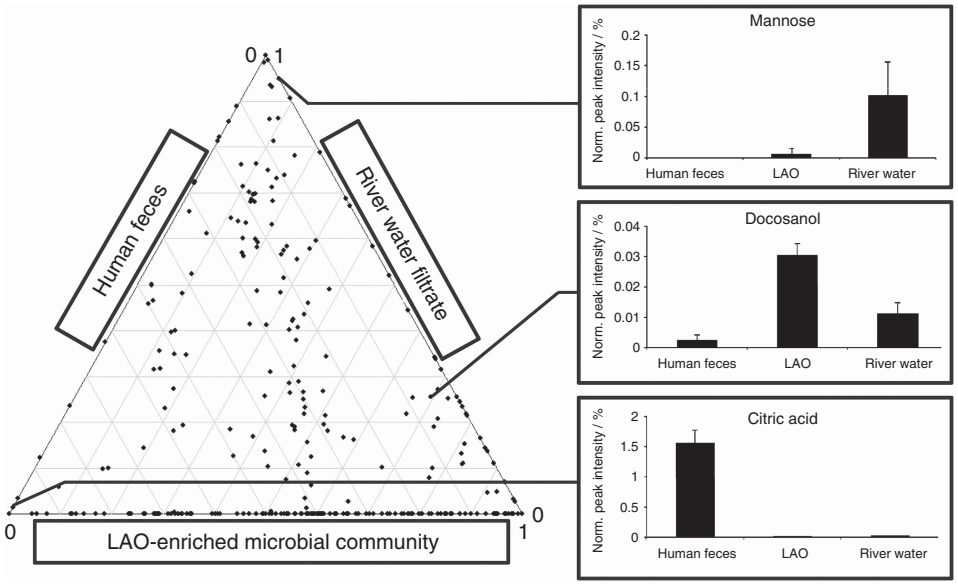

Figure 6 Three-way comparison of microbial community metabolomes obtained from a LAO-enriched microbial community, river water filtrate and human feces. Each dot represents the relative abundance of a metabolite in the three different environmental samples. Exemplary metabolite signatures for each microbiome are highlighted with their respective abundance levels on the right. Error bars represent s.d.

the obtained DNA for further downstream analyses, each DNA fraction from the three environmental samples was subjected to PCR amplification with primers targeting the 16S rRNA gene. For each microbial community DNA extract, clear and distinct bands with the correct molecular size were obtained (Supplementary Figure 5), highlighting the possibility for specialised downstream analyses without the need for additional purification steps.

Use of the developed standardized methodology further allows comparative analysis of different microbiomes. For example, it has not escaped our notice that an apparent enrichment in small RNA exists in human fecal samples (Figure 4b). Furthermore, to highlight potential signature metabolites for the three analyzed microbial communities, we contrasted the relative abundances of detected metabolites for each microbiome. For this, the metabolomic data sets from the respective microbiomes were subjected to a three-way comparative analysis (Supplementary Materials and methods). Clear differences in the metabolite composition of the three microbiomes are apparent, allowing the identification of specific signature metabolites for three microbiomes (Figure 6). For example, the enrichment of docosanol, a saturated fatty alcohol, in the LAO-enriched microbial community from the wastewater treatment plant may be explained by the wide use of this molecule as an emollient, emulsifier and thickener in cosmetics, nutritional and pharmaceutical products. Consequently, its presence may be expected in domestic wastewater and through its chemical properties would be enriched in LAOs. The pronounced enrichment of mannose, a simple sugar, in river water filtrate may be a direct result of photosynthesis by the dominant phototrophic organisms. Finally, a comparatively strong enrichment in citric acid was found in the human fecal samples. An increase of all metabolites involved in energy metabolism including citric acid was previously found in the serum of mice conventionally raised versus germ-free mice (Velagapudi *et al.*, 2010). Our present results indicate that this elevated level may directly result from the metabolic activities of microbial communities in the gastrointestinal tract.

By applying the NA-based workflow to these two additional mixed microbial communities, we demonstrate that the methodological framework can be applied to a range of different samples and, thus, represents a standardized biomolecular isolation procedure for comparative eco-systems biology investigations on a range of different microbial communities in the future.

Discussion

With the advent of high-throughput omic technologies, powerful and sensitive methods are available for the analysis of nucleic acids (DNA and RNA), proteins and small molecules obtained from biological samples including microbial communities (Jansson *et al.*, 2012). Molecular eco-systems biology investigations, centered around the integration of omic data from microbial communities, are faced with major challenges arising from the complexity, dynamics and heterogeneity of microbial consortia. In the present work, we demonstrate that microbial community metabolomes do not allow certain biological and technical replicates to be discriminated, which is a direct consequence of the extensive

heterogeneity encountered in these systems. Consequently, major artifacts may be introduced into coupled high-resolution omic experiments by sample splitting before dedicated isolation of the respective individual biomolecular fractions. Such approaches will result in disparate omic results and, thus, will not fulfill the premise of systematic measurements. Furthermore, they will not allow meaningful *in silico* reconstructions and modeling of community-wide processes and interactions.

Here, we describe a methodological framework that allows for the sequential extraction and purification of all known biomolecular fractions from single unique samples. The methodology is based on the combination of chromatographic spin column-based biomacromolecular isolation with prior extraction of concomitant polar and non-polar metabolite fractions. We demonstrate that the devised methodology yields equivalent or even better results compared with those obtained with dedicated extraction protocols, and that it is applicable to a range of different microbial community samples. Furthermore, the methodology may prove successful on a range of other biological samples, in particular those that are precious or which are characterized by extensive within-sample heterogeneity, for example, human tumor samples. The utilization of spin column chromatography for biomacromolecular separation minimizes the risk of exposure to harmful chemicals and, thus, allows routine laboratory use. Finally, the methodological framework is potentially compatible with sequential biomacromolecular isolation methods other than those used in the present work, for example, the GE Healthcare illustra triplePrep kit.

The reported methodological framework provides the basis for conducting molecular eco-systems biology investigations on a range of different mixed microbial communities in the future. The methodology will facilitate meaningful data integration, analysis and modeling of metabolomic, transcriptomic, genomic and proteomic data. Furthermore, it will allow exploitation of the inherent heterogeneity through spatial and temporal sampling of microbial communities to allow deconvolution of community-wide, population-wide and individual-level processes using statistical approaches. Consequently, it will allow the determination of associations between distinct biomolecules that may provide pointers towards the discovery of previously unknown metabolic processes linked to specific community members, a need which is highlighted by the relatively large number of uncharacterized metabolites detected in the metabolomic experiments carried out as part of the present work. Finally, by providing a standardized workflow, the methodology lays the foundation for future comparative molecular studies of different microbial ecosystems leading to further discovery of functional microbiome-specific biomolecular signatures.

Acknowledgements

The scientists and technical staff of the Centre de Recherche Public Gabriel-Lippmann, particularly Dr Christian Penny, Cécile Walczak and Sébastien Planchon, and from the Luxembourg Centre for Systems Biomedicine, particularly Dr Pranjul Shah and Laura Lebrun, are thanked for their assistance and support. Mr Bissen and Mr Di Pentima from the Syndicat Intercommunal à Vocation Ecologique (SIVEC) are thanked for their permission to collect samples and gain access to the monitoring platform of the Schifflange wastewater treatment plant. Rudi Balling and Anders Andersson are thanked for helpful comments on the manuscript. The present work was supported by an ATTRACT programme grant to PW (ATTRACT/A09/03) and Aide à la Formation Recherche (AFR) grants to HR (PHD-MARP-04) and EM (PRD-2011-1/SR), all funded by the Luxembourg National Research Fund (FNR).

Author contributions

PW and HR designed the research; HR, PW, TC and KH planned the experiments; HR, EM, PW, JR and TC carried out the analyses; HR, PW and EM analyzed the data; and HR, EM and PW wrote the manuscript.

References

Chey S, Claus C, Liebert UG. (2011). Improved method for simultaneous isolation of proteins and nucleic acids. *Anal Biochem* **411**: 164–166.

Chomczynski P. (1993). A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques* **15**: 532–536.

Denef VJ, Mueller RS, Banfield JF. (2010). AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. *ISME J* **4**: 599–610.

Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N *et al.* (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* **6**: 1060–1083.

Fischer CR, Wilmes P, Bowen BP, Northen TR, Banfield JF. (2011). Deuterium-exchange metabolomics reveals N-methyl lyso phosphatidylethanolamines as abundant lipids in acidophilic mixed microbial communities. *Metabolomics* advance online publication doi:10.1007/s11306-011-0344-x.

Fleige S, Pfaffl MW. (2006). RNA integrity and the effect on the real-time qRT-PCR performance. *Mol Aspects Med* **27**: 126–139.

Florens L, Carozza MJ, Swanson SK, Fournier M, Coleman MK, Workman JL *et al.* (2006). Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **40**: 303–311.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al.* (2008). Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci* **105**: 3805–3810.

He S, Wurtzel O, Singh K, Froula JL, Yilmaz S, Tringe SG *et al.* (2010). Validation of two ribosomal RNA removal

methods for microbial metatranscriptomics. *Nat Meth* **7**: 807–812.

Hummon AB, Lim SR, Difilippantonio MJ, Ried T. (2007). Isolation and solubilization of proteins after TRIzol extraction of RNA and DNA from patient material following prolonged storage. *Biotechniques* **42**: 467–472.

Jahn CE, Charkowski AO, Willis DK. (2008). Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *J Microbiol Methods* **75**: 318–324.

Jansson JK, Neufeld JD, Moran MA, Gilbert JA. (2012). Omics for understanding microbial functional dynamics. *Environ Microbiol* **14**: 1–3.

Kitano H. (2001). *Foundations of Systems Biology*. The MIT Press: Cambridge, Massachusetts, London, England.

Levantesi C, Rossetti S, Thelen K, Kragelund C, Krooneman J, Eikelboom D *et al.* (2006). Phylogeny, physiology and distribution of '*Candidatus* Microthrix calida', a new *Microthrix* species isolated from industrial activated sludge wastewater treatment plants. *Environ Microbiol* **8**: 1552–1563.

Li M, Wang B, Zhang M, Rantalainen M, Wang S, Zhou H *et al.* (2008). Symbiotic gut microbes modulate human metabolic phenotypes. *Proc Natl Acad Sci* **105**: 2117–2122.

Lin LI. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45**: 255–268.

Manchester KL. (1995). Value of A_{260}/A_{280} ratios for measurement of purity of nucleic acids. *Biotechniques* **19**: 208–210.

Morse S, Shaw G, Larner S. (2006). Concurrent mRNA and protein extraction from the same experimental sample using a commercially available column-based RNA preparation kit. *Biotechniques* **40**: 54–58.

Radpour R, Sikora M, Grussenmeyer T, Kohler C, Barekati Z, Holzgreve W *et al.* (2009). Simultaneous isolation of DNA, RNA, and proteins for genetic, epigenetic, transcriptomic, and proteomic analysis. *J Proteome Res* **8**: 5264–5274.

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

Ram RJ, VerBerkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC II *et al.* (2005). Community proteomics of a natural microbial biofilm. *Science* **308**: 1915–1920.

Tolosa JM, Schjenken1 JE, Civiti TD, Clifton VL, Smith R. (2007). Column-based method to simultaneously extract DNA, RNA, and proteins from the same sample. *Biotechniques* **43**: 799–804.

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM *et al.* (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37–43.

Velagapudi VR, Hezaveh R, Reigstad CS, Gopalacharyulu P, Yetukuri L, Islam S *et al.* (2010). The gut microbiota modulates host energy and lipid metabolism in mice. *J Lipid Res* **51**: 1101–1112.

Weckwerth W, Wenzel K, Fiehn O. (2004). Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics* **4**: 78–83.

Wilmes P, Bond PL. (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ Microbiol* **6**: 911–920.

Wilmes P, Bowen BP, Thomas BC, Mueller RS, Denef VJ, VerBerkmoes NC *et al.* (2010). Metabolome-proteome differentiation coupled to microbial divergence. *mBio* **1**: e00246–00210.

Wilmes P, Simmons SL, Denef VJ, Banfield JF. (2009). The dynamic genetic repertoire of microbial communities. *FEMS Microbiol Rev* **33**: 109–132.

Xiong J, Yang Q, Kang J, Sun Y, Zhang T, Margaret G *et al.* (2011). Simultaneous isolation of DNA, RNA, and protein from *Medicago truncatula* L. *Electrophoresis* **32**: 321–330.

This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)

CHAPTER ELEVEN

Sequential Isolation of Metabolites, RNA, DNA, and Proteins from the Same Unique Sample

Hugo Roume, Anna Heintz-Buschart, Emilie EL Muller, Paul Wilmes[1]
Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
[1]Corresponding author: e-mail address: paul.wilmes@uni.lu

Contents

Abstract

In microbial ecology, high-resolution molecular approaches are essential for characterizing the vast organismal and functional diversity and understanding the interaction of microbial communities with biotic and abiotic environmental factors. Integrated omics, comprising genomics, transcriptomics, proteomics, and metabolomics allows

Methods in Enzymology, Volume 531
ISSN 0076-6879
http://dx.doi.org/10.1016/B978-0-12-407863-5.00011-3

© 2013 Elsevier Inc.
All rights reserved.

conclusive links to be drawn between genetic potential and function. However, this requires truly systematic measurements. In this chapter, we first assess the levels of heterogeneity within mixed microbial communities, thereby demonstrating the need for analyzing biomolecular fractions obtained from a single and undivided sample to facilitate multi-omic analysis and meaningful data integration. Further, we describe a methodological workflow for the reproducible isolation of concomitant metabolites, RNA (optionally split into large and small RNA fractions), DNA, and proteins. Depending on the nature of the sample, the methodology comprises different (pre)processing and preservation steps. If possible, extracellular polar and nonpolar metabolites may first be extracted from cell supernatants using organic solvents. Cells are homogenized by cryo-milling before small molecules are extracted with organic solvents. After cell lysis, nucleic acids and protein fractions are sequentially isolated using chromatographic spin columns. To prove the broad applicability of the methodology, we applied it to microbial consortia of biotechnological (biological wastewater treatment biomass), environmental (freshwater planktonic communities), and biomedical (human fecal sample) research interest. The methodological framework should be applicable to other microbial communities as well as other biological samples with a minimum of tailoring and represents an important first step in standardization for the emerging field of Molecular Eco-Systems Biology.

1. INTRODUCTION

1.1. From multi-omics to integrated omics: The necessity for systematic measurements

The field of microbial ecology has advanced rapidly since the introduction of molecular biology methodologies, which have allowed the exploration of microbial diversity and function in a multitude of different environments. Over the past two decades, a wide variety of techniques have been developed to describe and characterize microbial assemblages. More recently, revolutionary improvements and advances in high-throughput "omic" technologies, ranging from metagenomics to metabolomics, are allowing resolution of community- and population-level processes from genetic potential to final phenotype. However, in order to collect such high-resolution data in a truly systematic fashion and facilitate meaningful data integration and analysis, robust sampling, sample preservation, and biomolecular isolation methodologies are essential.

In *Foundation of Systems Biology*, Kitano defined the ideal systematic measurement as the "simultaneous measurement of multiple features for a single sample" (Kitano, 2001). This consideration is particularly important in ecosystems biology as a vast microbial diversity is a hallmark of natural settings

(Caporaso et al., 2010) and even within-population genetic heterogeneity is very pronounced in natural microbial communities (Wilmes, Simmons, Denef, & Banfield, 2009). To assess the extent of heterogeneity at each biomolecular information layer, a comparative experiment was carried out on islet forming spatially resolved lipid-accumulating organisms (LAO)-enriched microbial community samples from a biological wastewater treatment plant (Roume et al., 2013). We measured the phylogenetic composition using 454 pyrosequencing of 16S rDNA amplicons (Muller et al., in preparation), the functional genetic potential using an Illumina HiSeq platform (Pinel et al., in preparation), and the community-wide metabolic profile using gas chromatography coupled to mass spectrometry (GC–MS; Roume et al., 2013). We generated data of the concomitant biomolecular fractions for four biological replicates (i.e., space resolved LAO islets) and four technical replicates of one single islet (i.e., different subsamples derived from the same islet after sample splitting), sampled at the same point in time (Fig. 11.1). Extensive sample-to-sample variation is apparent at each level of biomolecular information. Interestingly, the ability to discriminate between biological and technical replicates decreases from phylogenetic profiles (Fig. 11.1A) to metagenomes (Fig. 11.1B) and lastly metabolite profiles (Fig. 11.1C). This finding reflects

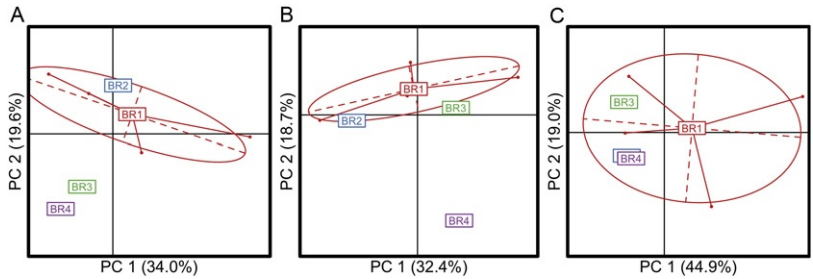

Figure 11.1 Within- and between-sample heterogeneity as determined from a biological wastewater treatment plant microbial community. Scatter plots of the results of principal coordinate analysis of Bray–Curtis dissimilarity indices of four biological replicates (BR1–BR4), from a unique sampling date, including one biological replicate (BR1) split into four technical replicates. Each biological replicate is indicated by a rectangle, and each technical replicate is represented by a dot. The ellipse covers two-thirds of the technical replicates, which are derived from the same biological replicate, with the rectangle indicating their common center of gravity. (A) Community compositions of the samples as revealed by 16S rDNA amplicon sequencing (Muller et al., in preparation), (B) corresponding functional capacities determined using KEGG orthologous group annotations of metagenomic sequence data (Pinel et al., in preparation), and (C) corresponding polar metabolomes (Roume et al., 2013).

increased complexity at each biomolecular information layer. The resulting fine-scale heterogeneity underlines the requirement for the isolation of concomitant biomolecules to fulfill the premise of systematic measurements and enable meaningful integration of space- and time-resolved microbial community omic datasets.

1.2. Methods for the isolation of concomitant biomolecules

Previously published methodologies for the simultaneous isolation of certain biomacromolecules, namely DNA, RNA, and proteins, from biological samples are primarily based on monophasic mixtures of water, phenol, and guanidine isothiocyanate (Chomczynski & Sacchi, 1987). The simultaneous extraction of metabolites, proteins, and RNA has previously been carried out on plant material (Weckwerth, Wenzel, & Fiehn, 2004). For safe routine laboratory use, chromatographic spin column-based methods have been developed for the combined extraction of RNA and proteins from a variety of cells and tissues (Morse, Shaw, & Larner, 2006). The additional isolation of DNA from tissues and cultured cells is the primary feature of the Qiagen AllPrep method (Tolosa, Schjenken, Civiti, Clifton, & Smith, 2007), and this has since been validated on an extensive set of medical samples (Radpour et al., 2009). Since then, several additional manufacturers, for example, GE Healthcare, Macherey-Nagel, Norgen Biotek, etc., have commercialized chromatographic spin column-based kits for the sequential isolation of DNA, RNA, and protein extraction from tissues or cultured cells.

The detailed protocols presented here provide a unique methodological framework for the reproducible isolation of high-quality genomic DNA, large and small RNA or total RNA, and proteins, as well as polar and nonpolar metabolites from single and undivided samples. This chapter focuses on biological wastewater treatment biomass, planktonic freshwater, and human fecal mixed microbial communities, which are representative of the differing sample characteristics one may encounter. However, the protocols can also be successfully applied to other samples with minor preprocessing modifications. Examples include frozen human feces, methanogenic reactor digestates, and diverse mouse tissue samples.

2. SAMPLING AND SAMPLE PREPROCESSING

2.1. Sample fixation

When collecting biological samples for systematic molecular analyses, it is crucial to immediately fix the samples in a state which best preserves the

information contained within the system at the time of sampling. If care is not given to this step, the information contained within the DNA, RNA, protein, and metabolites will rapidly change due to specific and nonspecific degradation, regulation of expression, and protein modifications. Sample fixation directly after sampling is therefore a key step in the experimental workflows presented for the three different microbial communities described below:

Biomass from a biological wastewater treatment plant (LAOs)

- Collect sample in a 50 ml tube and immediately snap freeze on-site in liquid nitrogen.

Freshwater filtrate

- Collect approximately 40 l of freshwater and preserve the sample at 4 °C. Proceed immediately to the preprocessing step.

Human fecal sample

- Cover 300 mg ± 10% of fresh human fecal sample immediately with 1.5 ml of RNAlater solution (Ambion) and carry on to the preprocessing step.

2.2. Preprocessing: Dealing with differing sample characteristics

Sample preprocessing is an important step, which varies significantly depending on the overall characteristics of samples (Fig. 11.2A). If possible, this step consists of the separation of the intra- and extracellular compartments and accompanying concentration of cells.

2.2.1 Biological wastewater treatment plant biomass or similar samples

For a sample which allows recovery of the extracellular medium, the sample preprocessing begins with slowly thawing the sample on ice to allow the subsequent separation of cells from the supernatant (extracellular compartment) by centrifugation. The detailed preprocessing protocol involves the following:

1. With a sterile spatula, reduce to powder around 200 mg ±10% of frozen LAO sample. To conserve the frozen state of the powder, cool down all equipment and refreeze the material regularly by briefly dipping it into liquid nitrogen. Float a weighing boat on the surface of the liquid nitrogen and pour the powdered sludge into this.
2. Transfer the frozen powder into a precooled 2 ml sample tube and place this into a liquid nitrogen bath. Optionally, it can be stored for few days at −80 °C until further processing.

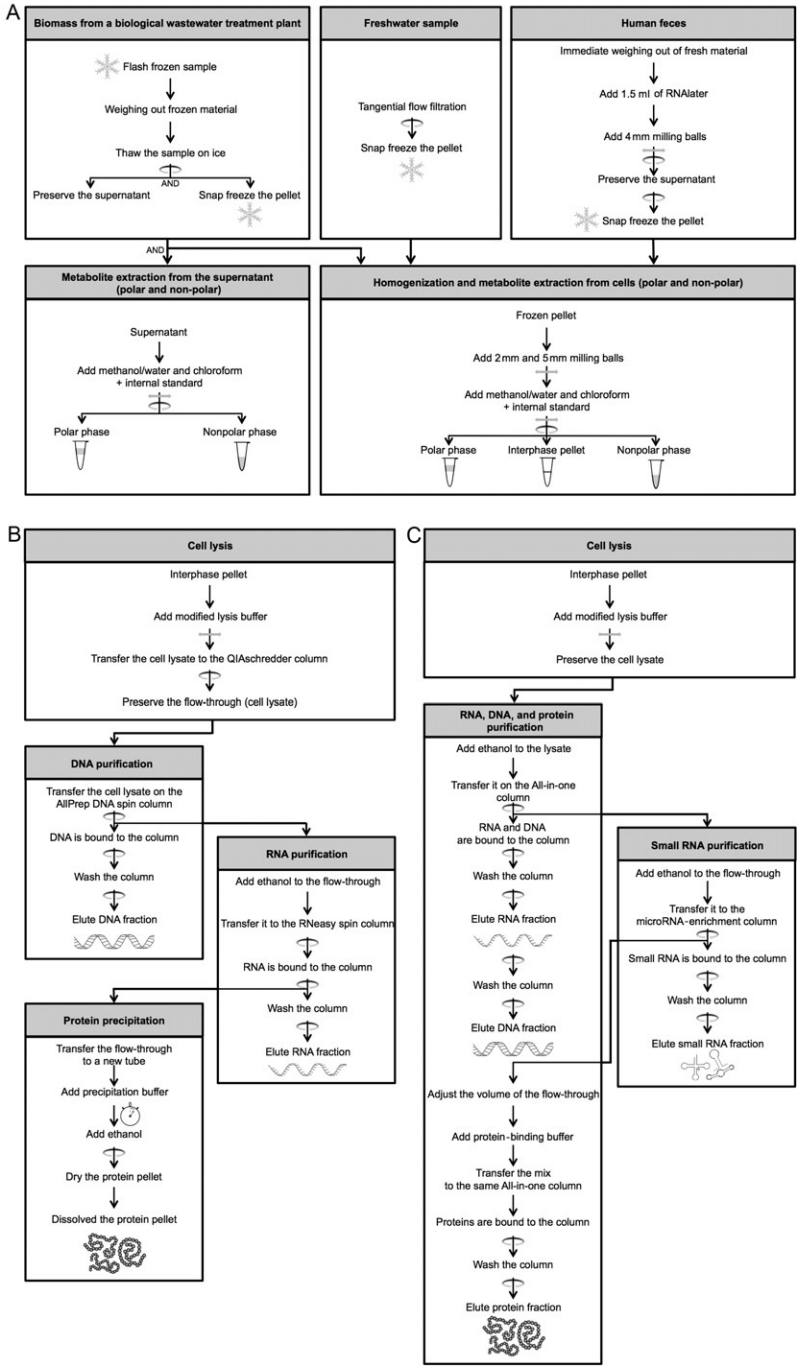

Figure 11.2 Synopsis of the described methodological workflow. (A) Preprocessing and metabolite extraction; (B) isolation of DNA, total RNA, and proteins using the Qiagen AllPrep-based procedure; and (C) isolation of DNA, small and large RNA, and protein using the Norgen Biotek All-in-One-based procedure. The circular arrows denote a centrifugation step, and the horizontal arrows symbolize a mix or bead beating step.

3. Thaw the sample on ice for approximately 10–15 min by placing the sample tube horizontally on the ice surface. Flip the tube regularly and verify the consistency of the sample every 5 min. It is highly recommended to place the thawed sample into ice immediately after the sample is completely thawed.
4. Centrifuge the thawed sample at 18,000 × *g* for 15 min at 4 °C and transfer the supernatant into a new 2 ml sample tube kept at 4 °C. Around 150 µl of supernatant for 200 mg of starting material may be expected. If visible material particles are still present in the supernatant, carry out an additional centrifugation for 5 min until a completely limpid supernatant is obtained. Proceed to Section 3.1 with the supernatant.
5. Snap freeze the cell pellet and preserve it at −80 °C until step 2 of Section 3.2.

2.2.2 Freshwater planktonic microbial communities

The preprocessing of a liquid sample with a low density of biomass consists of first concentrating the cells present within the sample before snap freezing:

1. Concentrate cells by tangential flow filtration using, for example, Sartocon silica cassettes (Satorius Stedim Biotek) with a filtration area of 0.1 m^2 and molecular weight cutoff of 10 kDa at a flow rate of ~1.5 l min^{-1}.
2. Pellet the concentrated cells by ultracentrifugation at 48,400 × *g* for 1 h at 4 °C.
3. Resuspend the resulting pellet in 1 ml of supernatant and transfer it to a 2 ml centrifugation tube.
4. Further concentrate the suspension by an additional centrifugation step at 14,000 × *g* for 5 min at 4 °C.
5. Snap freeze the resulting pellet and preserve it at −80 °C until step 2 of Section 3.2.

2.2.3 Human fecal sample

For a complex sample such as fresh feces, in which biomolecules are highly sensitive to degradation and which are rich in PCR inhibitors, the preprocessing step involves treatment of the sample to preserve the integrity of the biomolecules, especially RNA (Fitzsimons, Akkermans, de Vos, & Vaughan, 2003). The optimized workflow for fecal samples is thus as follows:

1. Add three autoclaved stainless steel milling balls with a diameter of 4 mm to the tube of fresh fecal sample overlaid with RNAlater.

2. Homogenize the mixture by cold bead beating in a Mixer Mill 400 (Retsch) for 5 min at 10 Hz. For this, the rack holding the sample tube needs to be precooled at 4 °C.
3. Centrifuge the homogenized sample at 700 × *g* for 1 min at 4 °C. This low-speed centrifugation step should result in a supernatant containing detached cells. Transfer the supernatant into a fresh 2-ml sample tube.
4. Pellet the cells contained within the supernatant at 14,000 × *g* for 5 min at 4 °C. Discard the supernatant.
5. Snap freeze the resulting pellet and preserve it at −80 °C until step 2 of Section 3.2.

3. SAMPLE PROCESSING AND SMALL MOLECULE EXTRACTION

Depending on the physical characteristics of the sample and the preprocessing, metabolite extraction can be done separately on supernatant and/or on cells (Fig. 11.2A).

3.1. Extraction of extracellular metabolites from cell supernatant

For metabolite extraction, the polar and the nonpolar solvent fractions should be present in equal volumes. The method for cold solvent extracting metabolites from the supernatant of a wastewater treatment plant biomass sample is as follows:

1. Add to one volume of the aqueous supernatant obtained in Section 2.2.1 at step 4, one volume (e.g., 150 μl) of cold methanol at −20 °C and two volumes of cold chloroform at −20 °C (e.g., 300 μl).
2. Mix the sample tube in a thermomixer (e.g., Thermomixer comfort, Eppendorf) for 30 min at 4 °C at maximum speed (e.g., 1400 rpm).
3. Centrifuge the tubes at 14,000 × *g* and 4 °C for 5 min.
4. Transfer the polar and nonpolar phases into separate 2 ml sample tubes. If required, the different phases can be preserved for a few hours at −20 °C at this step.

For a description of processing the polar and nonpolar metabolite fractions for subsequent GC–MS analyses, please see Roume et al. (2013).

3.2. Sample homogenization by cryomilling

The cryomilling (or cryogenic grinding) step involves breaking up the structure of the biological material and allows homogenization of the sample. The

frozen homogenous powder obtained at the end of this step guarantees a sufficiently large surface area for the solvents to extract the small molecules. The processing, described below, is of fundamental importance for efficient metabolite extractions from cellular biomass and subsequent representative metabolic profiling thereof (Lolo et al., 2007).

1. Cool autoclaved-sterilized milling balls (5×2 mm $+ 2 \times 5$ mm) by dipping the tubes containing them in liquid nitrogen.
2. Add the cold milling balls to the previously obtained frozen cell pellet (Sections 2.2.1, 2.2.2, or 2.2.3 at step 5).
3. Cryomill the frozen pellet for 2 min at 25 Hz in a Mixer Mill 400 (Retsch). The adaptor rack holding the sample tube needs to be precooled in liquid nitrogen to –80 °C prior to the milling step.
4. Dip the tube immediately into liquid nitrogen to preserve the sample in a frozen state.

At the end of the cryomilling step, the sample should comprise a frozen homogenous powder. If it is not the case, repeat the cryomilling step or empirically optimize the number and size of stainless steel milling balls according to the nature of the sample.

3.3. Extraction of intracellular metabolites from cell pellet

1. Add sequentially 300 µl of cold methanol/water (1:1, v/v) and 300 µl of cold chloroform. Vortex the solution until complete dissolution of the powdered sample into the solvent mixture.
2. Mill the sample tube in a Mixer Mill 400 (Retsch) for 2 min at 20 Hz. In order to avoid solvent leakage, wrap parafilm around the rim of the cap of the sample tube before starting the milling to ensure a tight seal.
3. Centrifuge the sample tube at $14{,}000 \times g$ and 4 °C for 5 min to separate the solvent mixture into a polar (top) phase, an interphase pellet (middle), and nonpolar (lower) phase.
4. Transfer the polar and nonpolar phases into new 2 ml sample tubes. If required, the different phases can be preserved for a few hours at −20 °C at this step.
5. Preserve the interphase with the milling beads in the original centrifugation tube on ice (until proceeding with step 1 of Section 4.1 or 4.2).

For an example of the processing of polar and nonpolar metabolite fractions for subsequent GC–MS analysis, please see Roume et al. (2013).

4. SEQUENTIAL BIOMACROMOLECULAR ISOLATION

Biomacromolecules, including large, small, or total RNA, DNA, and proteins, are isolated from the interphase pellet obtained following the metabolite extraction step (Section 3.3, step 5). Two sequential biomacromolecular isolation protocols, modified from commercially available kits, are described here. The Qiagen AllPrep DNA/RNA/Protein Mini Kit-based method (Fig. 11.2B) allows for sequential extraction of total RNA, DNA, and protein. The Norgen Biotek All-in-One Purification Kit-based method (Fig. 11.2C) facilitates the sequential extraction of large RNA, small RNA, DNA, and proteins. Although these two kit-based methods for biomacromolecular purification are described here, application of alternative methods to the interphase pellet should be entirely feasible but may require some minor adjustments.

Cell disruption, comprising both sample homogenization and cell lysis, is an early and fundamental step in any biomolecular isolation methodology. Both chemical and mechanical/physical methods are available for cell disruption. However, in natural microbial communities, due to the presence of many different microorganisms with vastly differing cellular properties and due to the presence of different interfering matrixes, chemical and/or enzymatic lysis by themselves are typically ineffective at comprehensive and reproducible cell lysis. In the methodology presented here, we combine the mechanical method of cryogenic grinding with chemical lysis to result in indiscriminate cell lysis.

The efficiency of cell lysis should be assessed for each new type of sample (Section 5.1). Furthermore, following sequential extraction, an aliquot of 5 μl for small and large or total RNA and 10 μl for DNA should be stored at 4 °C for quality assessment (Sections 5.3.1 and 5.3.2). These steps are essential to guarantee the efficiency of cell lysis as well as the quality of the obtained biomolecular fractions.

4.1. Qiagen AllPrep DNA/RNA/Protein Mini Kit-based method

The interphase pellet is overlaid with lysis buffer before bead beating with the same milling balls as used previously. Importantly, a modified lysis buffer is used to prevent RNA degradation. Following addition of the modified lysis buffer, the sample is subjected to shearing of high-molecular weight cellular components in order to reduce the viscosity of the lysates, to improve

the binding efficiency of nucleic acids to the chromatographic spin columns, and thereby to increase the overall yield and purity of the obtained biomolecular fractions. The methodology is summarized in Fig. 11.2B and is as follows:

1. Add 10 µl of β-mercaptoethanol per milliliter of RLT buffer (Qiagen).
2. Add 600 µl of cold (4 °C) modified RLT buffer (step 1) to the interphase pellet (step 5 of Section 3.3) and cover the rim of the closed tube cap with parafilm.
3. Resuspend the interphase in the modified lysis buffer by a quick vortexing of the sample tube.
4. Proceed to a bead beating in a Mixer Mill 400 (Retsch) for 30 s at 25 Hz. Note that the adaptor racks should be at 4 °C.
5. Transfer up to 700 µl of the lysate to a QIAshredder column (Qiagen) and centrifuge for 2 min at 12,100 × *g*. The entire lysate should pass through the QIAshredder column. If a pellet forms in the collection tube, it should be resuspended before loading the lysate onto the next column.

All further steps are described in the AllPrep DNA/RNA/Protein Mini Handbook (Qiagen, version 12/2007) in the section "Simultaneous purification of genomic DNA, total RNA, and total protein from animal and human cells" p. 22 step 4. In our own experience, the elution of DNA and total RNA in the dedicated buffers can be repeated in order to recover more nucleic acids.

Importantly, if the lysate has not passed through the AllPrep DNA spin column completely after 5 min of centrifugation at the maximum speed, transfer the liquid retained in the column to an additional AllPrep DNA column, which should then be processed along with the first tube. Both AllPrep DNA columns should be used for the subsequent DNA recovery steps.

4.2. Norgen Biotek All-in-One Purification Kit-based method

Similar to the Qiagen AllPrep-based method, the interphase pellet is resuspended in a modified lysis buffer, which prevents RNA degradation. Cell disruption is carried out in this buffer using the milling balls still present in the centrifugation tube following step 5 of Section 3.3. A few essential modifications to the manufacturer's instructions have been implemented and are detailed below. The workflow is also summarized in Fig. 11.2C.

1. Add 400 µl of cold lysis buffer supplemented with 10 µl ml^{-1} of β-mercaptoethanol and 100 µl of cold 1× Tris–EDTA buffer

(pH 7.2) to the interphase pellet. Cover the rim of the closed tube cap with parafilm.
2. Resuspend the pellet by quickly vortexing the sample tube.
3. Carry out cell lysis by bead beating the sample in the Mixer Mill 400 (Retsch) for 30 s at 25 Hz in a cold rack (4 °C).
4. Add 100 µl of pure ethanol to the lysate and mix by vortexing for 10 s.

In our own experience, increasing the proportion of ethanol in the lysate prior to loading of the column results in a higher yield of nucleic acids recovery.

Follow the procedure of the All-in-One Purification Kit protocol from (Norgen Biotek, 2008; PI24200-12) section 3: "Isolation of large RNA, Genomic DNA, microRNA, and Proteins" p. 17 3A.1.b.

Importantly, and as recommended by the manufacturer for bacterial starting material, two elutions of large RNA, DNA, and protein from the All-in-One column should be carried out in order to maximize the yield. We typically also carry out the column-based procedure for total protein purification.

5. QUALITY CONTROL

The different methods presented in this section are typically used in our laboratory to assess the efficiency of the protocol when applying it to new samples and to assess the quality of the recovered biomolecular fractions.

5.1. Cell lysis

Conservation of cell integrity until sample processing as well as representative cell lysis prior to biomolecular extraction are essential considerations for the methodology to result in reproducible and representative biomolecular fractions (Roume et al., 2013). The proportion of cells which are intact before and after cell disruption may be assessed using the Live/Dead *Bac*Light Bacterial Viability Kit (L7012, Invitrogen). Figure 11.3 shows that, in the case of LAO samples from wastewater, most cells in the snap-frozen sample still have intact cell membranes after 10 min of thawing on ice (Fig. 11.3A) and that the vast majority of the cells are lysed after cryomilling, metabolite extraction, and combined mechanical/chemical lysis in the modified Norgen Biotek lysis buffer (Fig. 11.3B). To assess the lysis efficiency:

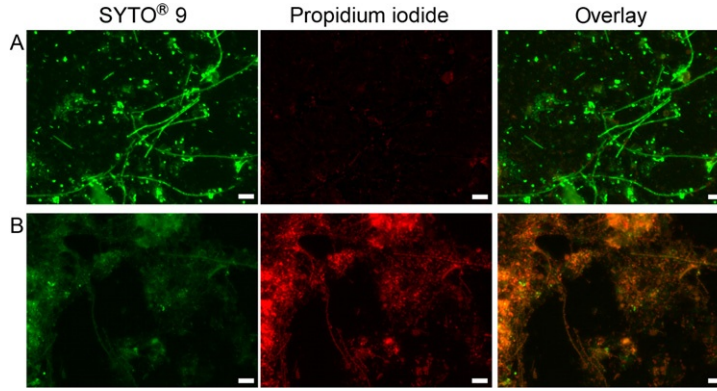

Figure 11.3 Cell integrity before and after cell lysis. Representative epifluorescent micrographs of microbial cells from a representative biological wastewater treatment plant sample stained with the Live/Dead stain. Propidium iodide specifically stains bacteria with a damaged cell membrane, whereas Syto 9 serves as the counterstain. Intact cells are highlighted in green and disrupted cells in red. Scale bar is equivalent to 10 μm. (A) Sample having undergone a single freeze–thaw cycle and (B) sample having undergone subsequent metabolite extraction followed by mechanical and chemical lysis using the Norgen Biotek All-in-One Kit's modified lysis buffer.

1. Pellet cellular material from the individual protocol steps for which you want to assess cell membrane integrity by centrifugation at 14,000 × *g* and 4 °C for 5 min.
2. Wash the pellet once with phosphate buffered saline solution (1 × PBS, pH 7) and redilute into 1 ml of 1 × PBS buffer.
3. Follow the manufacturer's instructions of the Live/Dead *Bac*Light Bacterial Viability Kit (Molecular Probes, rev. July 15, 2004, p. 3).

For the determination of the red and green pixel ratio, multiple red and green fluorescence micrographs can be processed as described in Roume et al. (2013).

5.2. Metabolite fractions

Quality control in metabolomics is a challenging task. The metabolites have to be extracted without compromising either their chemical structure or their relative abundance. Some metabolites are extremely unstable and therefore prone to degradation during sample preparation. Several

recommendations should be considered, from the sampling to the metabolomic analyses, to ensure the generation of reliable data.

5.2.1 Sampling and sample preprocessing

- The sample should be processed immediately or snap frozen in liquid nitrogen right away on site to stabilize the sample and stop any enzymatic activity.
- The sample should be preserved at a minimum of −80 °C, ideally −150 °C.
- The sample should be homogenized by cryomilling to facilitate efficient metabolite extraction from solid sample.

5.2.2 Validation of metabolomic analyses

- In the case of GC–MS, the analytical procedure usually needs to be optimized for each sample type with the fine tuning of several parameters, including, for example, the volume of metabolite fraction to be injected, the split or the oven program. The derivatization procedure may also need to be adapted.
- The polar and/or the nonpolar solvents used for the extraction should be supplemented with exogenous chemical compounds which should not be present endogenously within the sample and can serve as internal standards. During metabolomic analysis, the internal standard is a useful benchmark to assess metabolite recovery as well as the reproducibility of the analysis.
- When large numbers of samples are analyzed in a single analytical run, pool samples, which are mixtures of aliquots of all the samples included in the run and which, thus, reflect the entirety of metabolites detectable, should be included in the sample analysis sequence at regular intervals. For example, we typically include a pool sample as every seventh sample in a sequence, followed by a blank sample. The measurement results obtained from the pool samples can provide an indication of possible instrument drift during the run. Importantly, the pool samples, as well as the internal standards, can later also be used to normalize the data (Roume et al., 2013).

The metabolomics data obtained by GC–MS can be displayed as total ion chromatograms (TICs). We routinely assess the quality of the TICs to get an indication of the purity and representativeness of the metabolite extracts. For example, the metabolomic analysis of the three different samples, for

which the extractions are described, resulted in clear TICs comprised of numerous small and well-defined peaks (Fig. 11.4A).

5.3. Nucleic acids

5.3.1 Ribonucleic acid

Because RNA is a biomacromolecule prone to degradation, the accurate assessment of its integrity is one of the most critical steps for the success of any downstream analysis (Vermeulen et al., 2011), including ribosomal RNA removal, reverse transcription, and high-throughput cDNA sequencing (RNA-Seq). A commonly accepted large and total RNA quality and integrity indicator is Agilent's RNA integrity number (RIN; Fleige & Pfaffl, 2006). A RIN $\geq$7.0 ususally satisfies accepted quality requirements for most RNA-Seq protocols (Jahn, Charkowski, & Willis, 2008) and ensures a low contamination of the small RNA fraction with degradation products. If a sample exhibits a RIN of $<$7.0, we recommend not to use all other biomolecular fractions obtained from this sample and to carry out an additional extraction.

As shown in the representative electropherograms displayed in Fig. 11.4B and C, the extraction method described here facilitates the isolation of high quality and pure RNA. The large RNA fractions of the three different sample types were analyzed using the Agilent 2100 Bioanalyzer with the RNA 6000 Nano RNA Kit and 2100 Expert software (Fig. 11.4B). In order to compare different RNA fractions, fluorescent units of the traces were normalized (Roume et al., 2013). The traces exhibit distinct peaks between 100 and 4,500 nt. The two major peaks, at around 1,500 and 2,900 nt, show distinctively the 16S and the 23S rRNA, respectively. The broad peak around 100 nt indicates the small RNA fraction.

The electropherograms of small RNA traces (Fig. 11.4C), obtained using the Agilent RNA 6000 Small RNA Kit, are dominated by peaks around 60 nt representative of the transfer RNA (tRNA). Other small RNAs are represented by multiple peaks around 120 nt, including 5S rRNA. In eukaryotic systems, the "miRNA"-like region is typically defined as the broad peak ranging from 10 to 40 nt.

5.3.2 Deoxyribonucleic acid

The size, the quality (degraded vs. intact), and semiquantitative amount of DNA extracted can be determined by agarose gel electrophoresis as highlighted in Fig. 11.4D. The methodology described herein is able to yield DNA which exhibits intense and high-molecular weight DNA bands

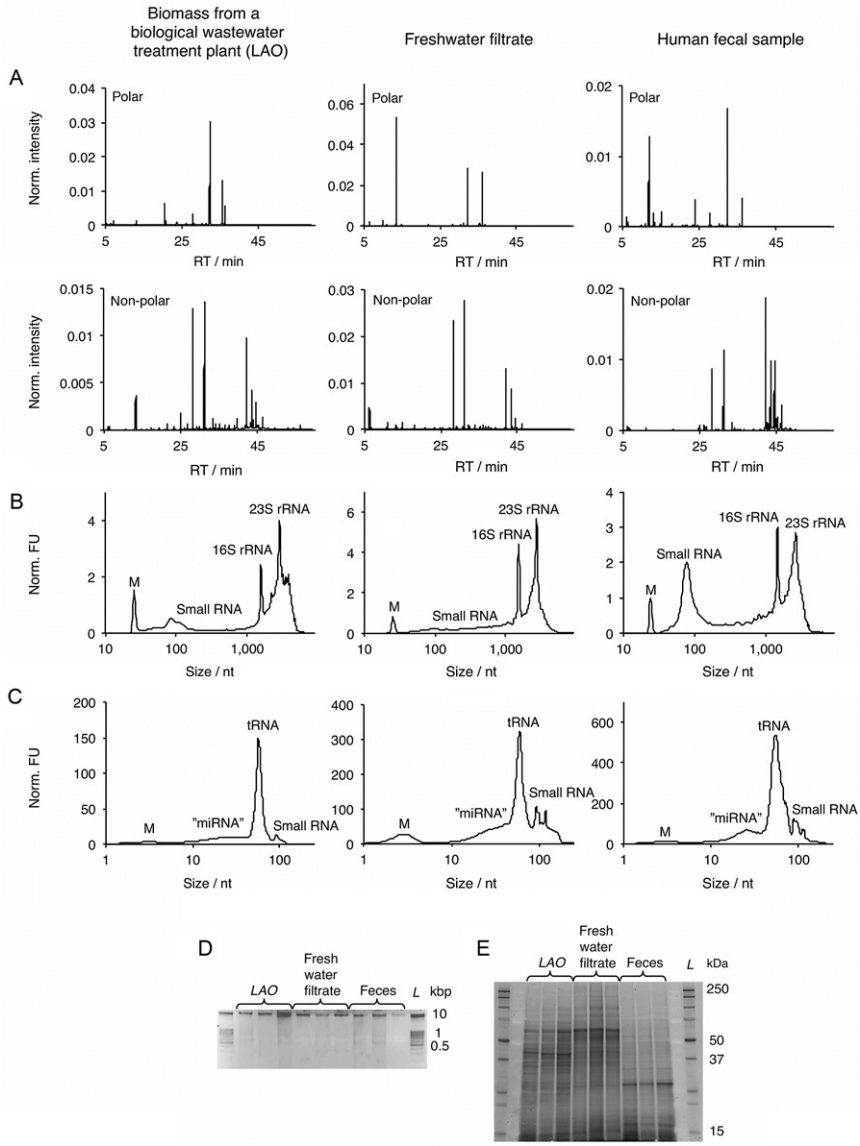

Figure 11.4 Representative quality control results for biomolecular fractions obtained using the Norgen Biotek All-in-One Purification Kit-based biomolecular isolation methodology when applied to a biological wastewater treatment sample (left hand panes), river water filtrate (middle panes), and human fecal samples (right hand panes). (A) Representative GC–MS total ion chromatograms of polar and nonpolar metabolite fractions. (B) and (C) Representative Agilent Bioanalyzer 2100 electropherograms of the obtained total RNA fractions and small RNA fractions, respectively. (D) Agarose gel electrophoresis image of genomic DNA fractions for three technical replicates for each of the samples. (E) SDS-PAGE image of the first protein elution for three technical replicates for each of the samples. Norm, normalized; FU, fluorescent unit; M, marker; L, ladder; nt, nucleotides; RT, retention time. *Figure reproduced from Roume et al. (2013) with permission from Nature Publishing Group.*

and can be directly used for library construction for high-throughput sequencing.

5.4. Proteins

The quality of obtained protein fractions is typically assessed by SDS-PAGE. The obtained gel should exhibit clear bands, as highlighted in Fig. 11.4E for the three different samples for which the biomolecular extraction protocol is described herein. The gel obtained can also be directly used for proteomics either by analysis of individually excised bands or entire lanes.

6. OUTLOOK

In microbial ecology, an increasing number of studies aim to integrate multi-omic datasets in order to obtain comprehensive high-resolution overview of microbial community structure and function. Considering recent technological improvements and the accompanying decrease in the cost of high-throughput molecular methods as well as associated progress in computational and statistical methodologies, these types of investigations will dramatically intensify in the coming years. As demonstrated in the present work, sample-to-sample variation is extensive for microbial communities at each biomolecular level. Consequently, the sequential isolation of biomolecules from an undivided sample is a clear prerequisite for meaningful multi-omic data integration and analysis (Muller, Glaab, May, Vlassis, & Wilmes, 2013). Application of the methodology presented here will allow truly systematic measurement of microbial consortia and will allow the deconvolution of microbial community-driven processes in unprecedented detail. Consequently, the described methodolgy and future iterations thereof provide the backbone of the emerging field of Eco-Systems Biology.

REFERENCES

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2010). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 4516–4522.

Chomczynski, P., & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*, *162*(1), 156–159.

Fitzsimons, N. A., Akkermans, A. D. L., de Vos, W. M., & Vaughan, E. E. (2003). Bacterial gene expression detected in human faeces by reverse transcription-PCR. *Journal of Microbiological Methods*, *55*(1), 133–140.

Fleige, S., & Pfaffl, M. W. (2006). RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine*, *27*(2–3), 126–139.

Jahn, C. E., Charkowski, A. O., & Willis, D. K. (2008). Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *Journal of Microbiological Methods*, *75*(2), 318–324.

Kitano, H. (2001). Systems biology: Toward system-level understanding of biological systems. In H. Kitano (Ed.), *Foundations of systems biology* (pp. 1–36). Cambridge, MA: The MIT Press.

Lolo, M., Pedreira, S., Vázquez, B., Franco, C., Cepeda, A., & Fente, C. (2007). Cryogenic grinding pre-treatment improves extraction efficiency of fluoroquinolones for HPLC-MS/MS determination in animal tissue. *Analytical and Bioanalytical Chemistry*, *387*(5), 1933–1937.

Morse, S., Shaw, G., & Larner, S. (2006). Concurrent mRNA and protein extraction from the same experimental sample using a commercially available column-based RNA preparation kit. *Biotechniques*, *40*(1), 54–58.

Muller, E. E. L., Glaab, E., May, P., Vlassis, N., & Wilmes, P. (2013). Condensing the omics fog of microbial communities. *Trends in Microbiology*, *21*, 325–333.

Muller, E. E. L., Pinel, N., Laczny, C., Roume, H., Lebrun, L., Hussong, R., et al. Integrated omics reveals that ecological success of a generalist is linked to fine-tuning of resource usage (in preparation).

Pinel, N., Muller, E. E. L., Narayanasamy, S., Laczny, C., Roume, H., Hussong, R., et al. Ecological dominance in the light of time-resolved, multi-omic community characterization (in preparation).

Radpour, R., Sikora, M., Grussenmeyer, T., Kohler, C., Barekati, Z., & Holzgreve, W. (2009). Simultaneous isolation of DNA, RNA, and proteins for genetic, epigenetic, transcriptomic, and proteomic analysis. *Journal of Proteome Research*, *8*(11), 5264–5274.

Roume, H., Muller, E. E. L., Cordes, T., Renaut, J., Hiller, K., & Wilmes, P. (2013). A biomolecular isolation framework for eco-systems biology. *The ISME Journal*, *7*(1), 110–121.

Tolosa, J. M., Schjenken, J. E., Civiti, T. D., Clifton, V. L., & Smith, R. (2007). Column-based method to simultaneously extract DNA, RNA, and proteins from the same sample. *Biotechniques*, *43*(6), 799–804.

Vermeulen, J., De Preter, K., Lefever, S., Nuytens, J., De Vloed, F., Derveaux, S., et al. (2011). Measurable impact of RNA quality on gene expression results from quantitative PCR. *Nucleic Acids Research*, *39*(9), e63.

Weckwerth, W., Wenzel, K., & Fiehn, O. (2004). Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics*, *4*(1), 78–83.

Wilmes, P., Simmons, S. L., Denef, V. J., & Banfield, J. F. (2009). The dynamic genetic repertoire of microbial communities. *FEMS Microbiology Reviews*, *33*(1), 109–132.