

# The effect of different organic substrates on the microbial communities of aerobic granular wastewater treatment sludge

**Thèse N° 9678**

Présentée le 30 août 2019

à la Faculté de l'environnement naturel, architectural et construit  
Laboratoire de biotechnologie environnementale  
Programme doctoral en génie civil et environnement

pour l'obtention du grade de Docteur ès Sciences

par

**Aline Sondra ADLER**

Acceptée sur proposition du jury

Prof. K. Schirmer, présidente du jury  
Prof. C. Holliger, directeur de thèse  
Prof. P. Junier, rapporteuse  
Prof. P. Halhjaer Nielsen, rapporteur  
Prof. R. Bernier-Latmani, rapporteuse

2019



# Acknowledgements

Cette thèse est le fruit de 4 années et 8 mois d'apprentissages, d'expériences au laboratoire et de traitement statistique et bioinformatique des données collectées. Elle est également le résultat de plusieurs collaborations et du travail de personnes qui ont activement contribué à ce projet. Je souhaite tout particulièrement remercier :

Mes superviseurs,

Prof Christof Holliger, pour m'avoir offert cette superbe opportunité, ainsi que la liberté et le support nécessaires pour réaliser ce travail de thèse dans les meilleures conditions. Pour l'organisation de collaborations avec d'autres laboratoires. Pour l'implication et la bienveillance.

Jacques Rougemont pour l'aide pertinente apportée en bioinformatique en début de thèse.

Les membres du jury,

Prof Kristin Schirmer, présidente du jury, l'examen s'est déroulé à merveille.

Prof Rizlan Bernier-Latmani, pour son implication en tant que membre du jury de thèse, pour ses remarques et questions pertinentes.

Prof Pilar Junier, pour son investissement en tant que membre du jury pour les examens de thèse et de candidature à la fin de la première année, pour ses remarques et questions pertinentes.

Prof Per Halkjaer Nielsen, pour son implication en tant que membre du jury de thèse, pour ses questions et remarques pertinentes. Pour m'avoir accueillie dans son laboratoire à Aalborg et donné l'opportunité de discuter et collaborer avec les membres de son équipe. J'ai beaucoup appris à cette occasion.

## Acknowledgements

---

Mes prédécesseurs, Samuel Lochmatter et David Weissbrodt qui ont posé d'excellentes bases sur lesquelles démarrer ce projet.

Les collègues,

Mathilde Willemin, Marie Vingerhoets, Arnaud Gelb et Lorenzo Cimmino pour le temps investi et les précieux conseils lors des préparations des présentations. Pour les bons moments.

Julien Maillard pour les relectures, les conseils pointus et le travail sur les différentes méthodes d'extractions d'ADN. Pour les goûts culturels alternatifs.

Marc Deront pour sa gestion efficace des soucis informatiques.

Filomena Jacquier pour l'organisation des voyages et événements, pour la gentillesse et les cakes.

Idriss Hendaoui pour la maintenance des bioréacteurs et la collecte de données, pour son enthousiasme et sa gentillesse.

L'équipe technique,

Stéphane Marquis pour l'enthousiasme scientifique, l'aide pratique et ingénieuse pour la construction et le bon fonctionnement des bioréacteurs.

Emmanuelle Rohrbach pour l'immense travail réalisé pour la préparation des échantillons pour le séquençage d'amplicons des expériences de l'EPFL et de l'Eawag.

Emmy Oppliger pour la maintenance des bioréacteurs et la collecte de données, pour la bonne humeur contagieuse.

Romain Savary et Dylan Magnin pour l'aide à la préparation des échantillons pour le séquençage.

Les étudiants master,

Valérie Berclaz pour la maintenance des bioréacteurs, la collecte de données et les analyses sur la précision des mesures. Pour la présence sympathique et la gentillesse.

Marie Horisberger pour la maintenance des bioréacteurs, la collecte de données et les expériences ingénieuses. Pour la présence sympathique et la bonne humeur.

Eva Reynaert pour la maintenance des bioréacteurs et la collecte de données à l'Eawag, pour les discussions inspirantes.

David Scheibler pour la maintenance des bioréacteurs et la collecte des données, pour la présence sympathique.

Antonio Hernandez pour la maintenance et la collecte de données à l'Eawag.

L'équipe du Department Process Engineering (Eawag) avec laquelle nous avons eu une collaboration intéressante, instructive et fructueuse, le résultat de cette collaboration constitue le chapitre III de cette thèse :

Manuel Layer pour toutes les données et les échantillons collectés à l'Eawag, pour la patience, la ténacité et la gentillesse.

Nicolas Derlon pour le soutien, l'engagement dans le projet et la sympathie.

Eberhard Morgenroth pour l'expertise, l'engagement dans le projet et les remarques pertinentes.

Le Swiss Institute of Bioinformatics,

Marco Pagni pour ses conseils inspirés en matière de bioinformatique et analyse de données, pour l'enthousiasme.

Emanuel Schmid-Siegert pour le temps gagné grâce aux conseils sur les analyses PacBio.

Robin Engler pour l'installation de programmes et les dépannages sur le cluster.

Le Center for Microbial Communities d'Aalborg,

Marta Nierychlo pour l'accueil sympathique à Aalborg et les discussions intéressantes.

Erika Yashiro pour son aide en bioinformatique et les discussions intéressantes.

Eustace Fernando pour les révélations en avant premières sur *Tetrasphaera* et *Candidatus Accumulibacter*.

## Acknowledgements

---

Le Microbial Population Biology Group,

Otto Cordero pour l'accueil dans le labo de Zürich, les discussions intéressantes et les précieux conseils.

Ela Sliwerska pour le séquençage des granules individuelles.

Les colocataires,

Michel pour l'aide à la préparation des présentations, le soutien moral et la bonne humeur.

Benoît pour les délicieuses quiches, le soutien moral et la sympathie.

Åsa et Matt pour les encouragements et les boîtes de nourriture salutaires.

Toute l'équipe du Viet Vo Dao, qui a été une bulle d'air indispensable pendant ce travail de thèse.

Les amis,

Vincent pour les illustrations désopilantes.

Floortje pour l'aide avec Latex, les discussions sérieuses, les moins sérieuses et les bons moments.

Tous les autres amis pour le soutien moral et les bons moments, ils ont été précieux.

La famille, pour le soutien moral, la patience et les bons moments, en particulier,

Denise pour avoir été un modèle inspirant, pour le soutien nutritionnel et moral.

Jeanne pour la gentillesse.

Mes soeurs pour les encouragements et les bons repas.

# Abstract

Aerobic granular sludge (AGS) is a promising alternative wastewater treatment to the conventional activated sludge system, allowing space and energy savings. This process is particularly suited for biological phosphorus removal, avoiding use of coagulant chemicals. Basic understanding of this process has mainly been obtained in laboratory-scale studies with simple synthetic wastewater containing volatile fatty acids as main carbon source. Yet, the aspect and performance of granular sludge cultivated in such model systems are rather different from those obtained in systems treating real wastewater.

In order to make a step toward the comprehension of AGS treating municipal wastewater, two approaches were applied to investigate the impact of the wastewater composition on AGS bacterial communities, settling properties and nutrient removal performance. The first approach was to transform activated sludge performing enhanced biological phosphorus removal into AGS in four parallel lab-scale reactors fed with different wastewater types: simple and complex polymeric synthetic, as well as raw and clarified municipal wastewater. The second approach was to progressively change the composition of the wastewater treated by AGS acclimated to simple synthetic wastewater to complex polymeric wastewater with an intermediary step with complex monomeric wastewater as a control. The whole DNA of biomass corresponding to the three wastewater types was sequenced with PacBio and Illumina technologies. During the two experiments, the bacterial communities, settling properties and nutrient removal performance were monitored.

The bacterial communities in AGS treating simple wastewater were drastically different from the ones treating complex wastewater. Several taxa belonging to Actinobacteria and Saccharibacteria were largely underrepresented with the simple wastewater. Lower nitrogen removal, lower settling properties and higher proportions of flocs were observed with the polymeric wastewaters compared to the monomeric wastewaters. The lower concentrations of diffusible organic carbon rather than the bacterial community compositions were identified as the cause for these differences. Indeed, genes putatively involved in denitrification and biofilm formation were found in the AGS treating monomeric and polymeric wastewater. Moreover, different denitrification efficiencies and settling properties were observed with AGS having very similar bacterial communities but treating wastewater with a different concentration of organic carbon.

The phosphate accumulating organism (PAO) *Candidatus* (*Ca.*) *Accumulibacter*, abundant in most of the AGS samples, was highly diverse and different clade repartitions were found within the AGS treating different wastewater types. The fermentative PAO *Tetrasphaera* was

## Abstract

---

less diverse and mostly found in the AGS treating complex wastewater. The co-occurrence of two PAO occupying distinct ecological niches likely participated to the quick recovery of the P-removal after the transient but sharp decrease of *Ca. Accumulibacter* observed during the transition from simple to complex monomeric wastewater and likely due to a bacteriophage attack.

The assembly of PacBio sequences associated to a binning based on composition and coverage of Illumina sequences produced nearly complete draft genomes attributed to poorly characterized taxa, thus providing information on their potential metabolism and a template for future metatranscriptomic analysis.

Keywords : Aerobic granular sludge, biological wastewater treatment, bacterial communities, phosphate accumulating organisms, glycogen accumulating organisms, biological phosphorus removal, nitrogen removal, metagenome-assembled genomes.

## Résumé

Les boues granulaires aérobie (AGS) constituent une alternative prometteuse aux boues activées utilisées pour traiter les eaux usées, permettant des économies d'énergie et de surface au sol. Ce procédé est bien adapté pour l'élimination biologique du phosphate, permettant d'éviter l'utilisation de coagulants chimiques. Une compréhension élémentaire de ce procédé a été établie principalement grâce à des études réalisées avec des eaux usées synthétiques simples contenant des acides gras volatiles comme principale source de carbone. Cependant, l'aspect et les performances des boues granulaires cultivées dans de tels systèmes sont différents de ceux obtenus dans des systèmes traitant des eaux usées réelles.

Deux approches ont été mises en oeuvre pour étudier l'impact de la composition de l'eau usée sur les communautés bactériennes, les propriétés de sédimentation et les performances d'élimination des nutriments des AGS. La première approche consistait à transformer des boues activées capables de déphosphatation biologique en AGS, dans quatre réacteurs alimentés par des eaux usées différentes : synthétiques simple et complexe polymérique, ainsi que municipales brute et clarifiée. La deuxième approche consistait à changer progressivement la composition de l'eau usée traitée par des AGS acclimatées à une eau simple vers une eau complexe polymérique, en passant par une eau complexe monomérique. L'ADN total de la biomasse traitant les trois types d'eau a été séquencé à l'aide des technologies PacBio et Illumina.

Les communautés bactériennes des AGS traitant des eaux simples étaient très différentes de celles traitant des eaux complexes. Plusieurs taxa appartenant aux Actinobacteria et Saccharibacteria étaient largement sous-représentés dans les échantillons traitant l'eau simple. Une plus faible élimination de l'azote, une moins bonne sédimentation et une plus grande proportion de floccs ont été observées avec les eaux polymériques. De plus basses concentrations de carbone organique diffusible, plutôt que les compositions des communautés bactériennes, ont été identifiées comme cause de ces différences. En effet, des gènes potentiellement impliqués dans la dénitrification et la formation de biofilms ont été trouvés dans les AGS traitant les deux types d'eau. De plus, des rendements de dénitrification et des propriétés de sédimentation différents ont été observés avec des communautés bactériennes similaires, traitant des eaux contenant des concentrations de carbone organique différentes.

L'organisme accumulant du phosphate (PAO) *Candidatus (Ca.) Accumulibacter*, abondant dans la plupart des échantillons, était très diversifié et une répartition différente des clades a été trouvée suivant le type d'eau traitée. Son pendant fermentatif, *Tetrasphaera*, était moins diversifié et a été principalement trouvé dans les AGS traitant les eaux complexes. La co-

## Résumé

---

occurrence de ces deux PAO occupant des niches écologiques différentes a probablement participé à la rapide récupération de l'élimination du phosphate après la diminution de *Ca. Accumulibacter* observée durant la transition de l'eau simple à l'eau complexe monomérique. L'assemblage des séquences PacBio associé à un binning reposant sur la composition des séquences et la couverture des séquences Illumina a permis de reconstituer des génomes attribués à des taxa peu connus, fournissant ainsi des informations sur leurs métabolismes potentiels ainsi qu'une base pour de futures analyses métatranscriptomiques.

Mots-clés : boues granulaires aérobie, traitement de l'eau biologique, communautés bactériennes, organismes accumulant le phosphate, organismes accumulant du glycogène, élimination biologique du phosphate, élimination de l'azote, génomes assemblés à partir de métagénomes.

# Contents

<b>Acknowledgements</b>	<b>iii</b>
<b>Abstract (English/Français)</b>	<b>vii</b>
<b>List of figures</b>	<b>xxi</b>
<b>List of tables</b>	<b>xxv</b>
<b>List of acronyms</b>	<b>xxv</b>
<b>I Introduction</b>	<b>3</b>
I.1 Aerobic granular sludge for wastewater treatment . . . . .	3
I.1.1 Wastewater treatment . . . . .	3
I.1.2 Aerobic granular sludge . . . . .	4
I.1.3 Research on aerobic granular sludge . . . . .	5
I.1.4 Bacterial communities in aerobic granular sludge . . . . .	6
I.1.5 Nitrogen removal . . . . .	8
I.1.6 Biofilm, extracellular polymeric substances and quorum-sensing . . . . .	10
I.1.7 Studying aerobic granular sludge through molecular methods . . . . .	11
I.2 Research questions and thesis overview . . . . .	11
<b>II High variability of microbial community composition among individual granules</b>	<b>15</b>
II.1 Introduction . . . . .	15
II.2 Materials and methods . . . . .	16
II.2.1 Sampling and sequencing . . . . .	16
II.2.2 Pre-processing of the sequences . . . . .	16
II.2.3 Estimation of the relative abundance of the different bacteria . . . . .	16
II.2.4 Estimation of the variation of the mean abundance of <i>Ca. Accumulibacter</i> depending on the number of granules considered . . . . .	16
II.2.5 Correlation calculations between bacteria and phages . . . . .	17
II.3 Results . . . . .	17
II.3.1 Bacterial community composition of individual granules . . . . .	17
II.3.2 Estimation of the variability of <i>Ca. Accumulibacter</i> . . . . .	18
II.3.3 Correlations between phages and bacteria . . . . .	20
II.4 Discussion . . . . .	22

## Contents

---

II.5 Conclusion . . . . .	23
<b>III Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater</b>	<b>27</b>
III.1 Preamble . . . . .	27
III.1.1 List of authors . . . . .	27
III.1.2 Authors contributions . . . . .	27
III.2 Introduction . . . . .	28
III.3 Materials and methods . . . . .	31
III.3.1 Experimental approach . . . . .	31
III.3.2 Experimental set-up . . . . .	31
III.3.3 Start-up approach . . . . .	31
III.3.4 Successful start-up definition . . . . .	32
III.3.5 Wastewater composition and sludge inoculum . . . . .	32
III.3.6 Physical sludge parameters . . . . .	33
III.3.7 Analytical methods . . . . .	34
III.3.8 Microbial community analysis . . . . .	34
III.4 Results . . . . .	36
III.4.1 Settling properties . . . . .	36
III.4.2 Sludge size fractions . . . . .	36
III.4.3 Evolution of the bacterial community composition from inoculation to stable state . . . . .	38
III.4.4 Bacterial communities of granules and flocs . . . . .	44
III.4.5 Nutrient removal performance . . . . .	45
III.4.6 Correlations between settling properties, nutrient-removal performances and microbial community composition . . . . .	46
III.4.7 Identification of correlations between the discriminant taxa and the sludge size distributions, the settling properties, and the nutrient-removal performances . . . . .	48
III.5 Discussion . . . . .	51
III.5.1 Diffusibility of organic substrates has significant influence on formation of AGS . . . . .	51
III.5.2 Start-up . . . . .	52
III.5.3 Stable state . . . . .	53
III.5.4 The role of flocs in AGS systems . . . . .	55
III.5.5 Implications for research and practice . . . . .	56
III.6 Conclusions . . . . .	57
<b>IV Transforming acetate-propionate-grown granules into AGS treating fermentable and polymeric organic substrates</b>	<b>61</b>
IV.1 Introduction . . . . .	61
IV.1.1 Effect of polymeric compounds on AGS . . . . .	61
IV.1.2 Resilience, resistance and multistability . . . . .	62

IV.1.3	The role of flocs in AGS . . . . .	63
IV.1.4	Hydrolysis in AGS . . . . .	63
IV.1.5	Objectives . . . . .	64
IV.2	Material and methods . . . . .	65
IV.2.1	Reactors operation . . . . .	65
IV.2.2	Wastewater composition . . . . .	67
IV.2.3	Nutrient-removal performance monitoring . . . . .	67
IV.2.4	Measures of SVI, TS and VS . . . . .	68
IV.2.5	Nematodes monitoring and controlling . . . . .	68
IV.2.6	Biomass sampling . . . . .	68
IV.2.7	DNA extraction and 16S rRNA gene sequencing . . . . .	69
IV.2.8	Taxonomic affiliation of 16S rRNA gene sequences . . . . .	70
IV.2.9	Statistical analysis . . . . .	70
IV.2.10	Determination of stable states and discriminant taxa . . . . .	70
IV.2.11	Comparison of the bacterial communities in flocs and granules . . . . .	70
IV.3	Results . . . . .	71
IV.3.1	Nutrient-removal performance . . . . .	71
IV.3.2	Settling properties . . . . .	73
IV.3.3	Evolution of the bacterial community composition during the transition from simple to complex monomeric and complex polymeric wastewater	79
IV.3.4	Stable states and discriminant taxa . . . . .	82
IV.3.5	Changes in the microbial communities during the transition to a complex polymeric and back to simple wastewater . . . . .	85
IV.3.6	Comparison of the bacterial communities in flocs and granules . . . . .	86
IV.3.7	Correlation between the bacterial communities and the nutrient removal performances . . . . .	89
IV.3.8	Multiple factor analysis . . . . .	91
IV.4	Discussion . . . . .	91
IV.4.1	General observations . . . . .	91
IV.4.2	Transition from simple to complex monomeric wastewater . . . . .	91
IV.4.3	Transition from complex monomeric to polymeric wastewater . . . . .	93
IV.4.4	Reversibility of the changes induced in the bacterial communities . . . . .	94
IV.4.5	Multiple stable states as the result of multistability? . . . . .	95
IV.4.6	The microbial communities in flocs and granules . . . . .	96
IV.5	Conclusion . . . . .	96
<b>V</b>	<b>Metagenomic analysis of different AGS samples</b>	<b>99</b>
V.1	Introduction . . . . .	99
V.1.1	The use of molecular techniques to study microbial communities . . . . .	99
V.1.2	Grouping DNA sequences according to their phylotype remains challenging	100
V.1.3	Long-read sequencing technologies facilitates the assembly of complete genomes . . . . .	101

## Contents

---

V.1.4	Denitrification is performed by microorganism of various phylotypes . . .	101
V.1.5	Phosphorus cycling related genes . . . . .	102
V.1.6	<i>Ca. Accumulibacter</i> is a highly diversified genus . . . . .	102
V.1.7	Objectives . . . . .	103
V.2	Material and methods . . . . .	103
V.2.1	Biomass sample collection . . . . .	103
V.2.2	Metagenomic DNA extraction from AGS biomass . . . . .	104
V.2.3	DNA sequencing . . . . .	105
V.2.4	Assemblies of PacBio long-reads . . . . .	105
V.2.5	Assemblies of Illumina reads and hybrid assemblies . . . . .	106
V.2.6	Hybrid assemblies of Illumina reads and PacBio long-reads . . . . .	106
V.2.7	In silico 16S rRNA gene extraction . . . . .	107
V.2.8	Estimation of completeness and contamination . . . . .	107
V.2.9	Binning of Pacbio contigs . . . . .	107
V.2.10	Annotation and search of homologous proteins . . . . .	108
V.2.11	Repartition of the COG in the main MAG . . . . .	109
V.2.12	Phylogenetic trees . . . . .	109
V.2.13	Coverages of the enzymes in the different samples . . . . .	110
V.3	Results . . . . .	110
V.3.1	Sequencing of metagenomic DNA . . . . .	110
V.3.2	De novo assemblies . . . . .	111
V.3.3	Evaluation of the quality of PacBio assembly with different number of polishing rounds . . . . .	113
V.3.4	Grouping of MAG . . . . .	113
V.3.5	Denitrification-related genes in MAG and reference genomes . . . . .	128
V.3.6	Phosphate metabolism-related genes in MAG and reference genomes . . . . .	130
V.3.7	Analysis of the most complete MAG . . . . .	135
V.3.8	Other MAG . . . . .	139
V.3.9	Coverage of hydrolysis- and biofilm-related genes in the different samples	140
V.4	Discussion . . . . .	144
V.4.1	Impact of the DNA extraction methods on the abundance of Actinobacteria	144
V.4.2	Evaluation of the assembly and binning . . . . .	144
V.4.3	The diversity of <i>Ca. Accumulibacter</i> complicates the assembly of the genomes from different clades . . . . .	146
V.4.4	The different clades of <i>Ca. Accumulibacter</i> present in the AGS . . . . .	146
V.4.5	Marker genes for PAO metabolism . . . . .	147
V.4.6	Biofilm formation . . . . .	148
V.4.7	Repartition of COG in the MAG . . . . .	148
V.5	Conclusions . . . . .	148

<b>VI General discussion</b>	<b>151</b>
VI.1 Bacterial communities in AGS fed with different wastewater types . . . . .	151
VI.2 Functional groups in AGS . . . . .	154
VI.2.1 PAO and GAO . . . . .	154
VI.2.2 Fermenting bacteria . . . . .	157
VI.2.3 Filamentous bacteria . . . . .	161
VI.3 Granulation and settling properties . . . . .	161
VI.4 Evolution of the PAO microdiversity . . . . .	163
VI.5 Nitrogen-removal according to the wastewater type . . . . .	166
VI.6 Functional groups in flocs and granules . . . . .	167
<b>VII Concluding remarks and perspectives</b>	<b>169</b>
<b>A Supplementary material</b>	<b>173</b>
A.1 Electronic supplementary material chapter II . . . . .	173
A.2 Supplementary information chapter III . . . . .	174
A.3 Supplementary information chapter IV . . . . .	182
A.3.1 Electronic supplementary material chapter IV . . . . .	185
A.4 Supplementary information chapter V . . . . .	186
A.4.1 Electronic supplementary material chapter V . . . . .	186
A.4.2 Fragment analysis reports and evaluation of the assembly after polishing	186
A.5 Supplementary information chapter VI . . . . .	194
<b>Bibliography</b>	<b>226</b>
<b>Curriculum Vitae</b>	<b>227</b>



# List of Figures

II.1	Bacterial community composition of the individual granules from metagenome O1 . . . . .	17
II.2	Bacterial community composition of the individual granules from metagenome O1 . . . . .	18
II.3	Distribution of the mean proportion of <i>Ca. accumulibacter</i> obtained from random <i>in silico</i> sampling of 3, 10 and 20 granules from metagenome O1 . . . .	19
II.4	Distribution of the mean proportion of <i>Ca. Accumulibacter</i> obtained from random <i>in silico</i> sampling of 3, 10 and 20 granules from metagenome O2 . . . .	19
II.5	Correlations between phages and bacteria-related sequences in the metagenome O1. . . . .	20
II.6	Correlations between phages and bacteria-related sequences in the metagenome O2 . . . . .	21
III.1	Conceptual model of carbon utilization and proposed desired undesired pathways in AGS systems . . . . .	30
III.2	Evolution of the sludge volume index SVI <sub>30</sub> and the SVI ratios (30/5 and 30/10) of the aerobic granular sludge of the four SBR. . . . .	37
III.3	Evolution of the sludge size fractions for the aerobic granular sludge fed with different influent composition . . . . .	39
III.4	Composition of bacterial communities from inoculation to stable state in the four AGS reactors treating different types of wastewater . . . . .	40
III.5	Principal coordinate plot based on the Bray-Curtis distance matrix of the bacterial OTU relative abundance in the sludge samples collected in the four reactors	41
III.6	Evolution of the Shannon diversity in the four reactors . . . . .	43
III.7	Average relative abundance of the main genera in the flocs and granules fractions collected in the four reactors during the stable state . . . . .	44
III.8	Multiple factor analysis performed on the data at bacterial stable state in the four reactors with settling characteristics, nutrient-removal and composition of the bacterial communities . . . . .	46
III.9	Principal component analysis based on the Bray-Curtis distance between the relative abundance of bacterial taxa, the granule size distributions and settling properties, and the nutrient-removal performances in the four reactors during bacterial stable state . . . . .	47

## List of Figures

---

III.10	Correlation heatmap between the discriminant taxa and <i>Ca. Accumulibacter</i> and the sludge size distribution, the settling properties and the nutrient removal efficiencies of the samples collected during the stable state, in the four reactors and average relative abundance of the taxa . . . . .	49
III.11	Boxplot of the Hellinger transformed abundance of the more discriminant genera, <i>Tetrasphaera</i> and <i>Ca. Accumulibacter</i> , in the four reactors and the inocula . . . . .	50
IV.1	Summarizing diagram of the reactor operation and the two experiments . . . . .	65
IV.2	$\text{PO}_4^-$ , $\text{NH}_4^-$ and TN-removal performance of the AGS during experiment 1 and process events susceptible to having had a negative impact on nutrient removal performance . . . . .	72
IV.3	$\text{PO}_4^-$ , $\text{NH}_4^-$ and TN-removal performances during the transition back from complex monomeric to simple monomeric influent wastewater and process event susceptible to having had a negative impact on nutrient removal performance . . . . .	73
IV.4	Evolution of the settling characteristics of the sludge during the transition from simple to complex monomeric wastewater and to complex polymeric wastewater and during the transition from complex monomeric back to simple wastewater . . . . .	75
IV.5	Pictures of the biomass sampled during experiment 1 and experiment 2 . . . . .	76
IV.6	Pictures of settling biomass taken during SVI measurements . . . . .	77
IV.7	Evolution of the bacterial community composition during the transition from simple to complex monomeric wastewater and to complex polymeric wastewater . . . . .	78
IV.8	Evolution of the bacterial communities during the transition back from complex monomeric to simple wastewater . . . . .	80
IV.9	Principal coordinate plot based on the Bray-Curtis distance matrix of the bacterial communities . . . . .	81
IV.10	The seven bacterial stable states detected during experiment 1 and 2. . . . .	82
IV.11	Principal component analysis based on the Bray-Curtis distance of the bacterial communities at stable states . . . . .	85
IV.12	Granules and flocs heatmap showing the average relative abundance of the abundant taxa during stable states 2, 4, 5 and 6 . . . . .	87
IV.13	Correlation heatmap between the discriminant taxa and the nutrient removal during experiment 1 and 2 and average Hellinger-transformed relative abundance of these taxa in the stable states corresponding to the different wastewater types . . . . .	88
IV.14	Multiple factor analysis performed on the nutrient-removal performance and the bacterial communities of the samples at bacterial stable state . . . . .	90
V.1	Relative abundance of the main bacterial genera estimated by the abundance of the 16S rRNA gene amplicon sequencing of the DNA extracted from sample from day 71 and 316 with four different extraction methods. . . . .	111
V.2	Phylogenetic tree of the 16S rRNA gene sequences extracted from the PacBio assembly and the hybrid assembly of Illumina and PacBio reads. . . . .	116

V.3	Phylogenetic tree of the 16S rRNA gene sequences extracted from PacBio contigs and selected reference genomes and indication of the sampling days . . . . .	123
V.4	Assembly graph of PacBio long-reads from sample d71 . . . . .	124
V.5	Assembly graph of PacBio long-reads from sample d322 . . . . .	125
V.6	Assembly graph of PacBio long-reads from sample d427. The contigs are colored according to the MAG they were classified to . . . . .	126
V.7	Assembly graph of PacBio long-reads from sample d740 . . . . .	127
V.8	Phylogenetic tree of the sequences matching the HMM model constructed from Pit sequences of <i>Ca. Accumulibacter</i> and <i>Tetrasphaera</i> . The tree is based on an alignment of 369 amino acids. . . . .	131
V.9	Phylogenetic tree of the sequences matching the HMM models of ppk1 and ppk2134	
V.10	Principal component analysis of the presence/absence of the COG in the main MAG. . . . .	135
V.11	Presence/absence of the COG related to inorganic ion transport and metabolism in the main MAG. . . . .	141
V.12	Presence/absence of the COG related to signal transduction mechanisms in the main MAG. . . . .	142
V.13	Average coverage of the selected EC number corresponding to enzymes related to hydrolysis and quorum-sensing detected in the DNA extracted from the biomass collected on day 71, 427 and 740 . . . . .	143
VI.1	Principal coordinate analysis of the Bray-Curtis distance of the bacterial communities at stable states in the two experiments . . . . .	152
VI.2	Relative abundance of PAO and GAO during the start-up from EBPR activated sludge to AGS in reactors treating different wastewater types . . . . .	155
VI.3	Relative abundance of PAO during the transition from simple monomeric, to complex monomeric and to complex polymeric influent wastewater in EPFL experiment. . . . .	156
VI.4	Relative abundance of GAO during the transition from simple monomeric, to complex monomeric and to complex polymeric influent wastewater in EPFL experiment . . . . .	157
VI.5	Composition of the guild of PAO in the Eawag and EPFL experiments, according to the different wastewater types. Functions were provided by MiDAS browser .	158
VI.6	Composition of the guild of GAO in the Eawag and EPFL experiments, according to the different wastewater types. Functions were provided by MiDAS browser.	158
VI.7	Relative abundance of fermenting and filamentous bacteria during the start up from EBPR activated sludge to AGS in reactors treating different wastewater types	159
VI.8	Relative abundance of fermenting bacteria during the transition from simple monomeric, to complex monomeric and to complex polymeric influent wastewater in EPFL experiment. The functions were attributed to taxa defined at genus level based on the functional information provided on MiDAS browser. . . . .	160

## List of Figures

---

VI.9	Relative abundance of filamentous bacteria during the transition from simple monomeric, to complex monomeric and to complex polymeric influent wastewater in EPFL experiment. The functions were attributed to taxa defined at genus level based on the functional information provided on MiDAS browser. . . . .	161
VI.10	Hellinger transformed proportions of <i>Ca. Accumulibacter</i> and <i>Tetrasphaera</i> related OTU during the transition from simple to complex monomeric and to complex polymeric wastewater . . . . .	164
VI.11	Relative abundance of functional groups in flocs and granules in the different bacterial stable states during the LBE and Eawag experiments . . . . .	168
A.1	Evolution of sludge morphology fed by 100%-VFA synthetic, complex synthetic, primary effluent and raw wastewater . . . . .	175
A.2	Evolution of SVI <sub>5</sub> , SVI <sub>10</sub> and SVI <sub>30</sub> and evolution of TSS and VSS in R1, R2, R3, R4	177
A.3	Three first axis of the principal component analysis of microbial communities structure evolution during the Eawag experiment . . . . .	178
A.4	Soluble COD and PO <sub>4</sub> <sup>3-</sup> -P concentrations after the anaerobic phase in R1, R2, R3, R4 . . . . .	180
A.5	Three first coordinates of the principal coordinate plot based on the Bray-Curtis distance matrix of the bacterial communities in the EPFL experiment . . . . .	182
A.6	Three first coordinates of the principal component analysis of the bacterial communities at stable states . . . . .	183
A.7	Fragment analysis report for the DNA extracted from sample d71 with the bacterial genomic DNA isolation CTAB protocol (extraction A) . . . . .	187
A.8	Fragment analysis report for the DNA extracted from sample d71 with the Maxwell® 16 Tissue DNA Purification Kit (extraction B) . . . . .	187
A.9	Fragment analysis report for the DNA extracted from sample d322 with the bacterial genomic DNA isolation CTAB protocol (extraction A) . . . . .	188
A.10	Fragment analysis report for the DNA extracted from sample d322 with the Maxwell® 16 Tissue DNA Purification Kit (extraction B) . . . . .	188
A.11	Fragment analysis report for the DNA extracted from sample d427 with the bacterial genomic DNA isolation CTAB protocol (extraction A) . . . . .	189
A.12	Fragment analysis report for the DNA extracted from sample d427 with the Maxwell® 16 Tissue DNA Purification Kit (extraction B) . . . . .	189
A.13	Fragment analysis report for the DNA extracted from sample d740 with the bacterial genomic DNA isolation CTAB protocol (extraction A) . . . . .	190
A.14	Fragment analysis report for the DNA extracted from sample d740 with the Maxwell® 16 Tissue DNA Purification Kit (extraction B) . . . . .	190
A.15	View on IGV of the mapping of PacBio long-reads and hybrid contigs from sample d427 on the PacBio contig utg000002c after 1 and 3 rounds of polishing	191
A.16	View on IGV of the mapping of PacBio long-reads and hybrid contigs from sample d427 on the PacBio contig utg000195l after 3 and 7 rounds of polishing	192

A.17 View on IGV of the mapping of PacBio long-reads and hybrid contigs from sample d427 on the PacBio contig utg000012l after 3 and 7 rounds of polishing 193

A.18 Principal coordinate analysis of the Bray-Curtis distance of the bacterial communities of the stable states of the two experiments on the 1st and 2nd axis (A), on the first and 3rd axis (B) and on the second and 3 axis (C). . . . . 194



# List of Tables

I.1	Typical nutrient concentrations in Swiss and Dutch urban wastewater and the effluent standards defined by the Swiss legislation. . . . .	4
III.1	Measured influent composition of the four SBR fed by 100%-VFA synthetic WW, complex synthetic wastewater, primary effluent wastewater and raw wastewater	33
III.2	Effluent concentrations and nutrient removal performances of the four AGS reactors fed with 100 %-VFA synthetic, complex synthetic, primary effluent and raw influent wastewater . . . . .	45
IV.1	Theoretical characteristics of the different influent synthetic wastewater . . . .	67
IV.2	Table of the discriminant taxa of the study . . . . .	83
IV.3	Variations in the relative abundance of the discriminant taxa of AGS across the different stable states observed with complex monomeric wastewater . . . . .	84
V.1	Characteristics of the PacBio and Illumina sequences before and after the filtering.	112
V.2	Main characteristics of the assemblies of the PacBio and/or Illumina sequencing data . . . . .	115
V.3	Information collected on the reference genome sequences. The estimated proportions of completeness and contamination, produced by CheckM and BUSCO analysis, the number of 16S rRNA gene sequences and the number of coding sequences where collected following the same procedure as for PacBio contigs	117
V.4	Selection of MAG obtained after the grouping of the PacBio contigs with the estimated taxonomy, the coverage of the Illumina reads obtained with extraction A and B, the number of contigs and the total length, the N50 statistic, the number of coding sequences detected, the estimated proportions of completeness and contamination, produced by checkM and BUSCO analysis . . . . .	118
V.5	Positive match with the HMM models for genes involved in denitrification found in the main MAG and selected reference genomes . . . . .	129
V.6	Positive match with the hmm models for genes involved in the phosphorus metabolism found in the main MAG and selected reference genomes) . . . . .	133
VI.1	Main taxa common to Eawag and EPFL experiments enriched in the simple wastewaters, the complex polymeric wastewaters or in both wastewater types.	153

**List of Tables**

---

A.1 Abundance of the phages- and bacteria-related sequences in the metagenome O1 173

A.2 Correlations between phages- and bacteria-related sequences in the metagenome O1 . . . . . 173

A.3 P-values of the correlations between phages- and bacteria-related sequences in metagenome O1 . . . . . 173

A.4 Abundance of the phages- and bacteria-related sequences in the metagenome O2 173

A.5 Correlations between phages- and bacteria-related sequences in the metagenome O2 . . . . . 173

A.6 P-values of the correlations between phages- and bacteria-related sequences in metagenome O2 . . . . . 173

A.7 Recipe of 100%-VFA synthetic and complex synthetic wastewater for reactors R1 and R2 . . . . . 176

A.8 P-values of the t-test comparing the relative abundance of the main genera in the floccular and the granular fraction of the sludge. . . . . 179

A.9 Mantel test comparing the Bray-Curtis distance matrices of the microbial communities, the settling and size characteristics of the sludge and its nutrient-removal performance, at bacterial stable state . . . . . 179

A.10 P-values of the t-test performed on the mean relative abundance of the taxa between different stable states or groups of stable states. The divergent, abundant and discriminant taxa are indicated . . . . . 181

A.11 Bray-Curtis distance matrices of the bacterial community compositions, the settling properties and size distribution of the sludge and its nutrient-removal performances at stable state . . . . . 181

A.12 Average relative abundance of the bacterial taxa (at genus level) in the inoculum and the four reactors during stable state . . . . . 181

A.13 Recipes for the preparation of the different synthetic wastewater used in this study 184

A.14 Bray-Curtis distance matrix of the bacterial communities in the AGS samples of experiment 1 and 2 . . . . . 185

A.15 P-values of the t-tests comparing the mean abundance of each taxa detected in the AGS of E1 and E2 in the three wastewater types (simple, complex monomeric and complex polymeric) . . . . . 185

A.16 Mean relative abundance of the taxa (genus level) in the AGS samples treating the different wastewater types (simple, complex monomeric and complex polymeric) 185

A.17 P-values of the t-tests comparing the mean abundance of each taxa detected in the AGS of E1 and E2 in the seven bacterial stable states . . . . . 185

A.18 P-values of the t-tests comparing the mean abundance of each taxa detected in the AGS of E1 and E2 in the seven bacterial stable states . . . . . 186

A.19 Complete table of the all the MAG obtained from grouping PacBio contigs. . . . 186

A.20 Distance between the characteristic tetranucleotide frequency vectors of selected reference genomes . . . . . 186

A.21 Number of selected P-related hmm profiles found in the MAG and not grouped PacBio contigs . . . . . 186

A.22 Number of selected N-related hmm profiles found in the MAG and not grouped  
PacBio contigs . . . . . 186



# List of accronyms

- AGS** Aerobic granular sludge
- BOD** Biological oxygen demand
- bp** Base pairs
- BUSCO** Bench-marking universal single-copy orthologs
- Ca.** Candidatus
- CDS** Coding sequence
- COD** Chemical oxygen demand
- COG** Cluster of orthologous genes
- DNA** Deoxyribonucleic acid
- EBPR** Enhanced biological phosphorus removal
- ENA** European nucleotide archive
- EPS** Extracellular polymeric substances
- EPV** EBPR podovirus
- FISH** Fluorescent in situ hybridization
- GAO** Glycogen accumulating organism
- GB** Giga byte
- HMM** hidden Markov model
- kbp** Kilobase pairs
- LBE** Laboratory for environmental biotechnology
- MAR** Micro auto radiography

## List of acronyms

---

- Mbp** Mega base pair
- MFA** Multiple factor analysis
- N** Nitrogen
- NGS** Next generation sequencing
- OGU** Operational genomic unit
- ORF** Open reading frame
- OTU** Operational taxonomic units
- P** Phosphorus
- PAO** Phosphate accumulating organism
- PBS** phosphate buffer saline
- PCoA** Principal coordinate analysis
- PCR** Polymerase chain reaction
- PF** Plug-flow feeding
- PHA** Polyhydroxyalkanoate
- Pit** low-affinity inorganic phosphate transporter
- Pst** high-affinity phosphate ABC transporter system
- ppk** Polyphosphate kinasse
- Q** Flow rate
- RAM** Random access memory
- rRNA** Ribosomal ribonucleic acid
- SBR** Sequencing batch reactor
- SND** Simultaneous nitrification denitrification
- SNP** single nucleotide polymorphism
- sRNA** small RNA
- SRT** Sludge retention time
- SS** Suspended solids
- SVI** Sludge volume index

<b>TN</b>	Total nitrogen
<b>TKN</b>	Total Kjeldahl nitrogen
<b>TP</b>	Total phosphorus
<b>tRNA</b>	transfer RNA
<b>TS</b>	Total solids
<b>TSS</b>	Total suspended solids
<b>V</b>	Volume
<b>v.</b>	version
<b>VER</b>	Volume exchange ratio
<b>VFA</b>	Volatil fatty acid
<b>VS</b>	Volatil solids
<b>VSS</b>	Volatile suspended solids
<b>WW</b>	Wastewater
<b>WWTP</b>	Wastewater treatment plant



À TABLE!





# I Introduction

## I.1 Aerobic granular sludge for wastewater treatment

### I.1.1 Wastewater treatment

Efficient and sustainable wastewater treatment is one of the major stakes of the 21<sup>th</sup> century [Infos-eau, 2009, WHO, 2019]. Nowadays, only three billion of people benefit from complete sanitation including wastewater treatment or safe disposal [WHO, 2019]. Even in high income countries, many wastewater treatment plants are overwhelmed due to densifying populations in urban area and needs to be upgraded often with the constrain of limited footprints. Wastewater treatment plants using activated sludge, based on the discovery of Ardern and Lockett in [1914], are commonly used in high-to-medium-income countries since 1930 [IWA, 2009]. Activated sludge systems are based on the observation that wastewater containing microorganisms can be treated with aeration, followed by a sedimentation and the reinjection of the settled microorganisms in the new wastewater [Ardern and Lockett, 1914]. Thus, the macro-pollutant mainly composed of organic carbon, ammonium and phosphate can be removed from the wastewater. Nowadays, this microbiological process is preceded with screening of coarse material, grit-removal, grease and other light material scumming and primary sedimentation. It is followed by a secondary sedimentation, after which the wastewater is generally discharged in the environment. The increasing concentrations of micro-pollutants in the wastewater are raising more and more concern and some states are including thresholds for micro-pollutant in the effluent wastewater. Micro-pollutants are not efficiently removed by traditional wastewater treatment processes and often require a tertiary treatment (eg. ozonation) [Margot et al., 2011].

Wastewater macro-pollutant concentrations can vary from one region to the other; it depends on a lot on factors as pluviometry, population customs, characteristics of the sewer system, and the wastewater origin. As an example, typical values for municipal wastewater in Switzerland and The Netherlands, two countries that have different topologies are shown in Table I.1.

In mountainous regions particulate slowly biodegradable chemical oxygen demand (COD)

## Chapter I. Introduction

can be higher than 50 % due to steep slope sewers and thus shorter retention time in these sewers [van Nieuwenhuijzen et al., 2004]. The macro-pollutants contained in wastewater are harmful for the environment and must therefore be properly removed before the water can be released in nature. Ammonium and nitrite are toxic for fish, nitrate and phosphate lead to eutrophication of water bodies and carbonaceous organic matter can favor the proliferation of pathogens [von Sperling, 2007]. In Switzerland, the ruling on waters protection [OFEV, 2011], sets the thresholds for suspended solids, COD, ammonium, nitrite, nitrate and phosphate in the treated water discharged in natural aquatic ecosystems (Table I.1). It is expected that if the macro-pollutant charge is sufficiently low, the aquatic ecosystem is capable of self-purification and will therefore not be altered [von Sperling, 2007]. In case of too high macro-pollutant discharges in aquatic ecosystems, they can be altered generally with a decrease of the biodiversity and ultimately eutrophication (overgrowth of aquatic plants transforming dramatically oxygen and nutrient concentrations, leading to the death of oxygen-dependent species [Pretty et al., 2003]).

Table I.1 – Typical nutrient concentrations in Swiss and Dutch urban wastewater and the effluent standards defined by the Swiss legislation.

Wastewater origin	SS [g m <sup>-3</sup> ]	COD [g O <sub>2</sub> m <sup>-3</sup> ]	BOD <sub>5</sub> [g O <sub>2</sub> m <sup>-3</sup> ]	TKN [g N m <sup>-3</sup> ]	NH <sub>4</sub> <sup>+</sup> [g N m <sup>-3</sup> ]	P [g P m <sup>-3</sup> ]
Dutch [M.N. et al., 2004]	237	450	171	42	NA	6.7
Swiss [Gujer, 2007]	200	340	170	30	20	6
Swiss thresholds [OFEV, 2011]	15-20	NA	15-20	NA	2 (90 %)	0.8 <sup>a</sup> (80 %)

<sup>a</sup> for wastewater discharged in sensitive environments

Aside conventional activated sludge systems, new technologies are developed in order to take up the challenge of energy-efficient, low-footprint and low-energy requiring wastewater treatment processes (eg. biotrickling filters, moving bed biofilm reactors, anammox) [Rusten et al., 1992, Strous et al., 1999, Togna and Singh, 1994]. Enhanced biological phosphorus- (P)-removal (EBPR) has demonstrated a high sustainability due to efficient P removal without addition of chemicals [Stokholm-Bjerregaard et al., 2017]. Consequent chemical savings and reduced sludge production resulting from biological phosphorus removal improves the environmental and economical sustainability of wastewater treatment [Coats et al., 2011]. It can however be subject to temporary failures due to reduced activity of the P-removing microorganisms [Martin et al., 2006].

### I.1.2 Aerobic granular sludge

Processes based on aerobic granular sludge (AGS) developed over the past 20 years offer an interesting alternative to conventional activated sludge systems. Aerobic granules were defined as “aggregates of microbial origin, which do not coagulate under reduced hydrodynamic shear and settle significantly faster than activated sludge” [de Kreuk et al., 2007]. AGS is obtained from EBPR activated sludge inoculum by applying a combination of conditions known to favor granulation while maintaining good biological nutrient performance : the alternation

## **I.1. Aerobic granular sludge for wastewater treatment**

---

of anaerobic feast and aerobic famine phases [Morgenroth et al., 1997], short settling time selecting for fast-settling biomass [Barr et al., 2016, Katherine Jungles et al., 2011, Liu and Liu, 2006, Robertson, Struan and van der Roest, Helle and van Bentem, André, 2015], high shear-forces, provided by mixing or superficial air velocity [Beun et al., 1999, Chen et al., 2007, Lochmatter and Holliger, 2014] and high concentrations of biodegradable organic COD in the influent wastewater [Nancharaiah and Reddy, 2018, Wang et al., 2009].

At full-scale, Nereda® is a technology based on AGS. It results from cooperation between public and private research. After the success of the first AGS pilot scale treating real wastewater in 2003 [de Kreuk and van Loosdrecht, 2006], the first full-scale plants using AGS were started in 2005. They are currently able to remove organic matter to reach the required pollutant load thresholds for the effluent. More than 50 full-scale municipal and industrial plants using Nereda® technology, sometimes in combination with other technologies, are thus far in operation or under construction worldwide [DHV, 2019].

Compared to activated sludge, AGS allows higher biomass concentration [Beun et al., 2000, de Kreuk et al., 2007, Etterer and Wilderer, 2001] simultaneous COD, nitrogen (N) and P removal [de Bruin et al., 2004] and shorter settling time [Basheer and Farooqi, 2014, Ni et al., 2009]. Different redox conditions can be present in granules at the same time providing the possibility to perform simultaneous nitrification-denitrification (SND) and thus allowing to avoid supplementary tanks for adequate denitrification conditions. All these characteristics enable the setup of compact plants with significantly reduced footprints, energy consumption and costs [Beun et al., 1999, de Bruin et al., 2004, Morgenroth et al., 1997, Pronk et al., 2015b, Swiatczak and Cydzik-Kwiatkowska, 2018]. Yet it is believed that the process can be improved further by understanding the function and interactions of the microorganisms composing AGS and adapting the operational conditions consequently [Pronk et al., 2015b].

### **I.1.3 Research on aerobic granular sludge**

Research on AGS was and still is often performed in lab-scale reactor with simple synthetic wastewater with volatile fatty acids (VFA) such as acetate and propionate as carbon source [Barr et al., 2016, Beun et al., 1999, Bin et al., 2015, de Kreuk et al., 2005, 2007, Sarma et al., 2017]. Substantial knowledge has been obtained from those studies. It provided the opportunity to study thoroughly the metabolism of model organisms with important functions [Barr et al., 2016] and their roles in granulation [de Kreuk and van Loosdrecht, 2004] and nutrient removal [Oehmen et al., 2010]. Importantly, it documented the negative impact of COD that was not consumed during the anaerobic phase and leaks in the aerobic phase, on granulation and nutrient removal performance [Weissbrodt et al., 2014a].

In real wastewater, generally more than 50 % of the COD is contained in the form of polymeric substances [Koch et al., 2000, Levine et al., 1991, Nielsen et al., 2010] and cannot be readily taken up by bacteria before undergoing hydrolysis. Hydrolysis is a slow process and was reported to be slower under anaerobic conditions compared to aerobic conditions [de Kreuk

et al., 2010, Goel et al., 1998b). Moreover, it is performed at the cell surface [Frolund et al., 1995, Goel et al., 1997] and particles above 1  $\mu\text{m}$  have a limited diffusibility across biofilms [Drury et al., 1993, Janning et al., 1998]. The particulate COD in the influent wastewater is therefore slowly taken up during the anaerobic phase and is likely leaking in the aerobic phase [de Kreuk et al., 2010]. This partial consumption of COD during the anaerobic phase is supposed to compromise the granulation and nutrient removal performance of the AGS [Bassin et al., 2012, Wagner et al., 2015b]. Indeed, a significant proportion of flocs seems to be constitutive of AGS treating real wastewater [Derlon et al., 2016, Pronk et al., 2015b]. Moreover, more porous or fluffy granules [de Kreuk et al., 2010] and unstable nutrient removal [Wang et al., 2018] have been reported in particular with low-strength polymeric wastewater [Wagner et al., 2015b, Winkler et al., 2013].

In consequence, the knowledge acquired on AGS treating simple monomeric wastewater cannot be simply transposed to AGS treating more complex wastewater containing polymeric compounds [Wagner et al., 2015b] and the operational conditions needs to be adapted to the composition of the wastewater. For example, with low-strength wastewater, a too harsh selection of fast settling biomass can lead to important washouts of the sludge. The volume exchange ratio and settling times can be tuned to allow granulation while keeping sufficient amount of biomass in the system [Ni et al., 2009, Wagner et al., 2015b].

Yet, to date, probably due to practical reasons (e.g., difficulties to have access to real wastewater), only a minority of studies on AGS are performed with wastewater containing substantial proportions of polymeric compounds [de Kreuk et al., 2010, de Kreuk and van Loosdrecht, 2006, Dulekgurgen et al., 2003, Lemaire et al., 2008, Lin et al., 2003, Schwarzenbeck et al., 2004].

### 1.1.4 Bacterial communities in aerobic granular sludge

Obviously, the structure of the microbial community in AGS influences the performance of the sequencing batch reactor (SBR) (e.g., COD, N, P-removal, sludge volume index (SVI)) [Lochmatter and Holliger, 2013, Tan et al., 2014, Weissbrodt et al., 2014a]. As they are able to incorporate polymers, protozoa participate to the degradation of particulate COD in the wastewater [de Kreuk et al., 2010]. Regarding the bacterial community, several key functional groups are crucial for good nutrient removal and settling properties. Fermenters (eg. *Tetrasphaera*, *Streptococcus*) and hydrolyzers (eg. *Curvibacter*) play a crucial role in COD removal, ammonium-oxidizing bacteria (eg. *Nitrospira*, *Nitrosomonas*), nitrite-oxidizing bacteria (eg. *Nitrospira*) and denitrifiers (eg. *Thauera*, *Azoarcus*, *Zoogloea*) participate in total nitrogen (TN) removal and phosphate-accumulating organisms (PAO) (eg. *Candidatus (Ca.) Accumulibacter*, *Tetrasphaera*) are responsible for most of the P-removal [Nielsen et al., 2010].

In EBPR activated sludge, the composition of bacterial communities is influenced by the type of wastewater used [Coats et al., 2017]. In particular, the composition of the guild of PAO is different in lab-scale reactors treating simple wastewater mainly composed of VFA and in

## I.1. Aerobic granular sludge for wastewater treatment

---

full-scale EBPR treatment plants [Mielczarek et al., 2013]. Microbial communities in AGS are probably slightly different from those in EBPR activated sludge [Bin et al., 2015], and to date, few studies have inspected their compositions in AGS treating polymeric wastewater [Cetin et al., 2018, Kang et al., 2018, Swiatczak and Cydzik-Kwiatkowska, 2018, Wang et al., 2018]. These microbial communities are therefore expected to depend on the type of influent wastewater, but to what extent and how is not clear yet.

In AGS, as in EBPR activated sludge, PAO is an essential guild. The first PAO candidate was *Acinetobacter* spp. but its abundance and its capabilities to remove phosphate in EBPR systems turned out to be too low [Wagner et al., 1994]. *Ca. Accumulibacter* was then considered as the bacteria responsible for most of the P-removal from wastewater [Hesselmann et al., 1999] and it became the model organism for PAO metabolism [Zilles et al., 2002]. To date, no pure culture of *Ca. Accumulibacter* can be grown but it was enriched up to 90 % in lab-scale reactors fed with VFA-based synthetic wastewater containing high phosphate concentrations [Lu et al., 2006, McMahon et al., 2007, Murray and Stackbrandt, 1995]. *Ca. Accumulibacter* has the ability to take up volatile fatty acids (VFA) anaerobically and store them in the form of poly-hydroxy-alkanoate (PHA) by using the energy provided by the hydrolysis of intracellular polyphosphate and glycogen reserves. In presence of oxygen (sometimes nitrate or nitrite), the PHA reserves are used for growth and the replenishment of large intracellular polyphosphate reserves. The growth of these bacteria leads to a net decrease of orthophosphate from the water [Mino et al., 1998, Oehmen et al., 2007].

Studies of the bacterial communities in EBPR wastewater treatment plants revealed that the participation of *Tetrasphaera* to P-removal was at least equivalent to the one of *Ca. Accumulibacter* [Hanada et al., 2002, Kristiansen et al., 2013, Muszynski and Zaleska-Radziwill, 2015, Nielsen et al., 2012b, Stokholm-Bjerregaard et al., 2017]. The metabolism of *Tetrasphaera* is to date not completely defined. It is slightly variable depending on the different species and differ from the one of *Ca. Accumulibacter*. *Tetrasphaera* can take up large amounts of polyphosphate in aerobic and possibly anoxic conditions after incorporation of sugars or amino-acids during the previous anaerobic phase, but not after the uptake of acetate [Barnard et al., 2017, Kristiansen et al., 2013]. Contrary to *Ca. Accumulibacter*, *Tetrasphaera* does not constitute substantial PHA reserves [Kristiansen et al., 2013]. The carbon source are likely fermented in anaerobic or aerobic conditions [Nielsen, 2017] and part the fermentation products including VFA are released [Barnard et al., 2017]. But whether part of those products are stored and under which form has not been fully elucidated yet. *Tetrasphaera* also have the capability to hydrolyze polysaccharides [Nielsen et al., 2010]. Several *Tetrasphaera* species have been cultivated in pure culture and the denitrification coupled to P-removal was observed for different members of this genus [Barnard et al., 2017, Kristiansen et al., 2013]. The co-occurrence of *Ca. Accumulibacter* and *Tetrasphaera* frequently observed in EBPR systems where they occupy different ecological niches is believed to increase the robustness of biological P-removal [Mielczarek et al., 2013].

Since then, other PAO or putative PAO affiliated to *Ca. Halomonas phosphatis* [Nguyen et al.,

2012], *Gemmatimonas aurantiaca* [Zhang et al., 2003], *Ca. Obsruribacter* [Soo et al., 2014] and *Dechloromonas* [Kim et al., 2013] have been identified and few other will possibly be discovered in the future. *Dechloromonas* is a putative PAO which metabolism is still to be elucidated. *Dechloromonas* seems able to use nitrate as a terminal electron acceptor [Zeng et al., 2003c].

Glycogen-accumulating organisms (GAO) have a metabolism similar to PAO without the capability to incorporate large amounts of polyphosphate. For this reason, they were often considered as the cause of biological P-removal failures [Burow et al., 2007, Cech and Hartman, 1993, Kong et al., 2006, Saunders et al., 2003]. However, numerous recent studies suggest that even in high abundance, GAO do not interfere with the uptake of organic carbon by PAO, which likely have a competitive advantage in case of high P-concentrations [Qiu et al., 2019, Stokholm-Bjerregaard et al., 2017]. Members of the Competibacteraceae family have a classical GAO metabolism, they can take up VFA anaerobically and store them in the form of PHA until further utilization for growth and glycogen storage reconstitution during aerobic and possibly anoxic conditions [McIlroy et al., 2015b].

An alternative GAO metabolism was observed in the Actinobacteria *Micropruina* [McIlroy et al., 2018]. This genus was detected in high abundance in EBPR treatment plants [Stokholm-Bjerregaard et al., 2017, Wong et al., 2005]. It can take up various carbon sources anaerobically, a part of it will be fermented, the other part will be stored as glycogen for subsequent utilization in aerobic conditions. Unlike classical GAO, they do not constitute substantial PHA reserves [McIlroy et al., 2018].

Due to their denitrifying capabilities, several GAO and PAO likely participate to the biological N-removal in EBPR and AGS systems. Classical GAO and PAO were associated with stable granulation [de Kreuk and van Loosdrecht, 2004] but the abundance of fermenting PAO and GAO in AGS treating polymeric wastewater or their impact on granulation is not known.

### I.1.5 Nitrogen removal

There are several pathways resulting in the removal of nitrogen from wastewater. The most common and likely major one in AGS is the nitrification-denitrification pathway [Guimaraes et al., 2017]. Ammonium released from the degradation of urea and proteins is oxidized into nitrite and then into nitrate during the typically two-step process of nitrification which requires oxygen but not COD. It is generally performed by two types of autotrophic microorganisms. Ammonium-oxidizing organisms (e.g., *Nitrosomonas*, *Nitrosospira*) [Lemaire et al., 2008] that are generally present in low proportions (< 0.5 %) [Swiatczak and Cydzik-Kwiatkowska, 2018, Xia et al., 2014] and located close to the granules surface (< 200  $\mu\text{m}$ ) [Lemaire et al., 2008, Soliman and Eldyasti, 2018] oxidize ammonium into nitrite. Nitrite can further be oxidized into nitrate by nitrite-oxidizing organisms (e.g., *Nitrospira*). Daims et al. [2015] and van Kessel et al. [2015] however discovered that some of the *Nitrospira* strains are capable to oxidize ammonium into nitrate. This process of complete oxidation of ammonium into nitrate by a

## I.1. Aerobic granular sludge for wastewater treatment

---

single bacterium was called comammox.

Nitrification can be followed by the denitrification pathway where nitrate and then nitrite are used as electron acceptors in absence of oxygen. Resulting nitric oxide can be reduced to nitrous oxide, a greenhouse gas, which can be further reduced to gaseous dinitrogen or leave the system as it is [Campos et al., 2016, Kampschreur et al., 2009, Metcalf & Eddy, 2014]. Indeed, it appears that some denitrifying bacteria do not have a functional nitrous oxide reductase and emission of  $N_2O$  can occur [Zeng et al., 2003a].

Denitrification requires COD, but occurs after nitrification when most of the COD has generally been consumed. In conventional wastewater treatment, several strategies can be set up in order to circumvent this problem. For example, part of the wastewater can be pumped back and mixed with incoming wastewater or a carbon source can be added to the nitrified wastewater. With AGS, different redox conditions are available inside the granule, thus nitrification and denitrification can theoretically be simultaneous. Moreover, if carbon was stored during the previous anaerobic phase by denitrifying bacteria (eg. PAO, GAO), it can be used for denitrification [Pochana and Keller, 1999, Third et al., 2003]. De Kreuk et al. [2005] found that SND was mainly function of the  $O_2$  saturation and the granule size. An  $O_2$  saturation of 40 % and a granule diameter of 1.3 mm were found to be optimal parameters for SND with their acetate-fed AGS. Lower or absent SND were reported in AGS treating polymeric wastewater [Kang et al., 2018] possibly due to the small granule size [Li et al., 2014, Ni et al., 2009] or the non-diffusibility of an important part of the COD that was consumed locally at the granules surface and was therefore not available in the anoxic regions in the granules [de Kreuk et al., 2010]. For both monomeric and polymeric wastewater, intermittent aeration was shown to enhance nitrate removal [Lochmatter and Holliger, 2014, Pronk et al., 2015a].

Denitrifying bacteria are generally represented by numerous taxa of diverse phylogenetic origin. Moreover, this phenotype is not homogeneously distributed in the members of a single genus. It is therefore difficult to identify this phenotype based on amplicon 16S rRNA genes information [Jiang et al., 2008, Olson et al., 2017]. Denitrifying PAO are particularly interesting for biological nutrient removal with low-to-medium strength wastewater where COD can be a limiting factor, since they can couple N-removal to P-removal without extra-COD consumption [Kuba et al., 1996]. Genomic and *in situ* evidence showed that most *Ca. Accumulibacter* and *Tetrasphaera* found in wastewater treatment systems are involved in denitrification at least via nitrate or nitrite reduction [Flowers et al., 2009, Gonzalez-Gil and Holliger, 2011, Kim et al., 2013, Kristiansen et al., 2013].

In anoxic conditions, ammonium and nitrite can be converted into gaseous nitrogen by anammox bacteria [Mulder et al., 1995]. Five anammox *Candidatus* genera belonging to the phylum Planctomycetes were identified to date. They produce a specific lipid (ladderane) which helps to confine the intermediate compound, hydrazine, produced during the anammox process [Oshiki et al., 2016]. This process, currently applied in specific full-scale plants, has the advantage to remove nitrogen from the water with fewer or no COD consumption than

the nitrification/denitrification, but it requires anoxic condition [Ma et al., 2016].

### I.1.6 Biofilm, extracellular polymeric substances and quorum-sensing

AGS are biofilms, hence a gathering of multiple microorganism populations embedded in a protective matrix composed of secreted compounds referred to as extracellular-polymeric substances (EPS). The cooperation of bacteria observed in biofilms and the specialization induced by gene differential expression can make them resembling multicellular organisms [Jefferson, 2004]. Due to diffusion limitations of gases and nutrients, biofilms provide different ecological niches with local changes in redox conditions, pH and nutrient availability [Cvitkovitch et al., 2003, Xavier et al., 2007]. Biofilm can favor exchanges of genetic material [Cvitkovitch et al., 2003] and provide an efficient protection against antibiotics and heavy metals [Lewis, 2001, Teitzel and Parsek, 2003].

The composition of the EPS matrix influences the physical characteristics of the biofilm such as its porosity and stability [Lin et al., 2013, Seviour et al., 2009, Sutherland, 2001]. AGS have a different EPS composition than floccular sludge. Some key gel-forming EPS have been identified in certain AGS (eg. Granulan, alignate, alginate-like) [Nielsen et al., 2012b, Seviour et al., 2011] but they seem to depend on the bacterial community constituting the biofilm [Nielsen et al., 2012b]. However, the identification of EPS can be biased by the extraction method [Seviour et al., 2012]. Therefore, the typical composition of the AGS EPS matrix is still to be investigated.

Almost all bacteria can form biofilms [Teschler et al., 2015]. The transition from a planktonic mode to a biofilm mode is triggered by external conditions and leads to the modification of the expression of large subsets of genes [Jefferson, 2004]. Quorum-sensing plays an important role in biofilm regulation [Merritt et al., 2003, Sakuragi and Kolter, 2007, Wang et al., 2012, Xia et al., 2012]. Some bacteria constitutively release small molecules named autoinducers. Their concentrations therefore inform bacteria on the density of other bacteria in their vicinity [McIlroy et al., 2017]. A specific bacterial population recognizes its own signal but also those from other populations [Federle and Bassler, 2003]. Tan et al. [2014] studied the implication of the gram<sup>-</sup> specific autoinducer N-acyl-homoserine-lactone (AHL) quorum sensing system in the formation of granules from floccular sludge. They observed that high concentrations of AHL were strongly correlated with granule formation and an increased EPS production. On the opposite, low AHL concentrations corresponded to a large proportion of flocs or granule collapsing. This was confirmed by [Lv et al., 2014]. They also identified a cluster of microbial populations whose abundance increased with AHL. The question whether particular communities are AHL-producers, recipient or both is still to be elucidated [Liu and Liu, 2006]. Quorum-sensing independent biofilm regulation were also detected, some of them implying sRNA, and can be the subject of further studies [Teschler et al., 2015].

Several ecological factors were identified as having an influence on biofilm formation, structure or composition. Besides chemical stress (e.g., pH, antibacterial molecules), physical stress

(e.g., shear forces) can favor biofilm formation [Yarwood et al., 2004]. Protozoa, through their movements inside the biofilm and their consumption of microorganism and EPS, can influence the biofilm structure, making it more porous and loose [Böhme et al., 2009]. Protozoa also have an influence on the bacterial community even though it was estimated lower than for planktonic communities [Wey et al., 2012].

The biofilm matrix does not protect the bacteria against phage attacks. Indeed, most phages can produce degradative enzymes (e.g., polysaccharide lyases) [Sutherland et al., 2004, Sutherland, 1995] able to break down the EPS matrix or access the bacteria through channels inside the biofilm. Moreover, an important cell lysis due to a phage attack is also likely to impact the biofilm structure as well as the bacterial community composition [Brockhurst et al., 2006, Middelboe, 2000, Stern et al., 2012]. Since the dominant bacterial populations are more susceptible to a phage attack, the latter tend to favor a high bacterial diversity [Thingstad and Lignell, 1997].

### I.1.7 Studying aerobic granular sludge through molecular methods

An important part of bacteria found in AGS systems were to date not isolated in pure culture [Amann et al., 1995, Luecker et al., 2010]. Molecular techniques such as MAR-FISH combining the utilization of radioactive substrate with fluorescent *in situ* hybridization or metagenomic, metatranscriptomic and proteomic [Lawson et al., 2017, McIlroy et al., 2014, McIlroy et al., 2017] approaches have helped to increase our comprehension of various wastewater treatment systems treating complex wastewater (eg. annamox, EBPR activated sludge) [Ju and Zhang, 2015, Nielsen et al., 2012b]. Such techniques can be useful to investigate the causes of the differences in the physical structure and nutrient removal performance of AGS treating monomeric and polymeric wastewater.

## I.2 Research questions and thesis overview

Since the majority of studies on AGS used simple synthetic influent containing mainly soluble substrates (e.g., VFA such as acetate and propionate), the general goal of this thesis was to identify potential differences between bacterial communities in AGS treating simple wastewater containing acetate and propionate as carbon source and bacterial communities in AGS treating complex wastewater containing fermentable and polymeric compounds. The influence of these communities and the wastewater composition on the AGS settling properties and nutrient removal performance was investigated in order to gain understanding of the differences observed between AGS treating different wastewater types. Another important objective of this thesis was to relate potential functions (genes) to specific bacterial populations composing the biomass of AGS reactors with different wastewater compositions. In particular, the homogeneity of functional capabilities (eg. denitrification) inside individual genera was investigated. The focus was on the functions linked to nutrient removal performance and granulation.

## Chapter I. Introduction

---

As a first step, the bacterial community diversity in individual granules of AGS treating simple synthetic wastewater was studied in order to assess the minimal sampling effort required for the following bacterial community analysis. The main results are presented in chapter II. Chapters III and IV present the results of two different approaches used to study the influence of wastewater composition on AGS bacterial communities, settling properties and nutrient removal performance. In Eawag experiment, the start-up of AGS from EBPR activated sludge was studied in four lab-scale reactors treating in parallel different wastewater types (simple synthetic and complex polymeric synthetic wastewater, two kinds of real wastewater). In particular the bacterial communities, settling properties and nutrient removal performance were monitored over one year and the kinetics of granulation were linked with the influent composition. The EPFL experiment was designed to study the evolution of the bacterial communities, the settling properties and nutrient removal performance on already formed AGS acclimated to a simple synthetic wastewater during a first transition to a complex monomeric wastewater containing 2/3 of fermentable compounds and a second transition to a complex polymeric wastewater containing 1/3 of polymeric compounds. The potential functions of the main bacterial populations in the AGS treating these three different synthetic wastewater were investigated by whole metagenomic sequencing and subsequent bioinformatic analysis (chapter V). In chapter VI, a general discussion synthesizes and compares the information collected in the four first chapters.

NE ME RENVOYEZ PAS  
DANS MON GRANULE,  
IL EST INFESTÉ DE  
BACTÉRIOPHAGES!

SECRÉTARIAT  
DU GRANULE  
AUX MIGRATIONS





# II High variability of microbial community composition among individual granules

## II.1 Introduction

Microbial community compositions are the results of multiple factors that can be categorized as intrinsic if they are related to the community itself or extrinsic if they are related to the environment in which the community thrives. It is generally difficult to dissociate the effect of these two kinds of factors due to the changing environmental conditions. Aerobic granular sludge (AGS) provides numerous community replicates subject to the same environmental conditions. However, it is not known whether individual granules contain very similar microbial communities or whether the latter can be quite different among individual granules.

In the framework of a collaborative project, the metagenomes of 142 individual granules collected at two different time points from the routine AGS reactor at the laboratory for environmental biology (LBE) were sequenced in order to study the evolution and diversity of microorganisms in multiple replicates receiving the same environmental conditions. The results of these separate investigations showed that community replicates represented by individual granules varied much more in their composition at genus-level resolution than would be expected from neutral assembly processes [Leventhal et al., 2018]. For the purpose of this thesis, the sequence information of these two metagenome data-sets was used to identify the minimal sample size in order to have an amount of biomass that is well representative of the overall reactor biomass.

Phages are known to have a potentially big influence on the bacterial communities [Koskella and Brockhurst, 2014]. In particular, decreases in *Ca. Accumulibacter* abundance have been linked with the attack of bacteriophages [Barr et al., 2010b, Skennerton et al., 2011]. Based on this information, the correlations between phages and bacteria in the two metagenome data-sets have been investigated by Otto Cordero and the results are presented in this chapter.

## **II.2 Materials and methods**

### **II.2.1 Sampling and sequencing**

In December 2014, 46 granules were sampled from the LBE routine AGS SBR lab-scale reactor by Ela Sliwerska and Otto Cordero (Microbial Population Biology Group, IBP - D-USYS, Swiss Federal Institute of Technology in Zurich). These granules were sequenced individually with Illumina Mi-Seq technology in paired-end mode, providing an average of 2x 190 282 sequences per granule with reads up to 300 base pairs (bp) long [Leventhal et al., 2018]. This first metagenome will be referred to as the metagenome O1. In May 2015, a new set of 96 granules was sampled from the same reactor. These granules were sequenced individually with Illumina Hi-Seq technology in paired-end mode, providing on average 2x 855 435 sequences of 100 bp per granule and will be referred to as the metagenome O2.

### **II.2.2 Pre-processing of the sequences**

The same pipeline was used with the two metagenomes (O1 and O2). The sequences were trimmed and filtered using Trimmomatic v. 0.32 [Bolger et al., 2014] with a sliding window of 10 bp, a minimum average quality threshold of 15 and a total minimum size of 50 bp. Identical sequences were de-replicated using the script dedup.sh [Bushnell, 2014]. In order to merge the overlapping paired sequences, every pair was aligned using muscle v. 3.8.31 [Edgar, 2004]. The paired sequences with an overlap bigger or equal to 20 bp and less than 5 mismatches were merged.

### **II.2.3 Estimation of the relative abundance of the different bacteria**

The pre-processed sequences were mapped against the 16S rRNA gene databases MiDAS v. S123\_2.1.3 [McIlroy et al., 2017] using bowtie2 [Langmead and Salzberg, 2012] in local mode with a maximum of 100 different mapping for each query sequence, an insert size going from 0 to 3000 bp and one mismatch allowed during multiseed alignment. The mapping with an alignment score over query length ratio lower than 1.71 were discarded and the number of remaining primary mapping per genus was computed.

### **II.2.4 Estimation of the variation of the mean abundance of *Ca. Accumulibacter* depending on the number of granules considered**

An estimation of the variability of *Ca. Accumulibacter* in function of the sampling effort was made based on the relative abundance of this genus in the individual granules from the metagenomes O1 and O2 separately. A defined number of granules were randomly picked (e.g., 3, 10 or 20), *in silico*, and the mean of the *Ca. Accumulibacter* proportions in these granules was computed. This was repeated 1000 times for each metagenome.

## II.2.5 Correlation calculations between bacteria and phages

The pre-processed sequences were mapped against the NCBI phage database (2018), complemented with 6 EBPR podovirus sequences [Skenner et al., 2011], using bowtie2 with the parameters mentioned above. The number of hits was merged with the number of hits with the bacterial 16S rRNA genes. The proportions of hits were divided with the total of hits for phages and bacteria. Pearson's correlation matrices and the corresponding p-values were computed in R [R Development Core Team, 2008] with the package Hmisc [Harrell Jr et al., 2018].

## II.3 Results

### II.3.1 Bacterial community composition of individual granules

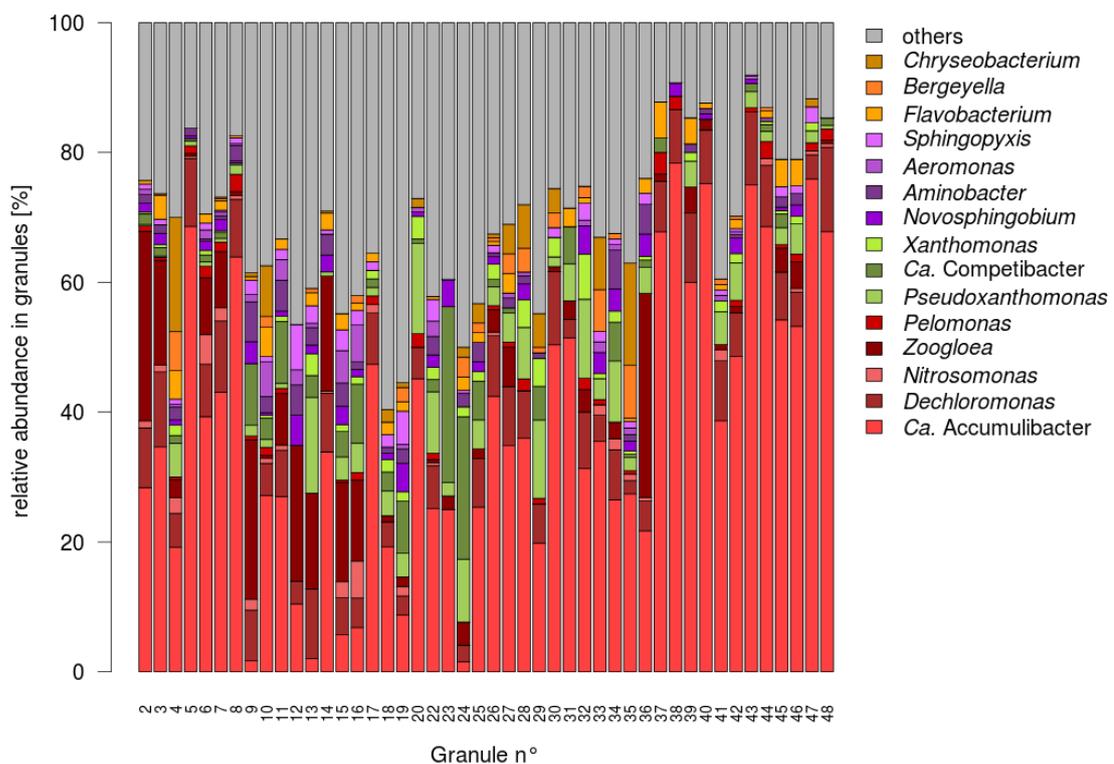


Figure II.1 – Bacterial community composition of the individual granules from metagenome O1. The 15 most abundant genera are shown.

The sequences from the metagenomes O1 and O2 were mapped on the 16S rRNA bacterial genes of the MiDAS database in order to estimate the proportion of the different genera in the individual granules. In the metagenome O1, the Gammaproteobacteria was the most represented class, in particular, the Betaproteobacteria were representing more than 33 % of the sequences in most of the granules (Figure II.1). The classes of Alphaproteobacteria and Flavobacteriia were also abundant in this metagenome.

## Chapter II. High variability of microbial community composition among individual granules

The bacterial community composition was very different from one granule to another, especially concerning the PAO *Ca. Accumulibacter* whose relative abundance was more than 60 % in some granules and less than 3 % in others. The proportion of the GAO *Ca. Competibacter* was also very different from one granule to the other and was higher in the granules where the proportion of *Ca. Accumulibacter* was low.

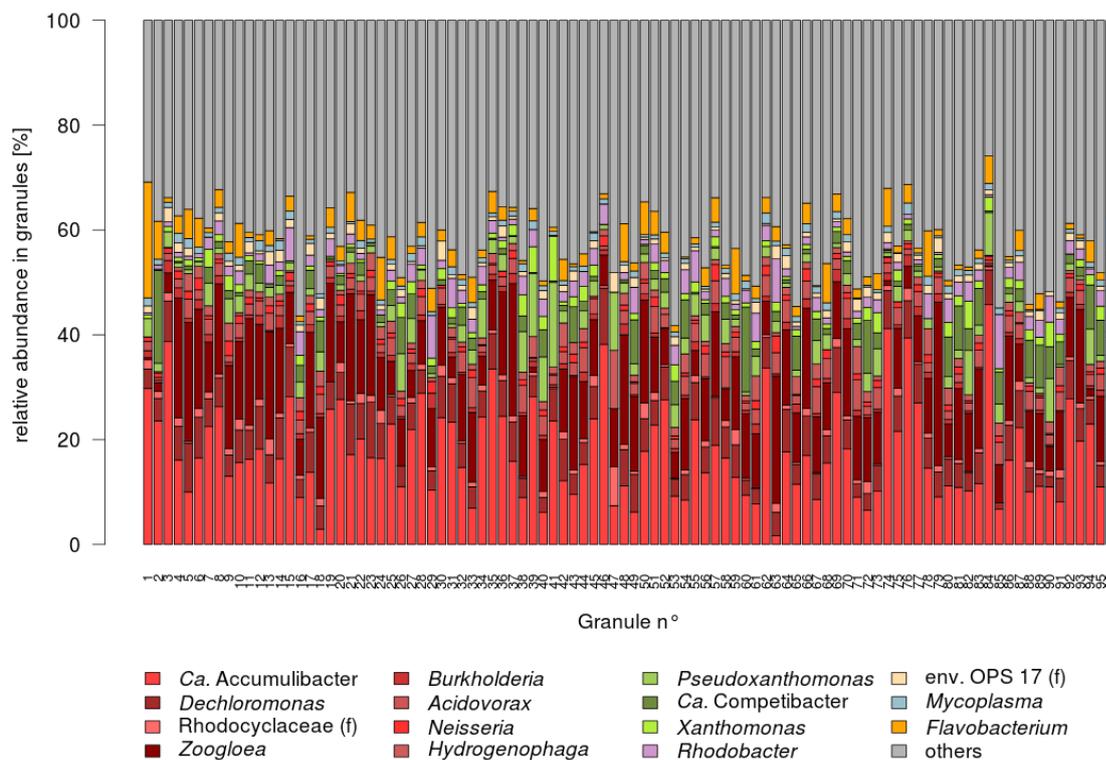


Figure II.2 – Bacterial community composition of the individual granules from metagenome O2. The 15 most abundant genera are shown.

In the metagenome O2, the microbial composition of the granules was similar to the one of O1 with few differences (Figure II.2). The Gammaproteobacteria was, also in this metagenome, the dominant class, with an important proportion of Betaproteobacteria. The classes of Alphaproteobacteria, Flavobacteriia, Sphingobacteriia, and Mollicutes had representatives in the most abundant genera of O2. The proportion of *Ca. Accumulibacter* varied between the different granules, yet with less amplitude than in the granules of the metagenome O1. The diversity in the order of Betaproteobacteriales is higher in these granules, with 8 genera from this order among the most abundant ones. *Zoogloea* represents around 10 % of the bacterial sequences in O2, while its abundance was more variable in O1.

### II.3.2 Estimation of the variability of *Ca. Accumulibacter*

In order to estimate the reproducibility of the estimation of the bacterial community, *in silico* sampling were performed among the granules of metagenomes O1 and O2 separately. This

sampling simulation was performed on the proportion of *Ca. Accumulibacter* because it was the most variable across each metagenome.

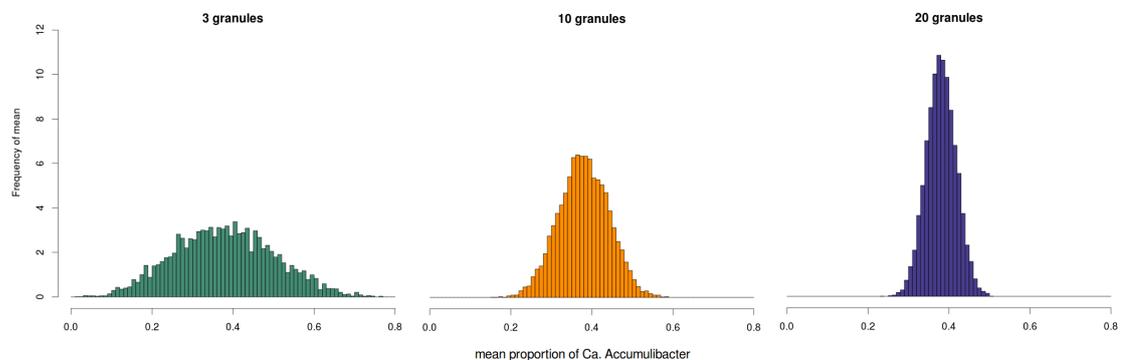


Figure II.3 – Distribution of the mean proportion of *Ca. Accumulibacter* obtained from 1000 random *in silico* sampling of 3, 10 and 20 granules from metagenome O1.

The distribution of the mean proportion of *Ca. Accumulibacter* calculated from 3, 10 or 20 granules from the metagenome O1 is shown in Figure II.3. The reproducibility of an estimation of the proportion of *Ca. Accumulibacter* from 3 granules is very weak and the probability to have a difference higher than 20 % between this value and the 'real' mean proportion was estimated at 56 %. With 10 granules, the estimation of the mean proportion of *Ca. Accumulibacter* is still not reliable, with 22 % chance to have an 'error' bigger than 20 %. With 20 granules, the estimation of the mean proportion of *Ca. Accumulibacter* starts to become stable with a probability to have an 'error' higher than 20 % dropping to 4 %.

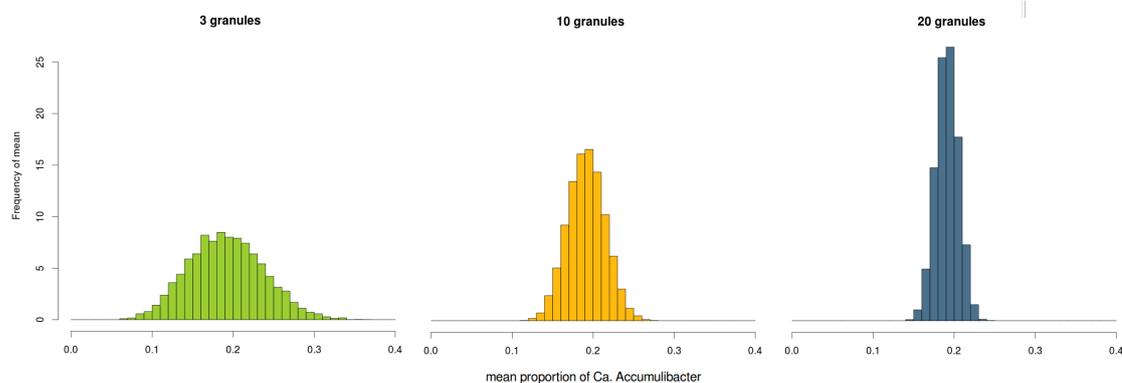


Figure II.4 – Distribution of the mean proportion of *Ca. Accumulibacter* obtained from 1000 random *in silico* sampling of 3, 10 and 20 granules from metagenome O2.

The distribution of the mean proportion of *Ca. Accumulibacter* calculated from 3, 10 or 20 granules from the metagenome O2 is shown in Figure II.4. The variability of this proportion is significantly lower in this metagenome than in O1. Yet, the probability to have an 'error' higher than 20 % on the estimation of this proportion is around 10 % with 10 granules. This probability drops to 0.4 % with 20 granules.

II.3.3 Correlations between phages and bacteria

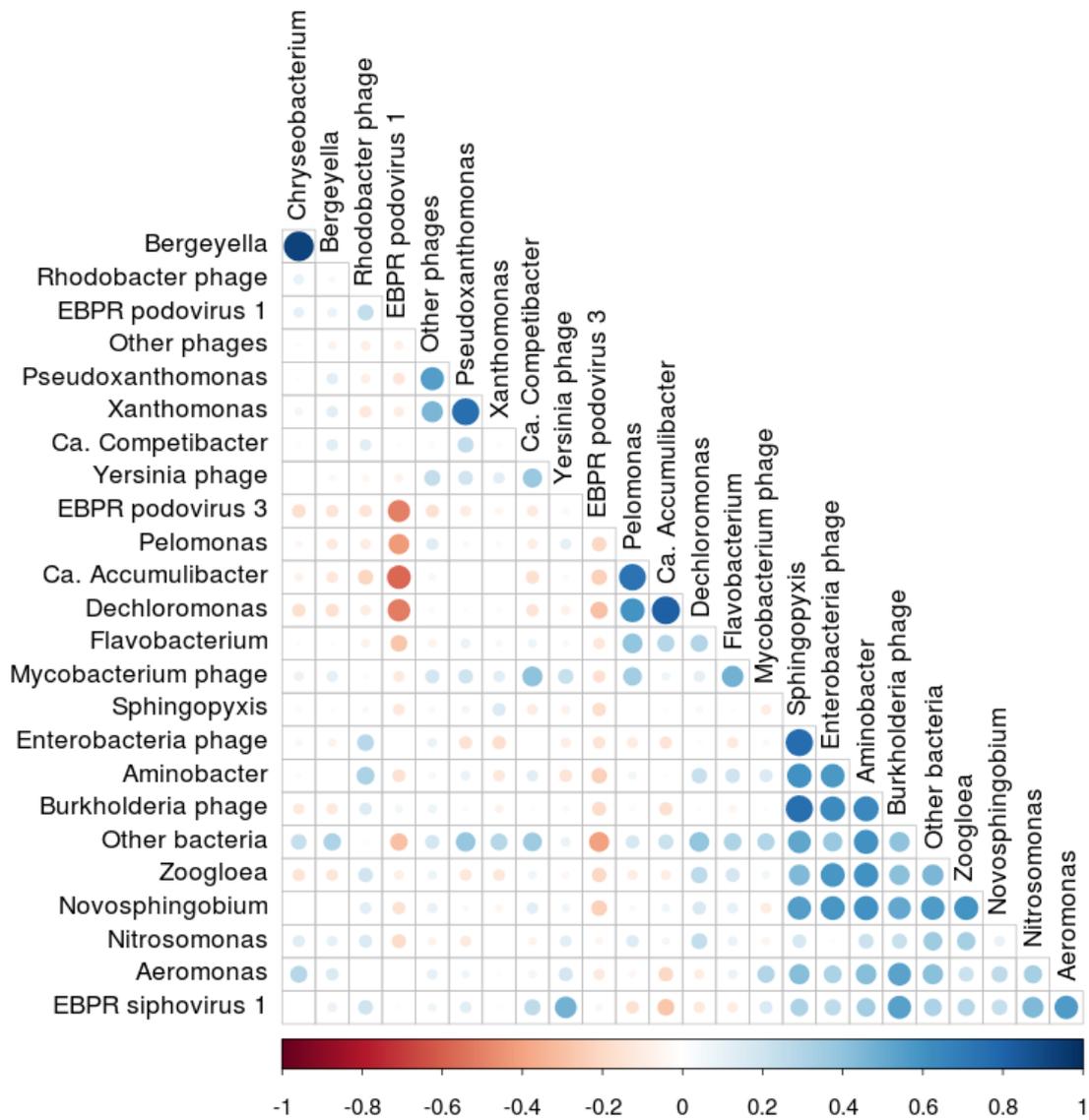


Figure II.5 – Correlations between phages and bacteria-related sequences in the metagenome O1. The tables of abundance (Table A.1), numerical values (Table A.2) and the p-values (Table A.3) are in provided in electronic supplementary material.

Significant positive correlations were found between phylogenetically related bacteria in metagenome O1 (Figure II.5): *Chryseobacterium* and *Bergeyella* ; *Ca. Accumulibacter*, *Dechloromonas* and *Pelomonas* ; *Pseudoxanthomonas* and *Xanthomonas* ; *SpHINGOPYXIS*, *Aminobacter* and *Aeromonas*. The bacteria from the latest group has also significant positive correlations with *Zoogloea*, the *Enterobacteria* phage and the *Burkholderia* phage. EBPR podovirus 1 (EPV1) was significantly negatively correlated with *Ca. Accumulibacter*, *Dechloromonas*, *Pelomonas* and EBPR podovirus 3 (EPV3). Between 7 (granule n°37) and 16555 (granule n°19) sequences mapping EPV1 and between 2 (granules n°3 and 12) and

12731 (granule n°36) sequences mapping EPV3 were found in metagenome O1. EPV3 was negatively but not significantly correlated to all the main bacteria.

The correlation between the estimated proportions of the most abundant phages and bacteria were computed to identify putative positive or negative influences.

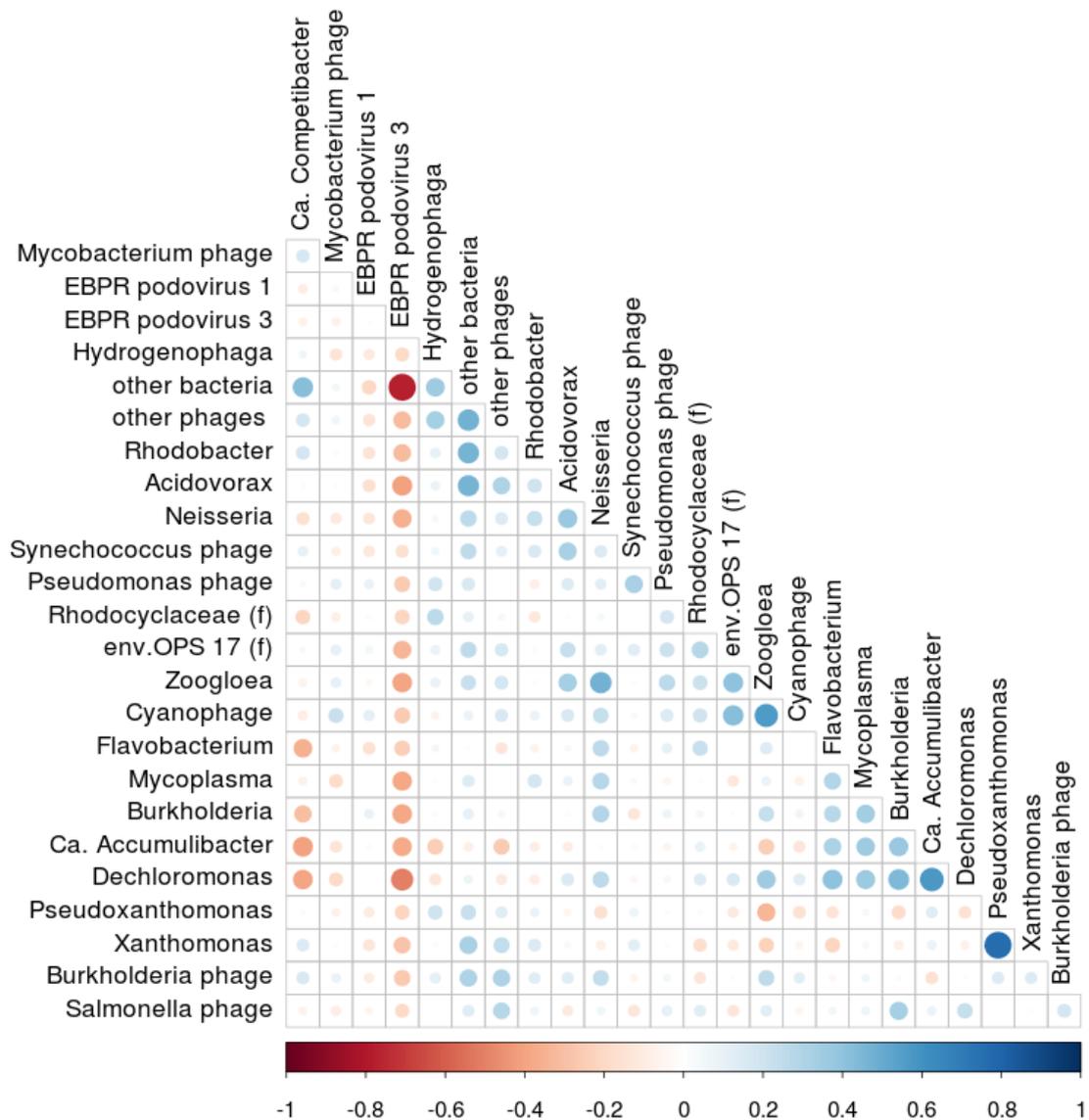


Figure II.6 – Correlations between phages and bacteria-related sequences in the metagenome O2. The tables of abundance (Table A.4), numerical values (Table A.5) and the p-values (Table A.6) are provided in electronic supplementary material.

The same positive correlations between phylogenetically related bacteria were found in the metagenome O2 (Figure II.6): *Pseudoxanthomonas* and *Xanthomonas*; *Ca. Accumulibacter* and *Dechloromonas*. On the other hand, there was no significant negative correlations between

## Chapter II. High variability of microbial community composition among individual granules

---

EPV1 and *Ca. Accumulibacter* or *Dechloromonas* in metagenome O2. Between 0 and 400 sequences mapping EPV1 and between 0 and 59332 (granule n°1) sequences mapping EPV3 were found in metagenome O2. A negative correlation between EPV3 and 'other bacteria' suggests that the bacteria which is responsible for this negative correlation is not a member of one of the 15 most abundant genera. In this metagenome, *Ca. Accumulibacter* and *Ca. Competibacter* have a significant negative correlation.

### II.4 Discussion

Working with granular sludge has the advantage that the quantity of biomass is generally not a limiting factor for DNA analysis. A single granule of 0.2 mm would already provide enough DNA for molecular analysis such as whole genome sequencing. The bacterial communities in the individual granules, assessed at the genus level, were highly variable, in particular in the metagenome O1. This brings the question of minimal amount of granules to collect and homogenize in order to assess the bacterial community of the AGS. *In silico* sampling showed that the probability to estimate the 'real' mean proportion of *Ca. Accumulibacter* with an 'error' higher than 20 % from three replicates is higher than 50 %. With 20 granules, this probability is below 5 % and therefore a minimum of 20 granules, was set as the minimal quantity to sample and homogenize to have a representative sample for the forthcoming analysis of this study. This *in silico* sampling was a first rough estimation based on the hypothesis that all the granules provide the same quantity of DNA. Yet it highlighted the importance of the quantity of biomass that are homogenized for molecular analyses. For activated sludge, a very precise protocol for DNA extraction from activated sludge is publicly available [MiDAS, 2015], giving precise standard volumes from the sample collection and homogenization. Unfortunately, nothing similar is available for granular sludge, even though the differentiation of the microbial community is likely higher in granular sludge than in conventional activated sludge. Moreover, the majority of publications on granular sludge microbial communities do not mention the number of granules or the volume of biomass homogenized before DNA extraction. The few studies providing information about this quantity reveal heterogeneous habits concerning the quantity of starting material [Lawson et al., 2017, Oyserman et al., 2016, Zhang et al., 2017a]. Yet, since this factor potentially has a big impact on the reproducibility of the results, it would be important to be described in detail in future reports.

The assessment of the bacterial community composition was performed by mapping the reads of the metagenome on the MiDAS 16S rRNA gene database. The results must be handled with caution. If the metagenome can provide quantitative values [Xia et al., 2014], the number of 16S rRNA genes contained in the genome of a single bacteria was shown to vary between 1 and at least 15 [Acinas et al., 2004]. The choice of the database also influences the results, several databases were tested (data not shown) and the MiDAS database which is based on the SILVA 123 NR99 including full length 16S rRNA genes sequences from microorganisms abundant in wastewater treatment, did provide a higher resolution than other databases such as green genes [DeSantis et al., 2006] and Silva [Quast et al., 2012].

The microbial community revealed by these two metagomes was typical of an AGS from a lab-scale reactor fed with a wastewater containing acetate and propionate as carbon source. This type of synthetic wastewater favors Gammaproteobacteria and in particular the PAO *Ca. Accumulibacter* and the GAO *Ca. Competibacter* [de Kreuk and van Loosdrecht, 2004, Gonzalez-Gil and Holliger, 2014, Leventhal et al., 2018]. No Actinobacteria were detected in the 15 most abundant genera of the two metagenomes. These bacteria are known to be difficult to lyse and are often underestimated by sequencing-based methods due to too gentle DNA extraction methods [Albertsen et al., 2015]. The extraction method used with these metagenomes did not have a bead beating step and is therefore likely to extract less efficiently the DNA of gram-positive bacteria. It is also possible that Actinobacteria were not abundant in this AGS because they prefer other substrate than acetate and propionate [Kristiansen et al., 2013, McIlroy et al., 2018].

The bacterial community in the individual granules was more variable in the metagenome O1 than in O2. In the two metagenomes, the proportion of *Ca. Accumulibacter* had the largest variation amplitude. This phenomenon could be due to the presence of a lysogenic phage, EPV1, whose abundance in the granules was also highly variable and negatively correlated with the proportion of *Ca. Accumulibacter*. EPV1 is thought to be specific to *Ca. Accumulibacter* and has the genomic potential to repress the CRISPR-Cas defense system of its host [Skennerton et al., 2011]. A total of 16 555 sequences mapping EPV1 were found in the metageome of one of the granules suggesting that the samples may have been collected in the middle of a phage infection event, with the *Ca. Accumulibacter* of several granules having already been infected, a lysogenic event being past, ongoing or not yet begun depending on the granules. The variability of *Ca. Accumulibacter* was lower in the metagenome O2 and there a maximum of 400 EPV1-related sequences in a granule were recovered. Moreover, no correlation was detected between *Ca. Accumulibacter* and EPV1 in this metagenome. This gives support to the hypothesis that a lysogenic event ongoing or past in the granules of O1 was increasing the variability of the proportion of the dominant genus *Ca. Accumulibacter* and consequently of the whole bacterial communities of the granules. The granules of O2 were then collected long enough after this bacteriophage attack, when the microbial communities in the granules had time to stabilize. Leventhal et al. [2018] revealed that the *Ca. Accumulibacter* population was mostly composed of a dominant strain related to clade IA in some granules or IIA in others. Yet, differences were noticed in the phage-defense system of the different *Ca. Accumulibacter* clades [Albertsen et al., 2012]. It is thus possible that the different clades are more or less sensitive to the different bacteriophages.

## II.5 Conclusion

The composition of the bacterial communities of individual granules can be highly variable. *In silico* sampling in the estimated bacterial communities proved to have a low reproducibility below 20 granules. This motivates the definition of a minimal sample size for future experiments presented in this thesis. At least 1 ml of biomass containing at least 20 granules were collected

## **Chapter II. High variability of microbial community composition among individual granules**

---

in order to obtain a representative sample of the overall bacterial reactor biomass. Moreover, the quantity of homogenized granules used as starting material for molecular analysis is worth to mention in publications on the microbial communities in granular sludge, since it can have a strong impact on the reproducibility of the results. Ideally, this quantity would be set up by a standard protocol.

The variability of the bacterial communities was particularly high in the first metagenome O1 mainly due to the variability of the dominant genus *Ca. Accumulibacter*. The negative correlation between sequences of the latter and sequences of the bacteriophage EPV1 indicated that this variability was due to an ongoing attack on *Ca. Accumulibacter* by this bacteriophage with different infection stages in the different granules.

POURQUOI EUX  
ONT-ILS DROIT  
À DU VRAI CACA?





# III Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater

## III.1 Preamble

This chapter is the result of our collaboration with Eawag and is similar to the postprint version of the article published in Water Research X in 2019, ( [doi.org/10.1016/j.wroa.2019.100033](https://doi.org/10.1016/j.wroa.2019.100033)).

### III.1.1 List of authors

The authors contributing to the above mentioned article are Manuel Layer<sup>a,b</sup>, Aline Adler<sup>c</sup>, Eva Reynaert<sup>a,c</sup>, Antonio Hernandez<sup>a,c</sup>, Marco Pagni<sup>d</sup>, Eberhardt Morgenroth<sup>a,b</sup>, Christof Holliger<sup>c</sup>, Nicolas Derlon<sup>a</sup>.

<sup>a</sup> Eawag Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland

<sup>b</sup> ETH Zürich, Institute of Environmental Engineering, 8093 Zürich, Switzerland

<sup>c</sup> EPFL Ecole Polytechnique Fédérale de Lausanne, ENAC IIE Laboratory for Environmental Biotechnology, 1015 Lausanne, Switzerland

<sup>d</sup> SIB Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

### III.1.2 Authors contributions

The candidate and Manuel Layer have contributed equally to this work.

ML, AA, EM, CH and ND designed the study. ML, ER and AH operated the AGS reactors, performed the sampling and the collection of data, AA did the molecular analysis of the

### Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater

---

microbial communities. ML and AA analyzed the data and wrote the manuscript. ER, AH, MP, EM, CH and ND contributed to the data analysis. AA performed the statistical analysis with the advise of MP. EM, CH and ND reviewed and provided valuable edits to the manuscript.

## III.2 Introduction

Aerobic granular sludge (AGS) systems have been developed over the past 20 years and now offer a relevant alternative over conventional activated sludge systems [Morgenroth et al., 1997]. Advantages of AGS include enhanced settling properties of the sludge, a high suspended solid concentration and the co-existence of different redox conditions across the granules, which result in significant energy, footprint and chemical savings [Khan et al., 2015]. Worldwide, more than 40 full-scale plants are now in operation, treating a wide range of municipal and industrial wastewaters [Pronk et al., 2017]. However, the performance and/or granulation process of AGS systems are often hampered by the wastewater composition [de Kreuk and van Loosdrecht, 2006, Guimaraes et al., 2017, Guimarães et al., 2018]. To optimize the performances of such systems, it is therefore required to understand the link between the influent composition and the granule formation.

Lab-scale sequencing batch reactors (SBR) have been extensively used to develop our fundamental understanding of AGS systems [de Kreuk and van Loosdrecht, 2004, He et al., 2016b, Weissbrodt et al., 2013]. Those studies were mainly conducted using high concentrations of volatile fatty acids (VFA) (e.g., acetate and/or propionate) and phosphorus. The key role of anaerobic feast and aerobic famine conditions on the granule formation was identified [de Kreuk and van Loosdrecht, 2004]. These conditions favour the growth of slow-growing organisms like polyphosphate- (PAO) and glycogen-accumulating organisms (GAO), which have been identified as key players in granulation [de Kreuk and van Loosdrecht, 2004]. The growth of PAO and GAO, and ultimately the granulation, is improved by the presence of soluble organic carbon (fermented or not) in the influent. The selective uptake of these substrates by PAO and GAO outcompetes ordinary heterotrophic organisms (OHO). OHO growth hampers the formation of granular biomass or the nutrient-removal performances, while it also promotes the formation of flocs [de Kreuk et al., 2010, Novak et al., 1993, Pronk et al., 2015a, Weissbrodt et al., 2014a]. If the growth of PAO and GAO is crucial for the formation of aerobic granules during treatment of VFA-rich influent, it is then intuitive that granulation might be hampered during treatment of municipal wastewater containing high particulate organic substrate ( $X_B$ ) and low VFA fractions.

A key aspect in understanding AGS systems is in characterizing the microbial community composition and understanding how it influences the granulation process. The microbial communities of AGS systems fed with 100%-VFA WW are well described in literature, and dominated by Gammaproteobacteria, in particular, the PAO *Candidatus (Ca.) Accumulibacter* and the GAOs from the Competibacteraceae family [He et al., 2016a, Henriet et al., 2016, Weissbrodt et al., 2013]. Most of these bacteria have a metabolism adapted to VFA uptake

under anaerobic conditions. It is also likely that those bacteria are not able to ferment most sugars and amino acids or to hydrolyze polymers [Kong et al., 2006, Marques et al., 2017]. So far, only few studies have characterized the microbial communities of AGS treating WW containing  $X_B$ . Hence, the core microbial community of these AGS has not been identified yet [Kang et al., 2018, Swiatczak and Cydzik-Kwiatkowska, 2018, Szabo et al., 2017b, Wang et al., 2018]. Fermentative and hydrolyzing bacteria are expected to be abundant in such systems, similarly to enhance biological phosphorus removal (EBPR) systems treating municipal WW [Kong et al., 2008]. The fermentative PAO *Tetrasphaera* does not store VFA in the form of intracellular polyhydroxyalkanoates (PHA) and is usually more abundant than *Ca. Accumulibacter* in Danish EBPR WW treatment plants [Mielczarek et al., 2013]. *Tetrasphaera* can take up orthophosphate aerobically after anaerobic storage of different carbon sources like amino acids and glucose [Nguyen et al., 2011]. *Micropruina* is also commonly found in EBPR activated sludge [Saunders et al., 2016, Stockholm-Bjerregaard et al., 2017]. *Micropruina* is a fermentative GAO, able to take up and ferment various carbon sources anaerobically to constitute glycogen reserves [McIlroy et al., 2018]. It is however unclear to what extent *Tetrasphaera* and *Micropruina* play a role in the formation of AGS during treatment of complex WW with a high  $X_B$  content. A key aspect of this study will be to characterize the microbial communities found in AGS systems fed with WW containing different fractions (and types) of  $X_B$ . Another objective will be to identify correlations between these communities and the sludge settling properties and nutrient removal performance.

If the WW composition influences the microbial community, it is reasonable to expect that the granulation process of AGS systems is also impacted (e.g. physical properties of biomass and start-up kinetics). A harsh selection of fast settling biomass in lab-scale reactors fed with 100%-VFA synthetic WW resulted in rapid granulation within two weeks in previous studies [Mosquera-Corral et al., 2011, Weissbrodt et al., 2013]. But nutrient-removal was impaired for weeks to months. Lochmatter and Holliger [2014] successfully started up an AGS system within 28 days without loss of nutrient removal, by applying more gentle washout of slow settling biomass and by adapting the organic loading during the early stages of start-up. Start-up of AGS systems with municipal WW can be significantly longer. Harsh selection pressure on slow-settling biomass can lead to fast granulation (e.g., 20 days) [de Kreuk and van Loosdrecht, 2006] but the time reported to transform activated sludge into AGS while maintaining the nutrient-removal performance have normally been much longer (40-400 days) [Derlon et al., 2016, Giesen et al., 2013, Liu et al., 2010]. Therefore, this study should aim to clarify the link between start-up kinetics and influent WW composition, while similar operating conditions are applied and the same inoculum activated sludge is used.

The diffusibility and uptake rate of organic carbon directly influences the microbial competition for substrate, and in turn the granulation (Figure III.1). A slow anaerobic conversion of non-diffusible  $X_B$  combined with a decreased substrate availability within the granule can result in carbon leakage (i.e., carbon available in aerobic conditions). Carbon leakage favours OHO growth to the detriment of PAO, GAO and fermenters, and ultimately results in floc formation [Jabari et al., 2016, Larsen and Harremoës, 1994, Morgenroth et al., 2002, Suresh

**Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater**

et al., 2018, Wagner et al., 2015b]. For municipal WW, non-diffusible  $X_B$  usually represents 50 % of the total influent chemical oxygen demand (COD) [Metcalf & Eddy, 2014]. Based on the current knowledge, the formation of AGS during treatment of  $X_B$  rich WW might be hampered. In addition, it remains unclear whether flocs are detrimental to AGS systems, when non-diffusible  $X_B$  represents a high proportion of the influent COD.

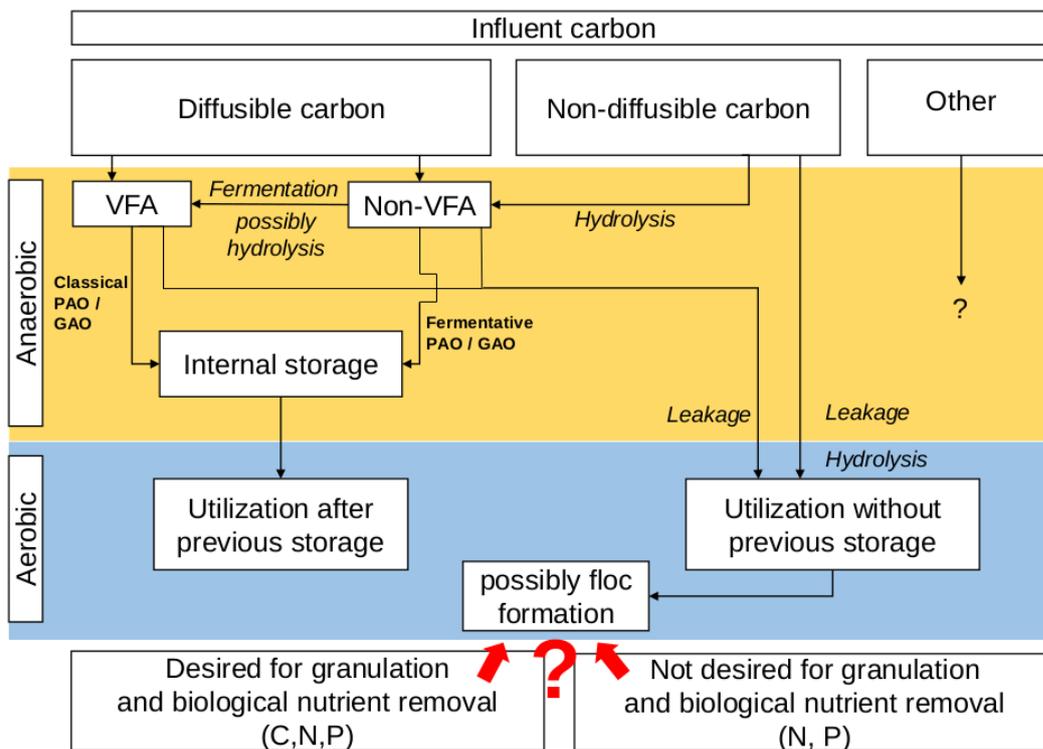


Figure III.1 – Conceptual model of carbon utilization and proposed desired/undesired paths in AGS systems, given plug-flow anaerobic feeding and subsequent aerobic fully mixed conditions.

The main goal of this study was to understand the link between influent WW composition, microbial community, physical AGS parameters and nutrient removal performance. The specific research questions were to better understand how the WW composition in terms of diffusible and non-diffusible organic substrates (i) influences the overall microbial community development, (ii) divides the microbial community between flocs and granules, (iii) governs nutrient removal, (iv) defines physical characteristics such as settling properties, sludge morphology, and (v) influences the success and duration of start-up of AGS systems when similar operating conditions are applied. Four lab-scale SBR were inoculated with the same activated sludge and operated for over 400 days in parallel. Four distinct WW were used: 100%-VFA synthetic, complex synthetic, municipal primary effluent, and municipal raw WW. The sludge properties (morphology, concentration, SVI, and size distribution), reactor performances (C, N, P and total suspended solids (TSS) removal), and microbial community

composition of the flocs and granules were monitored by 16S rRNA amplicon sequencing.

## III.3 Materials and methods

### III.3.1 Experimental approach

Four sequencing batch column reactors (SBR) were operated in parallel for 400 days and fed with four different wastewaters : 100%-VFA (acetate, propionate) synthetic WW (R1), complex synthetic WW (R2), primary effluent municipal WW (R3), and raw municipal WW (R4). Those four WW mainly differed with regards of the carbon source, i.e., concentrations in volatile fatty acids, soluble and particulate organic substrates (Table III.1). After approximately three months of operation, R4 was restarted due to complete sludge loss. Data of R4 (fed with raw WW) are thus shown for the first run (run#1) and the second run (run#2).

### III.3.2 Experimental set-up

The four SBR comprised a mixed liquor volume of 12.9 L (height-to-diameter ratio 8.4), and were operated in simultaneous fill-draw mode. The SBR cycles consisted of the following phases: (i) anaerobic phase (90 min), (ii) aerobic phase (240 min), (iii) settling (duration see description below) and (iv) selective excess sludge withdrawal (60 s), with a total cycle length of 5.6 hours (4.3 cycles per day). The anaerobic phase comprised an anaerobic plug-flow feeding (PF) and an anaerobic idle. The latter was changed to anaerobic mixing on day 357 and 289 for reactors R3 and R4 run #2, respectively. The WW upflow velocity during PF feeding ( $v_{ww}$ ) was set to  $0.25 \text{ m h}^{-1}$  (anaerobic PF + idle) and  $0.38 \text{ m h}^{-1}$  (anaerobic PF + mixing). The volume exchange ratio (VER) was set to 0.3. The oxygen concentration during the aerobic phase was controlled with a setpoint of  $2.00 \text{ mg O}_2 \text{ l}^{-1}$ . Mixing was provided by a mechanical stirrer during anaerobic PF+mixing, and by aeration during aerobic conditions. All SBR were equipped with oxygen sensors (Optical LDO, Endress & Hauser, Switzerland). Both sensors were connected to a programmable logic controller (PLC), which was controlled and monitored by a supervisory control and data acquisition (SCADA) system. All reactors were inoculated with activated sludge from the WW treatment plant (WWTP) of Thunersee, Switzerland, which performs biological carbon, nitrogen and phosphorus removal.

### III.3.3 Start-up approach

The start-up approach was based on the strategy developed by Lochmatter and Holliger [2014]. The selective pressure on slow settling biomass was first maintained at a low level, in order to prevent too high washout stress. This was achieved by slowly increasing the critical settling velocity ( $v_{crit}$ ) from  $1.7$  to  $5.1 \text{ m h}^{-1}$ . An increase in  $v_{crit}$  was iteratively reassured by SRT calculations (Equation III.1).  $V_r$  is the reactor volume (l),  $TSS_r$  is the TSS concentration in the reactor ( $\text{gTSS l}^{-1}$ ),  $Q_{ex}$  is the flow rate of excess sludge ( $\text{l d}^{-1}$ ),  $TSS_{ex}$  is the TSS concentration

### Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater

---

of the excess sludge ( $\text{gTSS l}^{-1}$ ),  $Q_{\text{eff}}$  is the flow rate of effluent ( $\text{l d}^{-1}$ ), and  $\text{TSS}_{\text{eff}}$  is the TSS concentration in the effluent ( $\text{gTSS l}^{-1}$ ).

$$\text{SRT} = \frac{V_r \cdot \text{TSS}_r}{Q_{\text{ex}} \cdot \text{TSS}_{\text{ex}} + Q_{\text{eff}} \cdot \text{TSS}_{\text{eff}}} \quad (\text{III.1})$$

If SRT was  $< 20$  d,  $v_{\text{crit}}$  was decreased again. The procedure was repeated on a weekly basis. Also, long anaerobic and aerobic phases were applied (total cycle duration 5.6 h), in order to improve anaerobic COD uptake and aerobic nutrient removal. Finally,  $v_{\text{ww}}$  was kept low ( $0.25 - 0.38 \text{ m h}^{-1}$ ), in order to provide a high substrate gradient during anaerobic plug-flow feeding and improve anaerobic COD uptake.

#### III.3.4 Successful start-up definition

We here define successful start-up of AGS on both physical properties and substrate removal. Specifically, the settling parameters  $\text{SVI}_{30} < 90 \text{ ml g}^{-1}$  and  $\text{SVI}_{30/10}$  -ratio  $> 0.8$ , the size fraction  $d > 0.25 \text{ mm}$  constituting at least 50 % of TSS, granule appearance based on microscopic images, and stable substrate and nutrient removal. The selection of the chosen parameters and values were selected based on previous experience on the on start-up of AGS systems during for the treatment of low strength municipal WW at Eawag [Derlon et al., 2016, Wagner et al., 2015b] and by other researchers/practitioners [Coma et al., 2012, de Kreuk and van Loosdrecht, 2006, Giesen et al., 2013, Liu et al., 2011, Ni et al., 2009, van der Roest et al., 2011, Wagner et al., 2015a]. The definition is in line with the original definition of AGS [de Kreuk et al., 2007]).

#### III.3.5 Wastewater composition and sludge inoculum

The detailed influent composition of R1, R2, R3 and R4 are shown in Table III.1. All WW were in the range typical of low-to-medium strength WW [Metcalf & Eddy, 2014]. Synthetic substrates comprised a total carbon:nitrogen:phosphorus ratio of approximately 100:7:1. Acetate (Ac) and propionate (Pr) were used as sole carbon source for the 100%-VFA synthetic WW (50 % of COD each). Complex synthetic WW was composed of 1/3 VFA (1/6 acetate + 1/6 propionate), 1/3 soluble fermentable substrates (1/6 glucose, 1/6 amino acids) and 1/3 particulate substrates (1/6 peptone, 1/6 starch). Particulate substrates were peptone from gelatin, enzymatic digest (Fluka Analytical, Switzerland), and starch made from wheat (Merck KGaA, Germany). Amino acids were composed of L-alanine, L- arginine, L-aspartic acid, L-glutamic acid, L-leucine, L-proline and glycine in equal COD-equivalents. These individual amino acids were chosen according to the most abundant amino acids present in the peptone used. Added nitrogen was composed of soluble  $\text{NH}_4$  -N for the system with 100%-VFA synthetic WW, but included nitrogen from peptone and amino acids for complex synthetic

### III.3. Materials and methods

Table III.1 – Measured influent composition of the four SBR fed by 100%-VFA synthetic WW, complex synthetic WW, primary effluent WW and raw WW, specific substrate recipe of R1 and R2 influent are given in supplementary material Table A.7 <sup>a</sup>

Reactor	100%-VFA synthetic WW R1	complex synthetic WW R2	primary effluent WW R3	raw WW R4 run #1	raw WW R4 run #2
Total COD [mg O <sub>2</sub> l <sup>-1</sup> ]	582 ± 65	503 ± 61	331 ± 97	808 ± 42	469 ± 151
Soluble COD [mg O <sub>2</sub> l <sup>-1</sup> ]	582 ± 65	457 ± 73	188 ± 76	271 ± 109	247 ± 121
Particulate COD [mg O <sub>2</sub> l <sup>-1</sup> ]	0 ± 92	46 ± 95	143 ± 123	537 ± 441	222 ± 194
VFA [mg O <sub>2</sub> l <sup>-1</sup> ]	582 ± 65 <sup>c</sup>	170 ± 26 <sup>c</sup>	26 ± 17 <sup>b</sup>	- <sup>d</sup>	40 ± 28 <sup>e</sup>
Ac + Pr [mg O <sub>2</sub> l <sup>-1</sup> ]	582 ± 65	170 ± 26	15 ± 9	- <sup>d</sup>	17 ± 11
Ac + Pr/ total <sub>COD</sub> ratio	1.00	0.33	0.05	- <sup>d</sup>	0.06
Total nitrogen [mg N l <sup>-1</sup> ]	43 ± 10	44 ± 11	33 ± 9	30 ± 7	41 ± 19
NH <sub>4</sub> -N [mg N l <sup>-1</sup> ]	40 ± 8	20 ± 5	24 ± 6	25 ± 4	29 ± 10
Total phosphorus [mg P l <sup>-1</sup> ]	5.4 ± 0.9	5.4 ± 1.7	3.3 ± 0.9	3.2 ± 0.5	4.4 ± 1.9
PO <sub>4</sub> -P	5.0 ± 1.1	4.7 ± 0.8	2.3 ± 0.5	2.6 ± 0.4	2.7 ± 0.8

<sup>a</sup> number of data 24-38 for R1, R2, R3, 13-15 for R4 run #1, 13-22 for R4 run #2

<sup>b</sup> VFA composition of municipal primary effluent WW: 16 % Acetate, 41 % Propionate, 43 % longer-chained VFAs (COD based)

<sup>c</sup> VFA composition of synthetic WW: 50 % acetate, 50 % propionate (COD based)

<sup>d</sup> VFA composition of municipal raw WW run #1 was not measured

<sup>e</sup> VFA composition of municipal raw WW run #2: 13 % Acetate, 52 % Propionate, 35 % longer-chained VFAs (COD based)

WW. Phosphorus was composed of soluble PO<sub>4</sub>-P species for both synthetic WW. In order to prevent bacterial growth in the synthetic substrate storage bottles, the phosphorus species and diluent water were stored separately from the nitrogen and carbonaceous species and mixed automatically before each cycle [Ebrahimi et al., 2010]. The nitrogen and carbonaceous species were prepared in 20-fold concentration in portions of 5 l and, after addition of 50 ml of a trace-element solution with the following concentrations : 16.22 g l<sup>-1</sup> of C<sub>10</sub>H<sub>14</sub>N<sub>2</sub>Na<sub>2</sub>O<sub>8</sub> · 2H<sub>2</sub>O, 0.44 g l<sup>-1</sup> of ZnSO<sub>4</sub> · 7H<sub>2</sub>O, 1.012 g l<sup>-1</sup> of MnCl<sub>2</sub> · 6 H<sub>2</sub>O, 7.049 g l<sup>-1</sup> of (NH<sub>4</sub>)<sub>2</sub>Fe(SO<sub>4</sub>)<sub>2</sub> · 6 H<sub>2</sub>O, 0.328 g l<sup>-1</sup> of (NH<sub>4</sub>)<sub>6</sub>Mo<sub>7</sub>O<sub>24</sub> · 4 H<sub>2</sub>O, 0.314 g l<sup>-1</sup> of CuSO<sub>4</sub> · 5H<sub>2</sub>O and 0.322 g l<sup>-1</sup> of CoCl<sub>2</sub> · 6H<sub>2</sub>O. Municipal WW from the city of Dübendorf, Switzerland, was used. Effluents of grit and fat removal (raw WW) and additional primary clarification (primary effluent WW) from the pilot-scale WWTP at Eawag were used for this study.

#### III.3.6 Physical sludge parameters

TSS, VSS and SVI<sub>5</sub>, SVI<sub>10</sub>, SVI<sub>30</sub> were quantified using standard methods [APHA, 2012]. Additionally, the SVI<sub>30/10</sub> and SVI<sub>30/5</sub> ratios were calculated. The sludge size fractions were separated by sieving the sludge at 1, 0.63 and 0.25 mm, respectively. Granules were associated with fractions d > 0.25 mm, flocs d < 0.25 mm. Size fractions were then quantified based on TSS measurements. Sludge morphology was observed by stereomicroscopy (Olympus, SZX10, Japan) on weekly-/bi-weekly basis.

## Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater

---

### III.3.7 Analytical methods

Samples of influent and effluent were analysed for COD, total nitrogen (TN) and total phosphorus (TP) using photochemical tests (Hach Lange, Germany, LCK 114, 314, 338, 238, 348, 349). Soluble COD (sCOD) was measured after filtration at 0.45  $\mu\text{m}$  (Macherey Nagel, Nanocolor Chromafil membranefilter GF/PET 0.45  $\mu\text{m}$ , Germany). Cations ( $\text{NH}_4^+$ -N) and anions ( $\text{NO}_3^-$ -N,  $\text{NO}_2^-$ -N,  $\text{PO}_4^{3-}$ -P) were measured using flow injection analysis (Foss, FIAstar flow injection 5000 analyzer, Denmark) and anion chromatography (Methrom, 881 compact IC, Switzerland), respectively. VFAs were measured using headspace solid-phase microextraction (HS-SPME) followed by gas chromatography coupled to flame ionization detection (GC-FID) (Trace 1300 GC, Thermo Scientific, USA) [Feng et al., 2008].

### III.3.8 Microbial community analysis

#### Biomass sampling, DNA extraction and sequencing

Both granules and flocs were collected for analysis of the microbial community composition after sieving at 250  $\mu\text{m}$ . Biomass samples of around 1 ml were centrifuged 5 min at 8392 x g (Nuair Awel CF-48R centrifuge, U.S.A) then washed twice by addition of 5 ml of ice-cold phosphate buffer saline (PBS) solution and then centrifuged again min at 8392 x g. Pellets were then re-suspended in 3 ml of PBS solution, homogenized with a glass homogenizer, distributed in cryotubes and stored at -80 °C until DNA extraction. 200  $\mu\text{l}$  of homogenized biomass were mixed with 400  $\mu\text{l}$  of elution buffer (T<sub>10</sub> E<sub>0.1</sub>) and 100  $\mu\text{l}$  of lysozyme solution (25 mg ml<sup>-1</sup>). After one hour at 37 °C, DNA was extracted using an automatic robot 16 DNA Purification System (Maxwell, Promega Corporation, Switzerland). The DNA concentration of each DNA extraction was measured with a spectrophotometer NanoDrop ND1000 (Witec AG, Switzerland). The bacterial 16S rRNA gene hypervariable regions V1-V2 were amplified by polymerase chain reaction (PCR) in a T3000 Thermocycler (Biometra GmbH, Germany), using the universal primers 27F and 338R (bold), respectively, (bold) with overhang adapters attached: forward (5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-**AGMGTTYGATYMTGGCTCAG**3') and reverse (5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACA-**GGCTGCCTCCCGTAGGAGT**3'), and the High-Fidelity Q5 polymerase (High-fidelity 2x Master Mix, Biolabs Inc., USA). For each sample, 5  $\mu\text{l}$  of DNA were mixed with 2.5  $\mu\text{l}$  of forward primer 27F (1  $\mu\text{M}$ ), 2.5  $\mu\text{l}$  of reverse primer 338R (1  $\mu\text{M}$ ), 25  $\mu\text{l}$  of QS High-Fidelity 2x master mix and 15  $\mu\text{l}$  of water, for a final volume of 50  $\mu\text{l}$ . The PCR runs were performed in a T3000 Thermocycler (Biometra GmbH, Germany) with the following steps: initiation (2 minutes, 95 °C), followed by 30 cycles of denaturation (45 seconds, 95 °C), annealing (45 seconds, 50 °C) and elongation (60 seconds, 72 °C) and a final extension step (5 minutes, 72 °C). The amplified DNA was quantified using the DNF-473 standard sensitivity NGS fragment analysis kit (Advanced Analytical Technologies Inc., U.S.A). The Lausanne Genomic Technologies Facility (University of Lausanne, Switzerland) performed secondary indexing PCR and multiplex sequencing by groups of 96 samples per run on an Illumina MiSeq platform in paired-end mode (2x250). The sequences were deposited at the

European nucleotide archive (ENA) under the study accession number ERP111727.

#### **Taxonomic affiliation of 16S rRNA gene sequences**

The amplicon sequences were demultiplexed and primers removed. The trimming and quality filtering of the sequences was performed using trimmomatic version (v.) 0.36 [Bolger et al., 2014] with a sliding window of 4 base pairs (bp), a quality score threshold of 15 and a minimal length of 100 bp. The paired-end reads were merged with Pear v. 0.9.11 [Zhang et al., 2014]. The sequences were then grouped with a minimum similarity threshold of 97 % using the clustering software cd-hit v. 4.6.1 [Fu et al., 2012]. Clusters with less than 5 sequences per sample on average were discarded. The cluster heads of the remaining clusters were compared with the 16S rRNA gene database MiDAS v. S123\_2.1.3 [McIlroy et al., 2017] using the blast software [Altschul et al., 1990]. For each cluster, the taxonomy of the best match with the cluster head was attributed to all the sequences of the cluster. The level of precision of the taxonomy was adjusted according to the percentage of similarity with the threshold sequence identity values given by [Yarza et al., 2014], 94.5 % for genus, 86.5 % for family, 82.0 % for order, 78.5 % for class and 75.0% for phylum. For example, if a sequence had 90 % of similarity with its best match, the taxonomy attributed to its cluster was precise only up to the family level.

#### **Statistical analysis**

All the statistical analyses and the related plots were performed with R program v. 3.5.0 [R Development Core Team, 2008] using the packages reshape2 and ggplot2 [Warnes et al., 2016, Wickham, 2007, 2016]. Bray-Curtis distance matrices, the associated principal coordinates analysis (PCoA) and Mantel-tests were done with the package vegan v. 2.5-2 [Oksanen et al., 2018]. A multifactorial analysis (MFA) was performed with the R-package FactoMineR [Le et al., 2008].

#### **Determination of the stable states and the discriminant taxa**

After visual inspection of the PCoA plots on the Bray-Curtis distance matrix of the bacterial operational taxonomic units (OTU, 97 %) relative abundance, the biomass communities of the different samples were separated in two states; a “transition state” and a “stable state”. For each reactor, a bacterial community was considered in the stable state if the maximum Bray-Curtis pairwise distance with the communities of all the following sample points was below 0.6 (the maximal distance between all the samples was 0.88). Five stable state data-sets corresponding to the stable states of the 4 reactors and of the inoculum were analyzed further. In order to extract the taxa that are discriminant between these five stable states, the mean relative abundances of the genera were compared two by two. After Hellinger transformations, the means were compared by using t-tests. The taxa are considered “divergent” if their mean in the stable data-set is significantly different (p-value 0.01 corrected for multiple testing using

### **Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater**

---

Bonferroni correction,  $p\text{-value} = 0.01/422 = 2.34 \times 10^{-5}$ ) between at least two stable states. Taxa are considered “abundant” if their average abundance during stable state is higher than 1 % in at least one stable data-set. The taxa being divergent and abundant are considered as “discriminant taxa” in the following analysis. There were 56 abundant, 273 divergent and 38 discriminant taxa on a total of 422 (at genus level).

#### **Comparison of the bacterial communities in flocs and granules**

The average proportion of the most abundant genera were compared in each reactor with t-tests, in order to evaluate potential differences in the microbial communities in flocs and granules. The results with a p-value lower than the Bonferroni corrected p-value of 0.01 ( $p\text{-value} = 0.01/20 = 5 \times 10^{-4}$ ) were considered as significant.

## **III.4 Results**

### **III.4.1 Settling properties**

A higher amount of diffusible organic substrate (VFA or fermentable) in the WW resulted in faster granulation and better settling properties (Figure III.2). Lower  $SVI_{30}$  and SVI-ratios closer to 1 were thus measured for the AGS of R1 and R2 (high content in diffusible organic substrate). Larger  $SVI_{30}$ , and SVI-ratios close to 0.8 were on the other hand measured for the AGS of R3 and R4 (high content in non-diffusible organic substrate).

Achieving good settling properties for the systems fed by municipal WW required a much longer period (several months to over 1 year). The  $SVI_{30}$  values of the AGS of R4 steadily decreased within the first 100 days of operation to  $50 \text{ ml g}^{-1}$  for both runs and stabilised at  $SVI_{30} < 70 \text{ ml g}^{-1}$  after 170 d for R4 run #2. AGS of R3 responded sensitively to changing operating conditions, which resulted in variable and high  $SVI_{30}$  values from day 0 to 200.  $SVI_{30} < 80 \text{ ml g}^{-1}$  and  $SVI_{30/10}$  ratio  $> 0.8$  were finally achieved within a few weeks quickly after introduction of anaerobic PF+mixing (from day 357) for R3. Successful start-up based on settling parameters ( $SVI_{30} < 90 \text{ ml g}^{-1}$  and  $SVI_{30/10}$  ratio  $> 0.8$ ) was achieved within the first two weeks for R1 and R2. A much longer start-up time was required for AGS of R4 and R3. Around 34 and 163 days were required to achieve successful start-up based on settling parameters for AGS of R4 run #2 and R4 run #1, respectively. R3 was successfully started-up only after 400 days.

### **III.4.2 Sludge size fractions**

The effect of the influent composition on the granulation process was confirmed by monitoring the different biomass size fractions (Figure III.3). The size of the granules greatly varied as a result of the presence of non-diffusible  $X_B$  in the influent. High fractions of large granules ( $d > 1 \text{ mm}$ ) were observed in R1 only, while smaller granules together with flocs were observed in

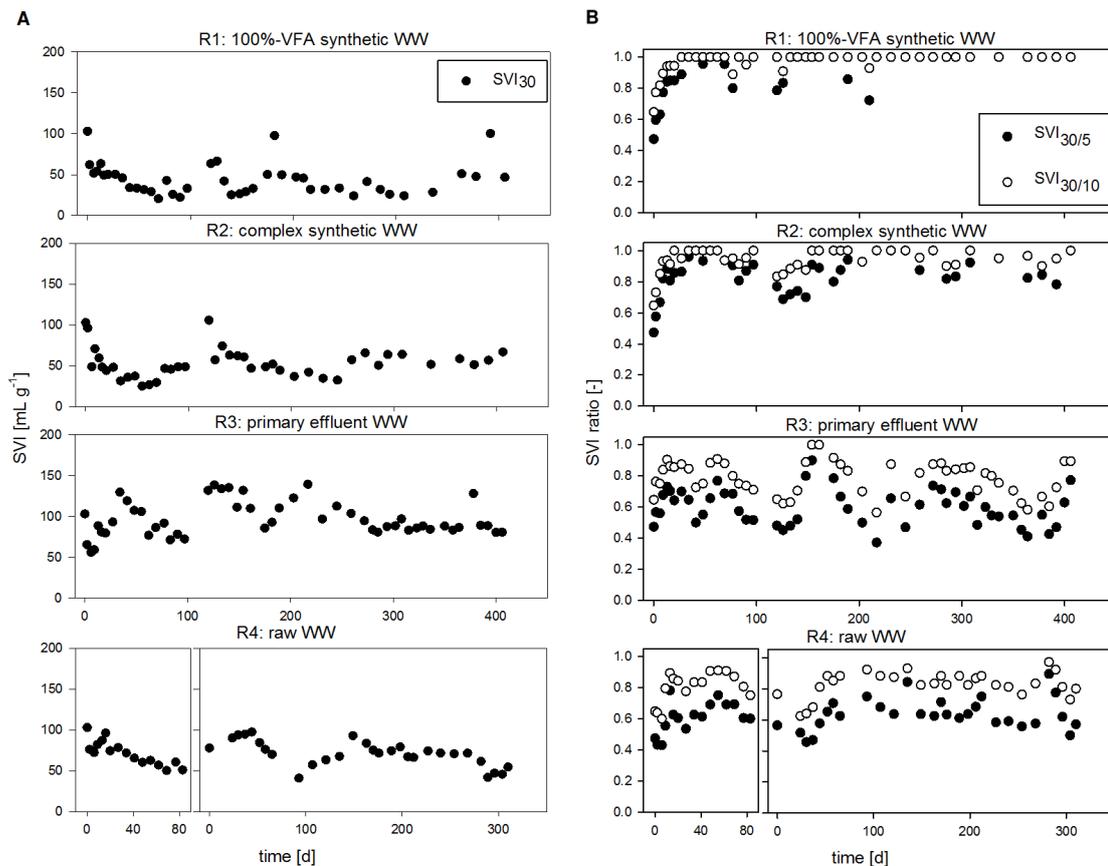


Figure III.2 – Evolution of (A) the sludge volume index  $SVI_{30}$  (measured after 30 minutes) and (B) the SVI ratios (30/5 and 30/10) of the aerobic granular sludge of the four SBR. All SVI-values are provided in supplementary information (Figure A.2).

all systems.

Medium and large diameter granules ( $d > 0.63$  mm) dominated the AGS in R1, while flocs ( $d < 0.25$  mm) represented only a minor fraction. The fraction of granules steadily increased in R1, while flocs simultaneously decreased to about 5 % after 50 days of operation. Larger granules gradually replaced small granules and dominated the sludge composition after 100 days of operation. The total fraction of granules ( $d > 0.25$  mm) in R1 remained steady throughout the entire reactor operation, despite some fluctuations in the individual fractions.

A major loss of large granules was observed after 300 days of operation in R1. Incomplete up-take of carbon during anaerobic conditions, possibly caused by pass of the settled sludge bed during plug-flow feeding, resulted in filamentous outgrowth of the granules (Supplementary information Figure A.1d). Filamentous outgrowth first resulted in an increase of sheared-off debris (indicated by an increase in  $d < 0.63$  mm size fractions), followed by breakage of large granules. After 390 days of operation though large granules started to develop again.

### **Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater**

---

AGS fed with complex influent WW were mainly composed of small granules (50-70%) and flocs (20-40 %). The AGS size fractions measured in the reactors fed with complex synthetic and municipal WW were very similar. Small granules represented the predominant size fraction, with 60-80 %. Almost no large granules developed in these systems (rarely above 10 % of total biomass). Also, the fractions of flocs decreased from 80 % to less than 40 % within the first 40 days. After day 40, 20-40 % of flocs remained in the systems until the end of the experiment. Large fluctuations of size fractions - mainly of flocs - were only observed in R2 and R3.

#### **III.4.3 Evolution of the bacterial community composition from inoculation to stable state**

The microbial communities of the different reactors were monitored over 426 days for the four reactors (267 samples, 12 millions of reads) (Figure III.4). The microbial communities developed differently in R1 (100%-VFA) in comparison to R2, R3 and R4, with the latter two being very similar to each other. The microbial community in R2 (synthetic complex WW) was rather similar to the ones of R3 and R4, although some differences could be observed. In R1, an initial increase of the Gammaproteobacteria with successive changes within this class was observed. *Dechloromonas* was progressively replaced by other Betaproteobacteriales such as *Azoarcus* or *Zoogloea* whose relative abundance fluctuated greatly during the experiment. The proportion of Actinobacteria, here comprising mainly putative fermenting bacteria, gradually decreased to below 2 % after day 119.

*Tetrasphaera* and *Ca. Accumulibacter* represented 8 % and 3 % of the inocula communities, respectively. In R1, the abundance of *Tetrasphaera* decreased progressively and was <0.5 % after 100 days whereas the abundance of *Ca. Accumulibacter* fluctuated between 0.1 % and 8 %. In R2, R3, and R4, Actinobacteria became abundant (30-50 %) during the two first weeks of operation and stabilized at around 10 % after 200 days of operation. The abundance of *Tetrasphaera* remained quite stable during the first 130 days in these reactors and then decreased to 1-3 % whereas abundance of *Ca. Accumulibacter* was always < 3 %.

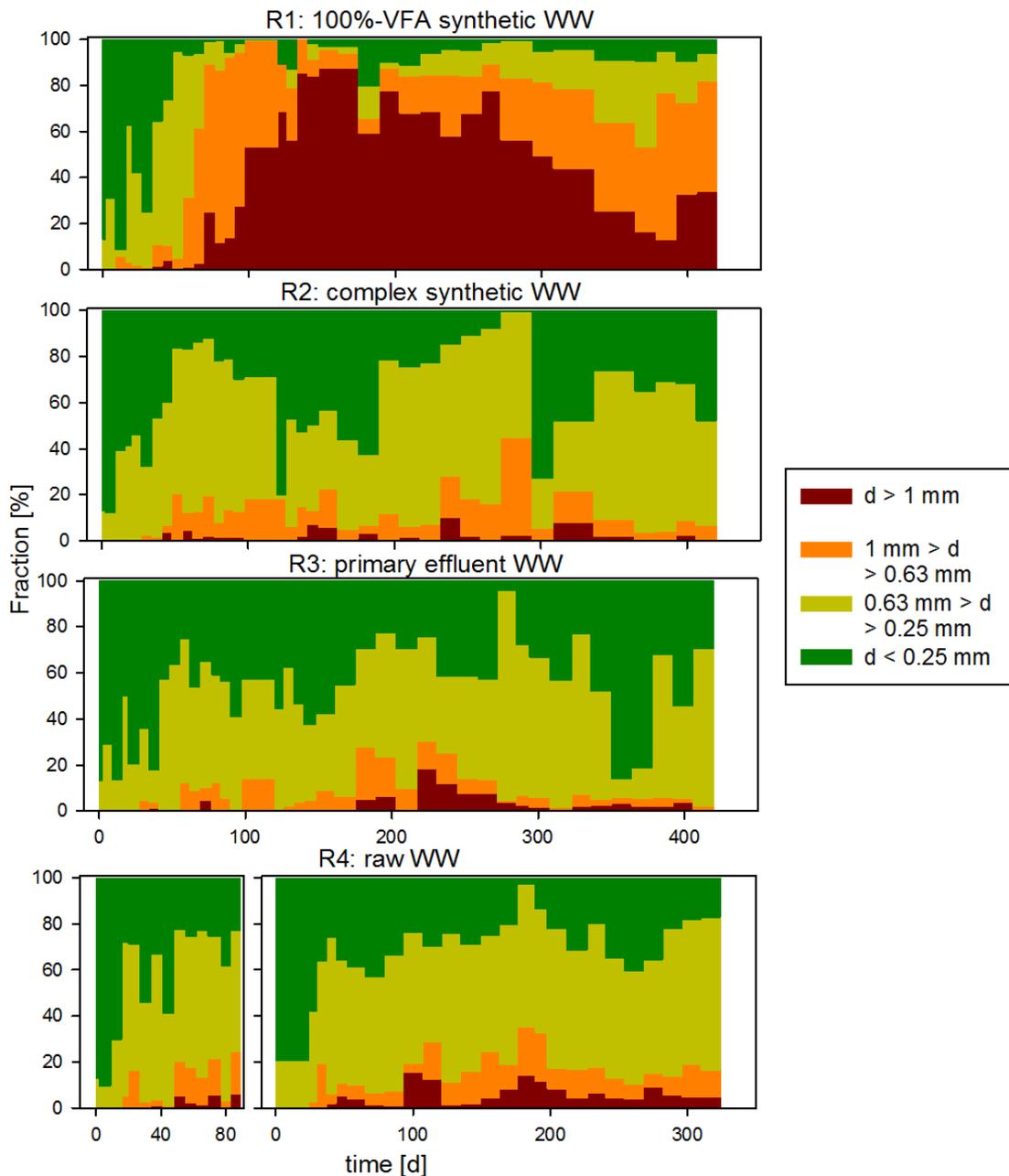


Figure III.3 – Evolution of the sludge size fractions (TSS based) for the aerobic granular sludge fed with different influent composition: 100%-VFA synthetic (R1), complex synthetic (R2), primary effluent (R3) and raw influent WW (R4) run #1 and run #2. Aggregates with  $d < 0.25 \text{ mm}$  are considered as flocs, aggregates with  $d > 0.25 \text{ mm}$  and  $d < 0.63 \text{ mm}$  as small granules, aggregates with  $d > 0.63 \text{ mm}$  and  $d < 1 \text{ mm}$  as medium granules and aggregates with  $d > 0.63 \text{ mm}$  as large granules.

### Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater

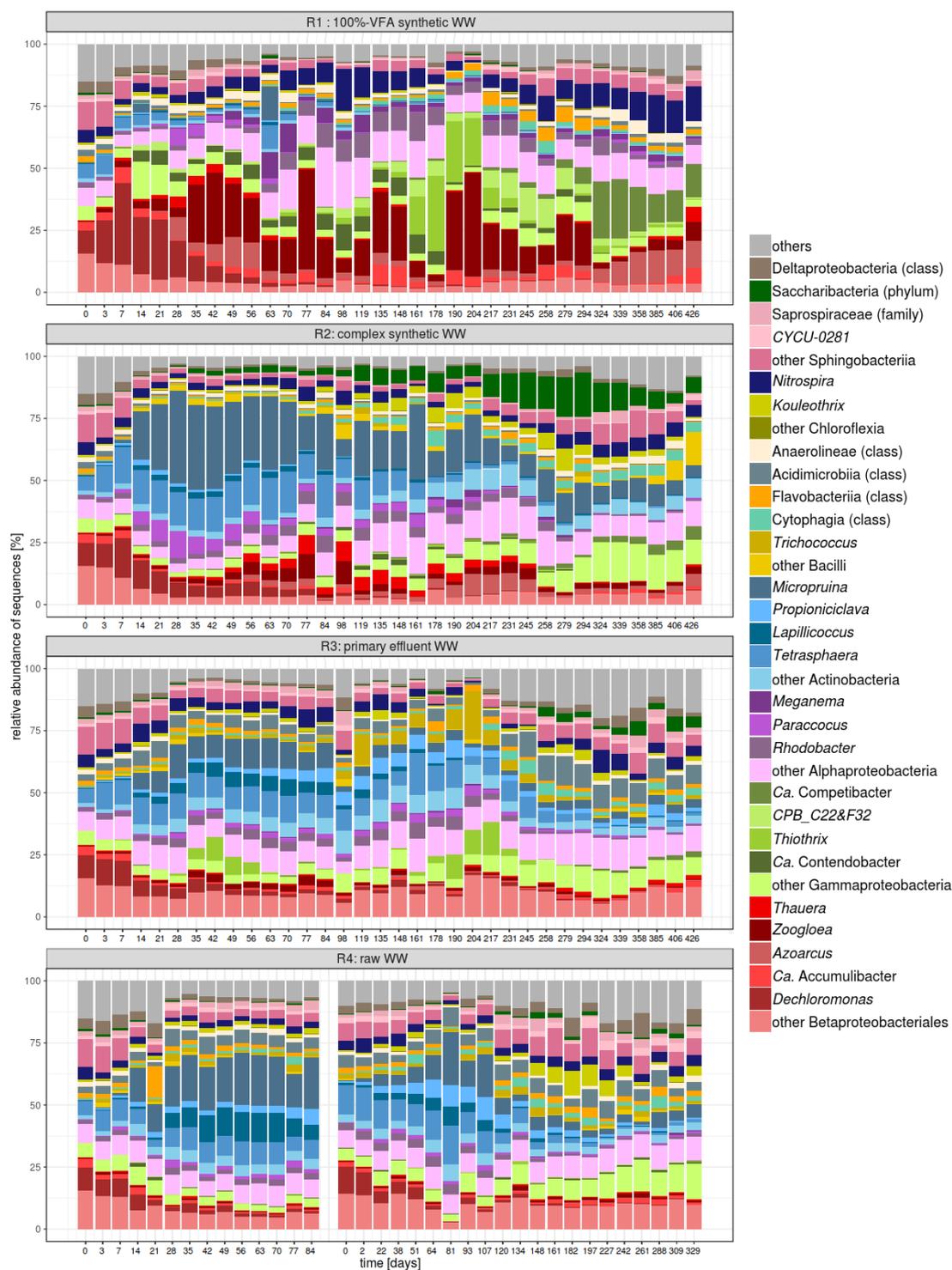
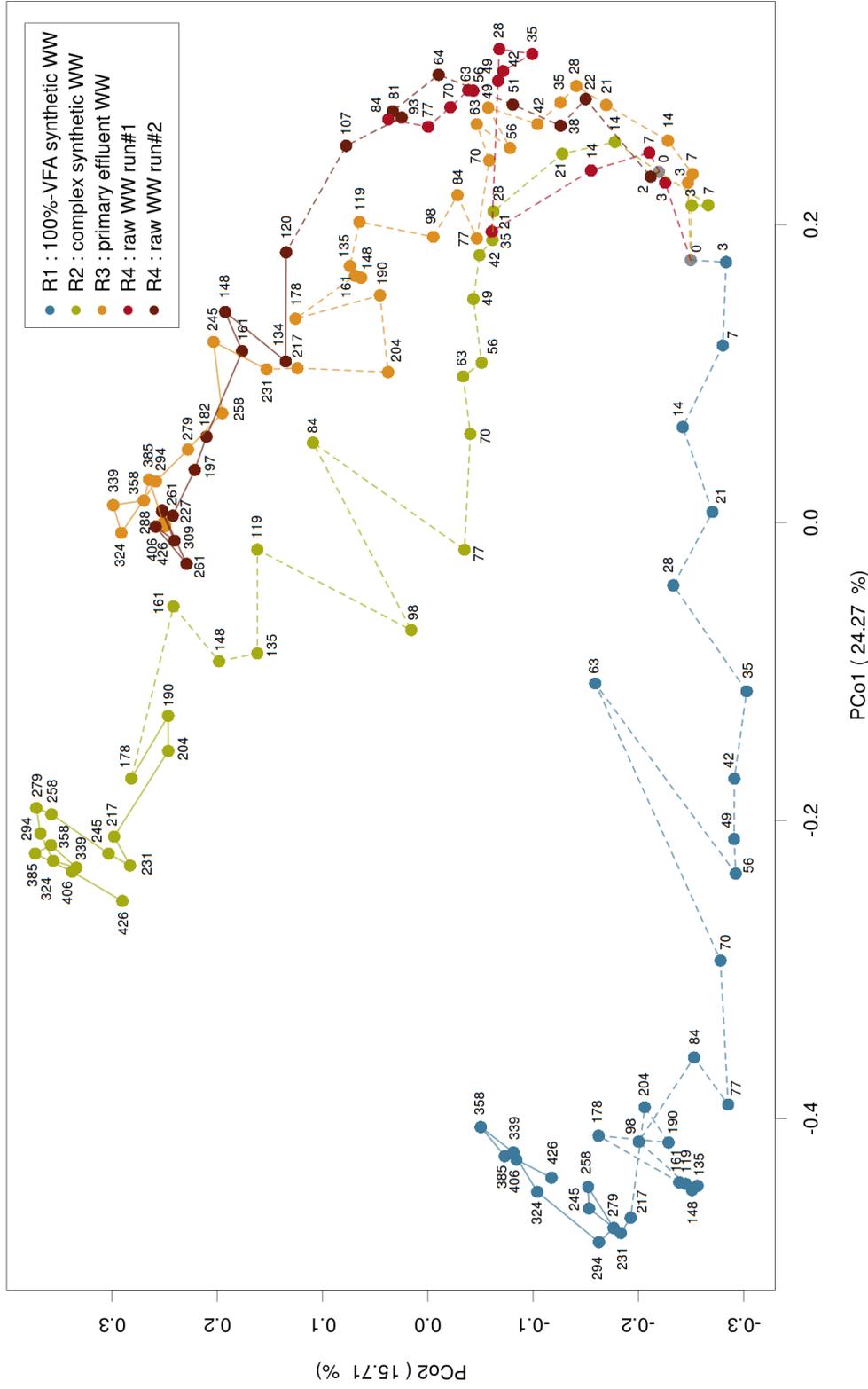


Figure III.4 – Composition of bacterial communities from inoculation to stable state in the four AGS reactors treating different types of WW. The most abundant taxa are shown in colors depending on the class they belong to. One exception is the order Betaproteobacteriales (colored in red) that has recently been included in the class Gammaproteobacteria [Parks et al., 2018]. The other taxa of the latter class are colored in green. The evolution of the bacterial community of the reactor treating raw WW (R4) is shown for both run #1 and run #2.



### Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater

---

The evolution of the bacterial communities according to the different WW are represented in the PCoA of the Bray-Curtis distance matrix of the bacterial OTU (97 %) relative abundance (Figure III.5). In all reactors, the bacterial community quickly changed first after inoculation (transition phase) and then stabilized 'stable state'. The bacterial communities of R3 and R4 (municipal WW), evolved very similarly towards a similar stable state. The bacterial community of R1 evolved very differently from R3 and R4. The evolution of the bacterial community of R2 was different from the one of R1 and quite close to the ones in reactors R3 and R4. The time to reach stable state significantly varied from one reactor to another. In R1 and R3 the bacterial community stabilized after 231 days, while 178 and 120 days were required for R2 and R4, respectively. The evolution of bacterial communities of R4 was similar during run #1 and run #2. The Shannon diversity index decreased during the transition phase, in particular for the reactors treating synthetic WW (Supplementary information Figure A.3). The index was higher in the samples of the stable states, in comparison to the samples of the transition phases, which supports the pertinence of the criterion applied to determine the stable state.

A comparison of the proportions of each genus was performed in order to identify the taxa responsible for the differences between the different microbial communities of the four AGS systems and the inocula (Supplementary information Figure III.11). Out of the total 422 taxa, 38 were identified as discriminant. Some of the discriminant taxa were characteristic for AGS fed with simple WW (R1) such as *Meganema* or members of the Rhodobacteraceae family (Supplementary information Figure III.6). Other taxa were mainly abundant in the AGS fed with real WW (inocula, R3 and R4) such as *Propioniciclava*, *Iamia*, *Acidovorax*, *Kouleothrix*, *Ca. Epiflobacter* and *Sulfuritalea*. Several taxa were more abundant in AGS treating complex WW, whether synthetic or municipal: *Microthrix*, the fermentative GAO *Micropruina* or the fermentative PAO *Tetrasphaera*. The following genera were abundant and present in the 'core community' of the systems but were not discriminant: the PAO *Ca. Accumulibacter*, the GAO *CPB\_C22&F32*, *Ca. Microthrix*, Saprospiraceae (f), *Rhodobacter*, *Thauera*, and *Thiotrix*. Finally, the proportions of the bacterial taxa of R1 and R2 were compared with those of R3 and R4 combined. This comparison was performed to assess the effect of synthetic vs. real WW on the bacterial communities of AGS. Among the 29 abundant taxa in R3 or R4, 23 were found in significantly lower proportions in R1. In R2, the number of underrepresented taxa dropped to only 15 (Supplementary information Table A.10).

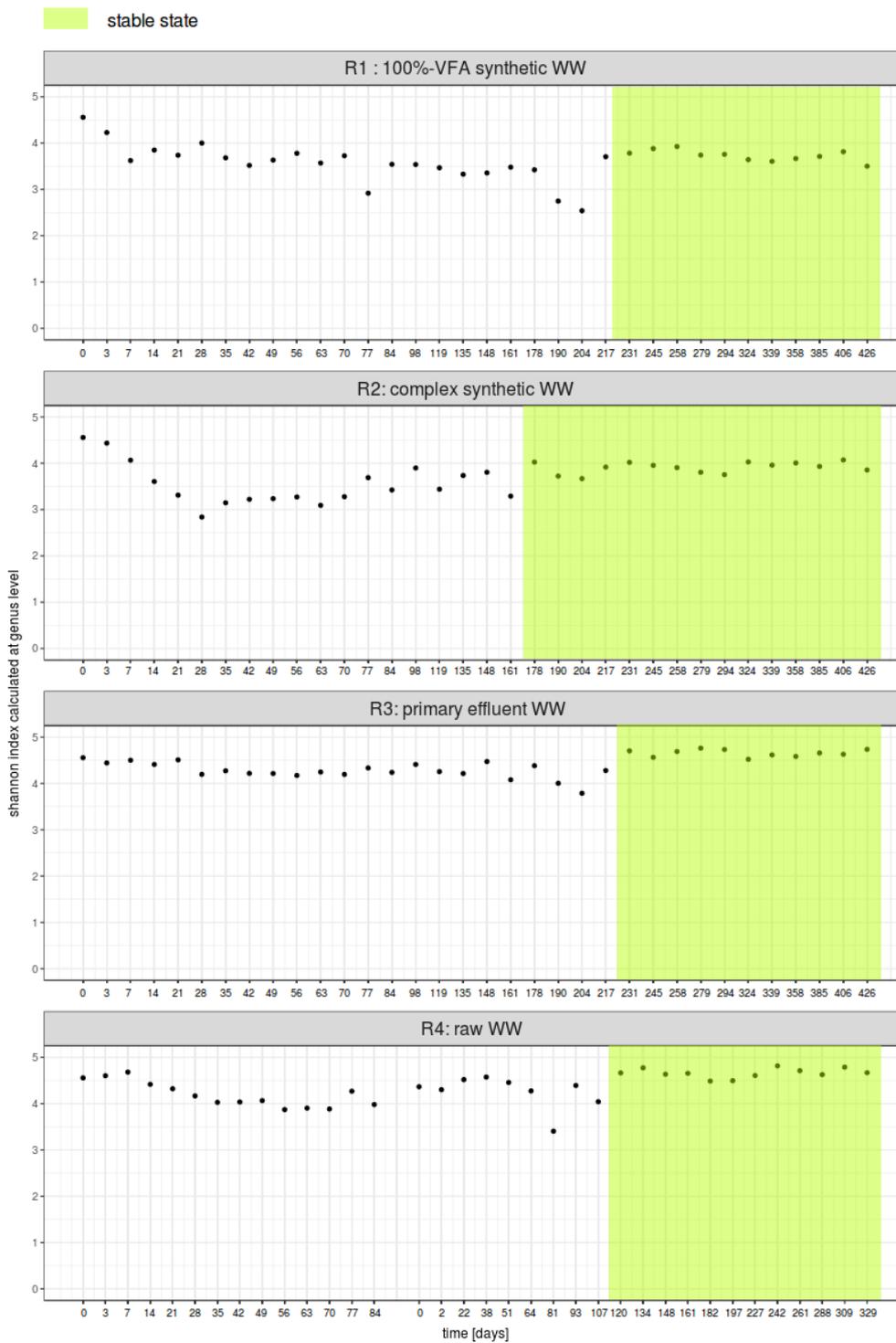


Figure III.6 – Evolution of the Shannon diversity index in the four reactors. The values corresponding to the reactor treating raw WW (R4) are shown for both run #1 and run #2. The periods corresponding to the stable state of the bacterial communities are indicated in green.

### Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater

#### III.4.4 Bacterial communities of granules and flocs

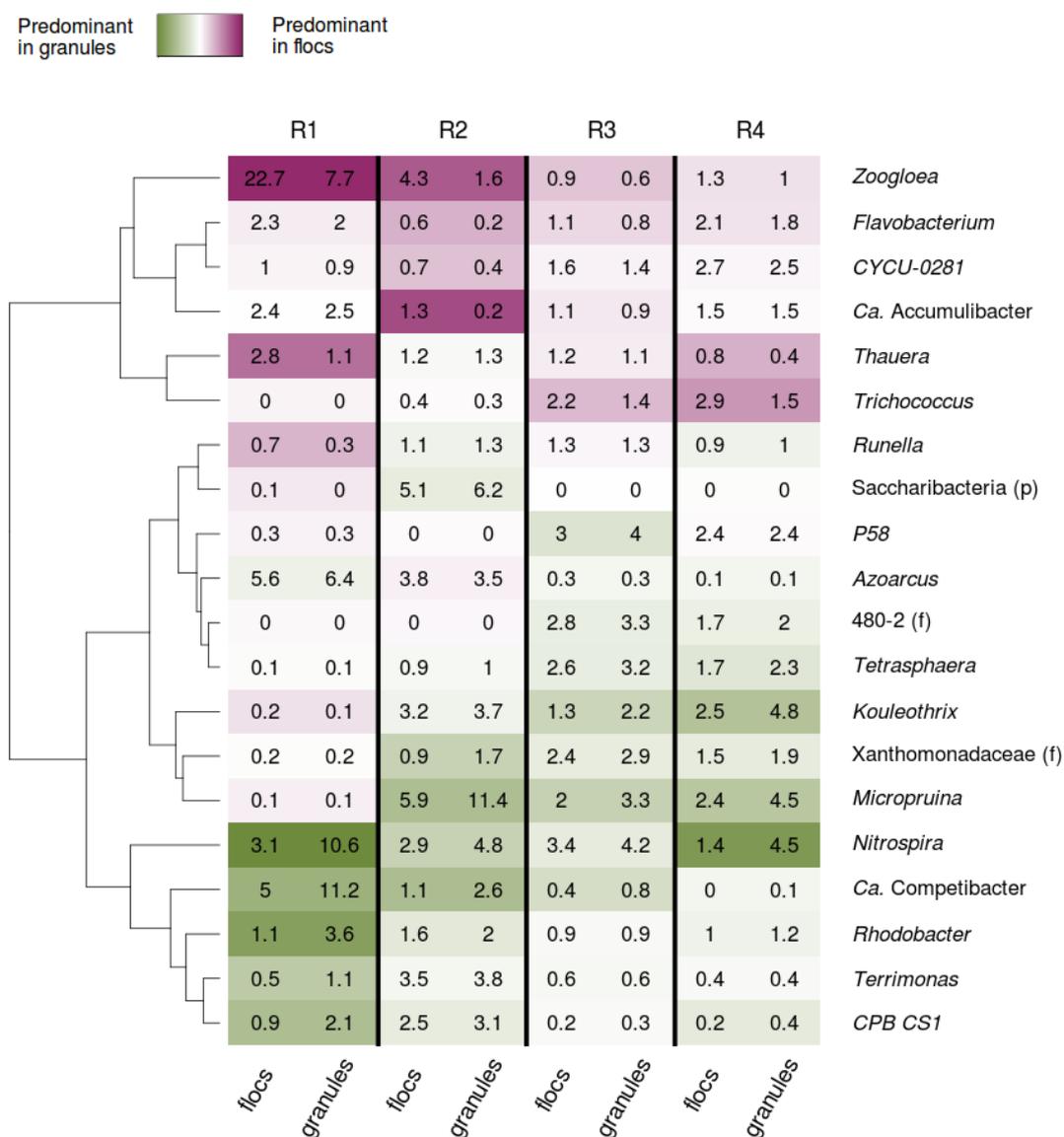


Figure III.7 – Average relative abundance of the main genera in the flocs and granules fractions collected in the four reactors during the stable state. Purple indicates a higher proportion in flocs while green indicates a higher proportion in granules. A pseudo-count of 0.5 % was added to each abundance to lower the possible effect of the noise in very low abundant genera.

The presence of non-diffusible organic substrate in the WW resulted in 20-40 % of flocs in the AGS. It is thus relevant to understand to what extent the microbial communities in flocs and granules are similar. The relative abundances of the main genera in flocs and granules were compared to detect potential enrichment of some genera between those two types of microbial aggregate (Figure III.7). At stable state, *Zoogloea*, *Flavobacterium*, *CYCU-0281*, *Thauera* and

*Trichococcus* were enriched in flocs, whereas *Nitrospira*, *Ca. Competibacter*, *Rhodobacter*, *Terrimonas* and *CPB\_CS1* were enriched in granules. However, only few of these differences were significant (Supplementary information Table A.8), when considered separately. This was the case for e.g. *Nitrospira* in R1 and R4 and *Thauera* in R4. Hence, in term of microbial community composition, flocs are quite similar to granules.

### III.4.5 Nutrient removal performance

Table III.2 – Effluent concentrations and nutrient removal performances of the four AGS reactors fed with 100 %-VFA synthetic (R1), complex synthetic (R2), primary effluent (R3) and raw influent (R4) WW run #1 and run #2. NO<sub>2</sub>-N in the effluent was in the range of 0.1-0.3 mg N l<sup>-1</sup> for all reactors.

Reactors	100%-VFA synthetic WW R1	complex synthetic WW R2	primary effluent WW R3	raw WW R4 run #1	raw WW R4 run #2
TSS effluent [mgTSSl <sup>-1</sup> ] <sup>a</sup>	13±15	17±22	43±43	15±15	34±34
COD removal[%]	91±7	93±5	83±11	92±3	88±8
TN removal [%]	77±14	60±16	45±20	47±12	63±16
NH <sub>4</sub> -N removal[%]	95±7	97±6	96±4	94±7	97±5
NH <sub>4</sub> -N effluent [mgNl <sup>-1</sup> ]	0.2±0.4	0.1±0.1	0.3±0.5	0.2±0.2	0.2±0.3
NO <sub>3</sub> -N effluent [mgNl <sup>-1</sup> ]	4±4	13±5	13±6	12±4	11±5
TP removal [%]	89±10	89±14	49±44	64±20	73±17
PO <sub>4</sub> -P removal [%]	92±11	96±7	64±28	63±24	79±18
PO <sub>4</sub> -P effluent [mgPl <sup>-1</sup> ]	0.4±0.6	0.2±0.3	1.2±1.1	1.0±0.7	0.7±0.9

<sup>a</sup> Calculated from measurements of samples taken during stable operation (no sludge washout events).

<sup>b</sup> The mean and standard deviation were calculated from 29-39 measurements for R1, R2, R3, 12-15 for R4 run #1, and 16-24 for R4 run #2

Overall, good substrate/nutrient removal and effluent quality was observed for all systems, despite significant differences in the sludge properties and granulation process (Table III.2). Very low TSS concentrations were measured in the effluent of R1 and R2 (< 20 mg TSS l<sup>-1</sup>), as well as in the effluent of R4 run #1 (15 mg TSS l<sup>-1</sup>). The highest TSS concentration was measured in the effluent of R3 (45 mg TSS l<sup>-1</sup>). During run #2 of R4, TSS effluent concentrations (34 mg TSS l<sup>-1</sup>) were lower than in R3, but still higher in comparison to R4 run #1 as well as R1 and R2.

Excellent COD- and NH<sub>4</sub>-N-removal efficiencies were observed in all reactors, except R3. For R1, R2, and R4, the COD- and NH<sub>4</sub>-N-removal efficiencies were consistently larger than 88 % and 95 %, respectively. The sludge loss in the effluent of R3 (43 ± 21 mgTSS l<sup>-1</sup>) increased the average COD concentration in its effluent. Effluent concentrations of NH<sub>4</sub>-N and NO<sub>2</sub>-N were consistently low in all reactors. High TN-removal (77 %) via simultaneous nitrification-denitrification (SND) was observed in R1 only, which was indicated by the lowest NO<sub>3</sub>-N effluent concentrations. Larger NO<sub>3</sub>-N effluent concentrations measured in the effluent of R2, R3 and R4 and resulted from a larger accumulation of NO<sub>3</sub>-N during the aerobic phase, compared to R1. Since full nitrification was observed in all systems, it can be concluded that

### Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater

the higher NO<sub>3</sub>-N effluent concentrations result from a lower (simultaneous) denitrification rate during aerobic bulk conditions.

PO<sub>4</sub>-P-removal was constant and high (> 90 %) for AGS of R1 and R2 only. AGS of R4 run #1 and run #2 showed lower PO<sub>4</sub>-P-removal of 61 and 78 % on average, respectively. The lowest PO<sub>4</sub>-P-removal was observed for AGS of R3. However, PO<sub>4</sub>-P-removal of R3 and R4 run #2 improved up to > 95 % after introducing anaerobic PF+mixing at day 357 and 289, respectively.

Overall, successful start-up of the reactors in terms of substrate/nutrient removal was achieved long before achieving good settling properties. Full COD and NH<sub>4</sub>-N removal were observed right after inoculation of the systems. High P-removal was observed without delay for R1 and R2 while stable and high biological phosphorus removal was reached after 77 and 93 days for R3 and R4 run #2, respectively. R4 run #1 was not able to recover high P-removal performance before the restart.

#### III.4.6 Correlations between settling properties, nutrient-removal performances and microbial community composition

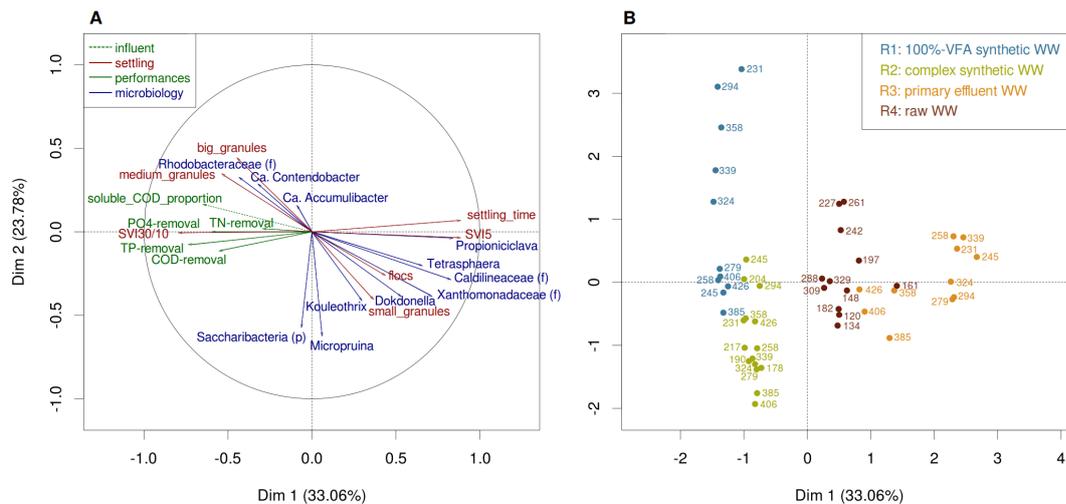


Figure III.8 – Multiple factor analysis (MFA) performed on the data at bacterial stable state in the four reactors with three groups of variables: settling characteristics, nutrient-removal and composition of the bacterial communities. These graphs show the contribution of the settling characteristics the nutrient-removal and the bacterial community compositions to the two first axis (A), and the projection of the corresponding sample points in this two-dimensional space (B).

Multiple factor analysis indicates a correlation between the proportion of soluble COD in the influent, a high proportion of medium to big granules and a SVI<sub>30/10</sub> ratio close to 1, suggesting good granulation and settleability (Figure III.8A). The proportion of soluble influent COD also correlated with good nutrient-removal performances, such as nitrogen, total phosphorus or

phosphate removal efficiencies (TN, TP and PO<sub>4</sub>-P-removal). The projection of the samples in the two-dimensional MFA space provides information about the global similarity between samples (Figure III.8B). Overall, samples of R2 are close to the ones of R1 and closer to samples of R4 than R3.

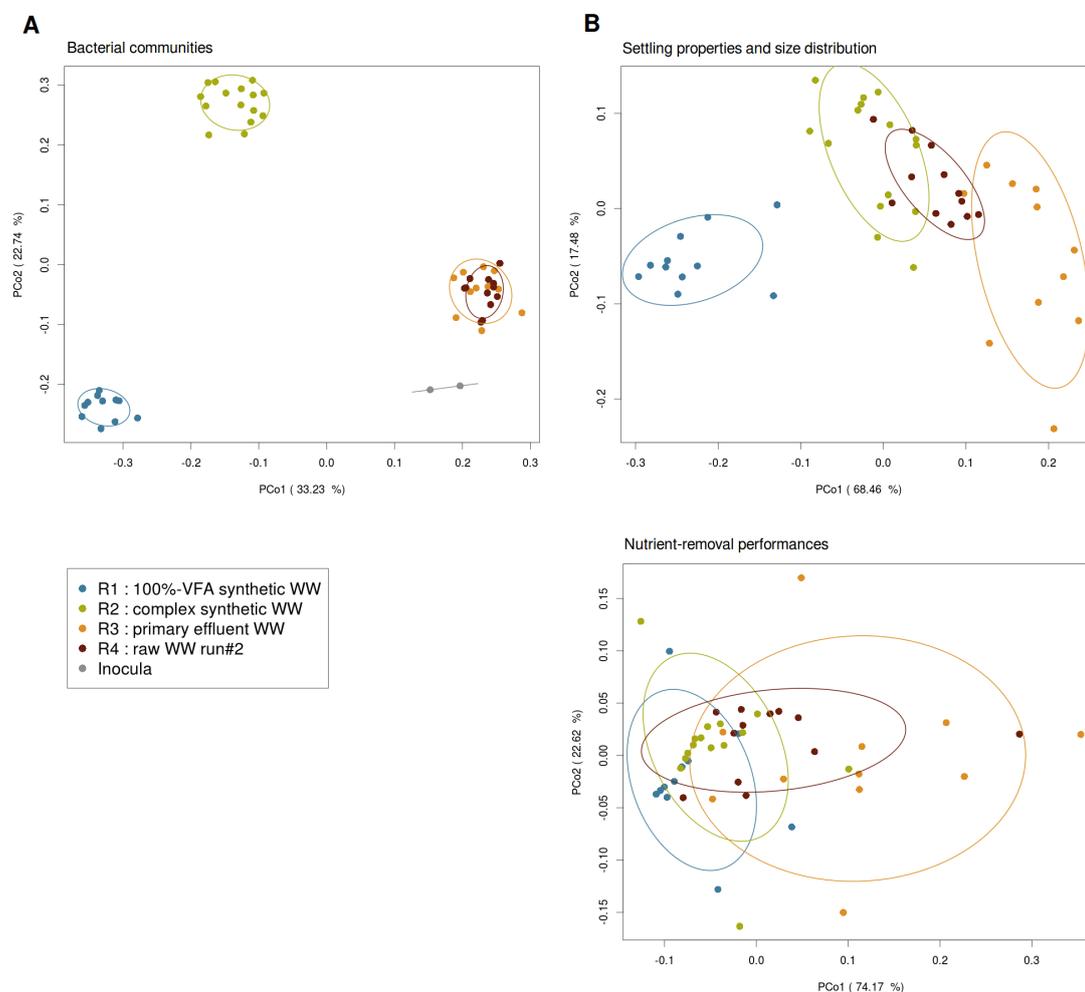


Figure III.9 – Principal component analysis based on the Bray-Curtis distance between the relative abundance of bacterial taxa (A), the granule size distributions and settling properties (B), and the nutrient-removal performances (C) in the four reactors during bacterial stable state. The ellipses indicate the 70 % confidence interval for the data related to each reactor.

At stable state, the bacterial communities, settling properties, size distribution of the sludge, and the nutrient removal performances show different individual distributions in the two-dimensional space (Figure III.9). Bacterial communities of AGS of R3 and R4 are very similar at stable state, and close to the communities of the inoculum (Figure III.9A). Bacterial communities of R3 and R4 are clearly distinct from the ones of the reactors treating synthetic WW, however closer to the ones of R2 than to those of R1. The projection of the samples forms a gradient from R1 to R3, with R2 and R4 in between based on settling properties and

### **Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater**

---

size distribution of the sludge (Figure III.9B). The links between the microbial communities and the size and density of the biomass is confirmed by the correlation of 0.65 between the two corresponding Bray-Curtis distance matrices (Supplementary information Table A.9 and Table A.11).

Based on the nutrient-removal performance data, a majority of samples mainly from R1, R2 and R4 grouped together. Outliers, mainly from R3, surrounded this group (Figure III.9C). This reflects the fact that, AGS of R3 had lower nutrient-removal performances than AGS of the other three reactors, even at stable state. Weak correlation (0.05) between the Bray-Curtis distance matrices of the bacterial communities and the nutrient-removal performances (Table A.9) is observed at stable state.

#### **III.4.7 Identification of correlations between the discriminant taxa and the sludge size distributions, the settling properties, and the nutrient-removal performances**

Three different clusters were identified by correlating the discriminant taxa with the different properties of the sludge (Figure III.10). Cluster I correlated with a high proportion of big and medium size granules, good settling properties and good nutrient removal. It includes the PAO *Ca. Accumulibacter* and its GAO competitors from the family Competibacteraceae (*Ca. Competibacter*, *Ca. Contendobacter*, *CPB\_CSI*). Aerobic filamentous (*Meganema*, *Zoogloea*), potentially filamentous (*Flavobacterium*) bacteria and the nitrifying *Nitrospira* also belong to this cluster. Numerous putative denitrifiers are also part of cluster I (*Zoogloea*, *Ca. Competibacter*, *Ca. Contendobacter*, *Ca. Accumulibacter*). Cluster II correlates with high proportions of small granules and flocs, relatively good settling properties, good P-removal and partial TN-removal. It is composed of the aerobic bacteria *Terrimonas*, which has the ability to hydrolyze various substrates, and *Amaricoccus*, which can store carbon in the form of PHA. It also comprises fermentative or putatively fermentative bacteria such as *Microthrix* and *Mesorhizobium* for which *in situ* physiology is not well described yet. The bacteria belonging to this cluster were relatively abundant in the reactor treating complex synthetic WW (R2). Cluster III contains taxa correlated with high proportions of small granules and flocs, poorer settling properties, and lower nutrient-removal. This cluster contains one third of novel or poorly characterized genera such as *P58* or *Dokdonella*. It also includes fermentative bacteria including *Tetrasphaera*, *Micropruina*, *Propioniciclava*, *Kouleotrix*, and *Trichococcus*. As cluster I, it contains various potential denitrifiers such as *Iamia*, *Sulfuritalea*, *Microthrix* or *Acidovorax*. Cluster III also comprises bacteria likely able to degrade macromolecules, e.g. *Ca. Epiflobacter*, *CYCU-0281* and members of the family Chitinophagaceae. These bacteria are more abundant in the systems treating municipal WW (Inocula, R3, R4). Members of the genera *Ca. Microthrix* and *Trichococcus*, and the family Caldilineaceae present in cluster III can be filamentous and thus impair the settleability of the sludge and be at least in part responsible for the poorer settleability associated with this cluster.

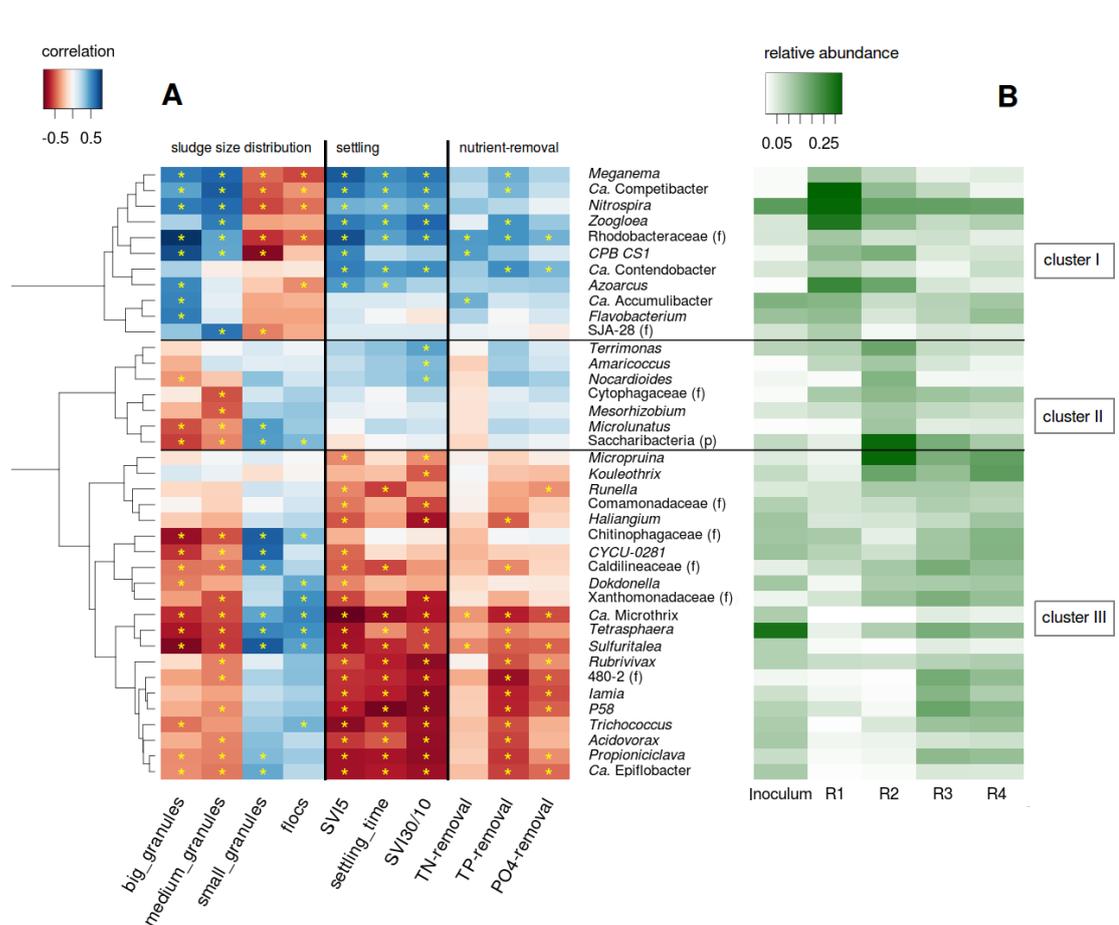


Figure III.10 – Correlation heatmap between the discriminant taxa and *Ca. Accumulibacter* and the sludge size distribution, the settling properties and the nutrient-removal efficiencies of the samples collected during the stable state, in the four reactors (A). The correlations having p-values lower than 0.01 are indicated with a yellow star. The different taxa were clustered together according to the similarity in terms of correlations with the different parameters of the sludge. The inverse values of SVI<sub>5</sub> and settling time were used for the construction of the correlation heatmap. The average relative abundance of the taxa (Supplementary information Table A.12) after Hellinger transformation, in the four reactors during the stable state is indicated in green (B).

**Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater**

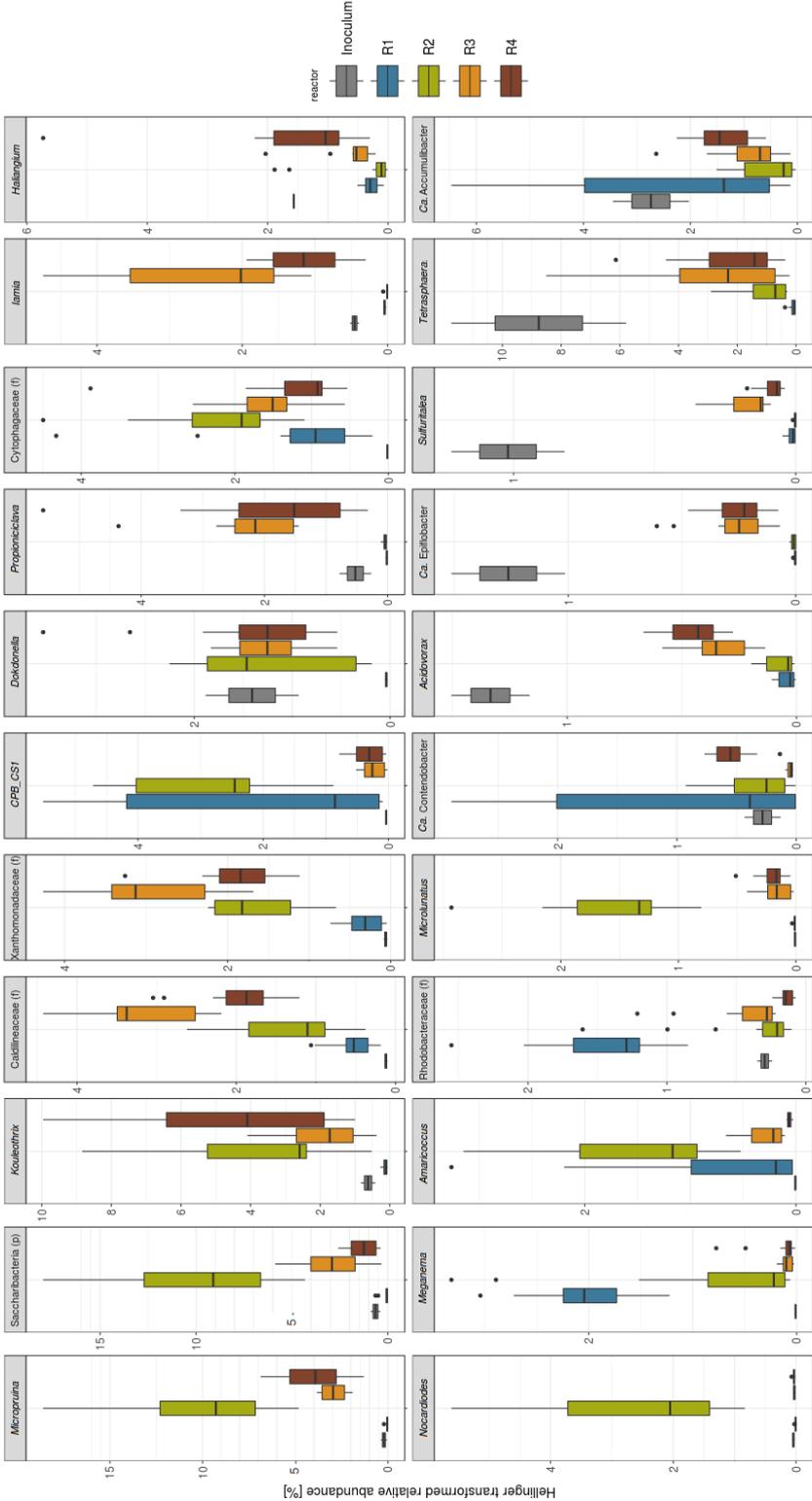


Figure III.11 – Boxplot of the Hellinger transformed abundance of the more discriminant genera (at least one p-value below  $1 \times 10^{-7}$  in the t-test), *Tetrasphaera* and *Ca. Accumulibacter*, in the four reactors and the inocula.

## III.5 Discussion

### III.5.1 Diffusibility of organic substrates has significant influence on formation of AGS

Wastewater composition, in terms of diffusible/non-diffusible organic substrate, significantly influenced the formation of AGS. A comparison between the four reactors fed with different amounts of diffusible/non-diffusible organic substrate help validating the conceptual model presented above (Figure III.1). Fast granulation was observed with WW containing high amounts of diffusible organic substrate (R1 and R2), thus resulting in excellent settling and stable nutrient-removal performances. On the other hand, a low amount of diffusible organic substrate resulted in a slow granulation, poorer settling properties, and often partial nutrient removal.

Granulation results from the selection of slower growing carbon-storing microorganisms over OHO [de Kreuk and van Loosdrecht, 2004, Vjayan and Vadivelu, 2017]. The high proportions of diffusible organic substrate in R1 and R2 promoted the growth of organisms able to store or use carbon under anaerobic conditions. Indeed, storing microorganisms such as *Ca. Accumulibacter* (classical PAO) and *Ca. Competibacter* or *CPB\_C22&F32* (classical GAO) represented an important part of the microbial community of R1 (Figure III.4). In R2, the fermentative PAO *Tetrasphaera* and the fermentative GAO *Micropruina* were abundant, in particular during granulation. The influent composition specifically favoured the growth of these organisms which coincided with rapid development of well-settling AGS. However, the size of the granules which developed in R1 and R2 was very different, with large granules dominating AGS of R1 while smaller granules were observed in R2. If diffusibility of substrate is very high, it promotes bacterial growth in the deep layers, ultimately leading to the formation of dense granules as in R1. Granule size is in theory associated with how deep substrate can penetrate into the granule, which is influenced by the substrate concentration gradient, uptake rate, and diffusibility of substrate [Eberhard Morgenroth, 2008, Rittmann and McCarty, 1981]. The nature and content of extrapolymeric substances (EPS) - associated with granule densification - is influenced by substrate type and loading, and has strong implications on granule size and settling [Rusanowska et al., 2019]. Deep substrate diffusion and conversion occurred in R1 as indicated by the high SND measured for this system. During the aeration phase, denitrification occurred in the anoxic zones of the granules due to previous anaerobic storage of diffusible organic substrate. Classical PAO and GAO enriched in the AGS of R1 were able to grow under both anoxic and aerobic conditions, thus favouring the densification of the granules. The lower fractions of diffusible organic substrate the R2 influent resulted in a limited growth in the core of the granules and ultimately in the formation of smaller granules. The different properties of the granules of R1 and R2 (settling, granule size and fractionation, etc.) can thus be explained by the amount and nature of the diffusible and non-diffusible substrates (VFA, fermentable soluble and particulate organic substrate). Granulation can therefore be linked to two aspects: (1) the diffusibility of the substrate governs its availability within the granules and (2) the nature of the substrates determines the microbial community

### **Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater**

---

composition and the aggregates densification.

On the contrary, the high proportions of non-diffusible  $X_B$  in the influent of R3 and R4 hampered the granulation process. Several phenomena lead to the conclusion that hydrolysis is only partial during anaerobic plug-flow feeding: (1) WW particles are often large, from several  $\mu\text{m}$  up to 1-2 mm [Dimock and Morgenroth, 2006, Levine et al., 1985], (2) hydrolysis is a very slow process [Benneouala et al., 2017, Jabari et al., 2016, Morgenroth et al., 2002], and (3) the anaerobic feeding duration of 1-2 h typically applied in AGS systems is insufficient to provide full hydrolysis of  $X_B$  [Jabari et al., 2016, Wagner et al., 2015b]. Therefore partial anaerobic conversion of  $X_B$  is unavoidable and results in a high availability of substrate during the subsequent aerobic phase, which in turn supports the growth of OHO [Wagner et al., 2015b]. OHO leads to poorer settling properties due to filamentous outgrowth, floc formation or a decrease of P-removal [de Kreuk et al., 2010, Novak et al., 1993, Pronk et al., 2015a, Suresh et al., 2018, Weissbrodt et al., 2014a]. Extensive OHO growth outcompeting slower growing storing organisms is the most likely reason for the slow development of granules, bad settling properties and overall lower nutrient removal performance observed for AGS of R3 and R4, in comparison to R1 and R2. Classical PAO and GAO were outcompeted by fermentative PAO and GAO in R3 and R4, and similarly in R2. The enrichment of fermentative PAO and GAO in those systems is the result of lower influent VFA concentrations combined with large amounts of fermentable substrates, stemming from both influent WWs and produced via hydrolysis of non-diffusible compounds.

#### **III.5.2 Start-up**

High concentrations of diffusible organic substrate, as found in the synthetic WW of R1 and R2, led to faster granulation (1 month) compared to granulation in R3 (> 1 year) and R4 (5 months). A main finding of our study is that different characteristic times were required to establish stable microbial communities, stable physical properties, or steady substrate/nutrient removal. Characteristic time to establish full substrate/nutrient removal was the quickest (around few days/weeks) while characteristic time to establish a stable microbial community was the slowest (several months).

Complete substrate and nutrient removal was indeed observed without any delay right after inoculation with activated sludge for all reactors. Maintaining high conversion rates after inoculation required an appropriate start-up strategy. Our selected start-up strategy relied on applying a low washout stress, to avoid a too harsh washout of slow growing organisms. Full nitrification without  $\text{NO}_2^-$  accumulation was observed straight after start-up, while no or very short loss of biological P-removal was noticed. A sustained nutrient removal after inoculation was also observed by Lochmatter and Holliger (2014), who applied a similar start-up strategy of low washout stress on slow-settling biomass. Low amounts of diffusible organic substrate in the WW must thus be balanced by a less harsh washout of flocs during start-up.

Aggressive washout of slow settling biomass as strategy to start-up AGS systems using either

100%-VFA or municipal WW as influent can result in very short start-up times down to 1 and 3 weeks, respectively, with settleability of  $SVI_{10} < 40 \text{ mlg}^{-1}$  [de Kreuk and van Loosdrecht, 2004, 2006]. The current study did not apply aggressive washout of slow settling biomass (flocs) during start-up resulting in significantly slower granule formation. Since granulation was achieved in all reactors, the start-up strategy in the current study is viable and relevant even in unfavourable influent WW conditions, such as low strength municipal WW. Maintaining high substrate/nutrient conversion rates over long-term ultimately resulted in the formation of granules. The different kinetics of AGS formation confirm the general trend observed in previous studies. Start-up is typically shorter with diffusible-only influent WW, consisting of 100%-VFA (1–4 weeks) [de Kreuk and van Loosdrecht, 2004, Lochmatter and Holliger, 2014, Weissbrodt et al., 2014a], than with WW containing non-diffusible polymeric compounds, such as municipal or industrial WW (3 weeks - more than 1 year) [de Kreuk and van Loosdrecht, 2006, Giesen et al., 2013, Liu et al., 2010]. The faster granulation of AGS in R4 compared to AGS of R3 confirms that higher organic loading rate tends to facilitate the granulation process [Li et al., 2008, Nancharaiyah and Reddy, 2018, Rusanowska et al., 2019].

Establishing a stable microbial community required between 4 and 8 months for all systems. During this period, different transient bacterial communities were observed. The transient bacterial community was dominated by *Zoogloea* in R1, fed with 100%-VFA WW. We propose that the presence of *Zoogloea* in R1 is an indirect consequence of the excellent settling properties of the AGS, which led to preferential flow and bypass of soluble COD into the aerobic phase (Supplementary information Figure A.4). The growth of *Zoogloea* during the early stage of granulation in R1 likely resulted from the presence of VFA in the aerobic phase [Weissbrodt et al., 2013]. Similar to the study of Weissbrodt et al. [2013], the high abundance of *Zoogloea* observed in R1 was associated with a high proportion of granules and a thin settled bed during feeding. Several studies suggested that *Zoogloea* plays a positive role in the formation of granules in VFA-rich influent WW by producing specific EPS [Kang et al., 2018, Larsen et al., 2008, Li et al., 2008]. Therefore the diffusible substrates that are residual at the beginning of the aerobic phase, such as VFA, must not in all cases be detrimental to granulation. In the reactors fed with complex WW (R2, R3 and R4), Actinobacteria dominated the transitional bacterial community. In particular, *Micropruina* was very abundant in these three reactors and was concomitant to the formation of granules. Its role in granulation in complex WW fed AGS systems is yet to be determined.

#### III.5.3 Stable state

The list of potential organisms that play a functional role in AGS systems can be extended based on this study. Especially considering their effect on nutrient removal, and taking into account the role of different influent compositions. However, a core microbial community of AGS systems cannot be established yet. The stable state operation primarily differed from the start-up phase mostly in terms of microbial community composition, which can help explain dynamics of nutrient removal performances. During stable operation, dominant species in R1

### Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater

---

were reported as abundant in other 100% VFA-fed AGS systems, such as *Zoogloea*, *Thauera*, *Rhodobacter*, *Meganema* and *Nitrospira*. In systems fed with VFA only, the PAO guild mainly consisted of *Ca. Accumulibacter* while the GAO guild was mostly composed of members of the Competibacteraceae family (e.g., *Ca. Competibacter*, *CPB\_C22&F32*) [He et al., 2016a, Henriet et al., 2016, Weissbrodt et al., 2013]. Such “simple” microbial community greatly differs from the ones of R2, R3 and R4. Abundant taxa detected in R2, R3 and R4 were previously detected in AGS fed with complex WW, like, e.g., *CYCU-0281*, *Dokdonella*, *Flavobacterium*, *Haliangium*, *Nitrospira*, *Rhodobacter*, *Thauera*, *Trichococcus*, unclassified genera related to Xanthomonadaceae, and *Zoogloea* [Kang et al., 2018, Swiatczak and Cydzik-Kwiatkowska, 2018, Szabo et al., 2017b]. Fermentative bacteria were present in both R2 and R3/R4, in particular the PAO *Tetrasphaera* and the GAO *Micropruina*. The sole presence of both diffusible and non-diffusible organic substrate - independent of their nature - resulted in relatively similar microbial communities. Yet, the remaining differences between those systems could stem from continuous inoculation of the sludge by bacteria present in the influent WW. Indeed, the similarity of the bacterial communities of R3 and R4 is stronger within AGS sampled during the same date than after the same number of days of reactor operation. Part of the microbial communities stem from immigration of bacteria *via* the influent WW [Saunders et al., 2016].

Filamentous OHO detected in R3 and R4 (e.g. *Trichococcus* and the family Caldilineaceae) can have a negative impact on the settling properties of the sludge. Those filamentous OHO are characterised by a high affinity for aerobic carbon degradation, which is consistent with the presence of slowly biodegradable substrates in the influent of those reactors. Yet, many genera known to have hydrolysing capabilities, such as *Ca. Epiflobacter*, *CYCU-0281* and *Kouleothrix*, correlated with low settling properties because they are linked to the presence of  $X_B$ , but not necessarily because they are filamentous.

During stable state, the bacterial communities selected in the different AGS systems had multiple taxa in common with the core communities of EBPR activated sludge systems (e.g., *Ca. Accumulibacter*, *Micropruina*, *Tetrasphaera*, *Zoogloea*) [Saunders et al., 2016, Stokholm-Bjerregaard et al., 2017]. But many taxa that are not yet characterized at the genus level were identified. They belong to the family of Xanthomonadaceae, Caldilineaceae, Cytophagaceae or to the phylum of Saccharibacteria. However, their function in AGS systems is yet to be determined.

The grouping of R1 vs. R2/R3/R4 based on microbial community also reflected differences in TN-removal, which was lower in the reactors fed with complex WW. TN-removal in the systems can occur either via (1) pre-denitrification of remaining  $\text{NO}_3^-$  from the previous cycle during the feeding and subsequent anoxic/anaerobic mixing phase, or (2) SND during aerobic bulk conditions. The latter process requires denitrifying bacteria, available COD and the presence of substantial anoxic zones within the granules. Numerous putative denitrifying bacteria were detected in AGS of all reactors, but only R1 gathered the two conditions to perform SND resulting in low effluent  $\text{NO}_3^-$ ; large granule size, and a high proportion of diffusible organic substrate in the influent. Further research is required to identify the influence of each of these

factors on SND, and to improve SND in AGS systems fed by more complex influent WW.

Decreased P-removal was observed in AGS systems treating municipal WW (R3, R4), possibly due to low influent diffusible organic substrate (especially VFA) in combination with carbon leakage, and low influent  $\text{PO}_4\text{-P}$  concentrations [Guimaraes et al., 2017]. Short-term loss of P-removal was also observed in R1 concomitantly to bypassing of substrate during PF feeding. In addition, low (ortho-) P concentrations of the municipal WW received by R3/R4 did not promote the growth of PAOs [de Kreuk et al., 2010]. Indeed, the proportion of PAO was above 5 % in the four reactors at stable state. These proportions are much lower than the average proportion reported in Danish EBPR WWTP (13 %) and may have made the four systems less robust in terms of P-removal [Nielsen et al., 2010]. Yet, the anaerobic carbon uptake, and therewith P-removal, were enhanced by the introduction of an anaerobic-mixed phase after PF feeding.

#### III.5.4 The role of flocs in AGS systems

Our results indicate that the presence of flocs (20-40 %) is representative of AGS systems fed with WW that contain non-diffusible  $X_B$  (Figure III.3). Flocs were observed for several months after the establishment of good/stable settling properties and substrate/nutrient removal. Flocs fractions ranging from 16 to 40 % of TSS were also reported in literature for pilot- and full-scale AGS plants [Derlon et al., 2016, Pronk et al., 2015b, van Dijk et al., 2018a]. Our results thus indicates that AGS systems are hybrid systems, composed of both flocs and granules, rather than biofilm (only) systems.

The presence of floc in AGS systems results from both short-term and long-term mechanisms. Short-term exposure of the granules to non-diffusible  $X_B$  triggers filamentous outgrowth de Kreuk et al. [2010].  $X_B$  attaches to the granule surface and is then partially hydrolyzed during the anaerobic phase [de Kreuk et al., 2010]. The fraction of  $X_B$  that is not converted anaerobically is then degraded under aerobic conditions, thus promoting the growth of OHO and filamentous outgrowth (finger-type) [Pronk et al., 2015a]. Such short-term mechanism might however not be relevant for real AGS systems, in which the sludge is exposed to  $X_B$  over long-term (several months/years). Over long-term, the presence of  $X_B$  in the influent favours the presence of flocs. Filamentous structures in turn do not develop at the surface of granules as  $X_B$  is likely captured/degraded by flocs Derlon et al. [2016], Wagner et al. [2015b].  $X_B$  attachment onto the granules surface is likely limited during anaerobic PF-feeding, while hydrolysis rate is in addition very low, especially under anaerobic conditions [Jabari et al., 2016]. For these reasons, it is expected that significant amount of  $X_B$  remains available during the mixed aerobic phase. We hypothesized that under mixed conditions, flocs would have a competitive advantage over granules for capturing and then degrading  $X_B$  [Derlon et al., 2016, Wagner et al., 2015b]. The surface-to-volume ratio of the flocs is much larger than the one of the regular and round shaped granules [Andreadakis, 1993, Mihciokur and Oguz, 2016]. We propose that the availability of  $X_B$  during the mixed aerobic phase combined with its selective

### **Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater**

---

capture by flocs is the main reason for their presence in AGS systems treating municipal WW. We ultimately suggest that flocs might in fact be beneficial during treatment of complex WW with elevated levels of non-diffusible  $X_B$  using AGS technology.

The microbial communities in the flocs and granules differed only for some particular genera at stable state. Thus the microbial community structures evolved in a similar manner in both fractions, probably due to a constant exchange of biomass between the two fractions [Liu et al., 2010, Zhou et al., 2014]. For the four reactors, our results indicate a higher fraction of *Zoogloea* in flocs than in granules. The enrichment of *Zoogloea* in flocs was particularly pronounced in the AGS of R1, where the proportion of flocs was the lowest (5 %) and where filamentous outgrowth was often observed. The presence of *Zoogloea* in flocs likely resulted from erosion of the granules surface [Szabo et al., 2017a]. Erosion of the granules surfaces might have been favoured by the low cohesion of filamentous structures resulting from the growth of *Zoogloea*. On the contrary, the higher abundance of slow growing organisms in granules, e.g., *Nitrospira* likely resulted from favourable growth conditions in the core of the granule. The SRT gradually increases over the granules depth, thus providing suitable growth conditions for slow growing organisms such as nitrifiers. High cohesion within granules also reduces the detachment rate and thus the exchange of bacteria from the granules to the flocs and vice versa. In this case, the differences between the bacterial communities of granules and flocs can become significant, because both fractions offer different niches and have different retention times [Winkler et al., 2012]. It is therefore expected that slow growing bacteria are progressively enriched in the granules. Such mechanism is confirmed by our measurements of relative abundance of *Nitrospira*, *Ca. Competibacter* or *CPB\_CS1*. Moreover, the lowest differences between the microbial communities of flocs and granules in R3 can be explained by granulation being more recent compared to the other systems.

#### **III.5.5 Implications for research and practice**

Our findings have relevant implications for both research and engineering practice. The complex synthetic WW resulted in the development of AGS that was more similar to AGS fed by municipal WW rather than AGS fed by 100%-VFA. Therefore we advise the use of complex synthetic WW (VFA, diffusible fermentable substrate and high non-diffusible  $X_B$  contents) as surrogate of municipal WW. The proportion and composition of  $X_B$  can be tuned and modified to address specific research questions.

In terms of implications for engineering practice, this study provides relevant information for the start-up of AGS systems. The superior performance of AGS in R4 over AGS in R3 indicates that an increased loading, was beneficial for granulation and nutrient-removal, despite the significant increased fraction of  $X_B$  in the influent. However, a balance must be found between applying a sufficient organic loading while limiting operating costs (e.g., aeration). Start-up time was likely extended due to the chosen start-up strategy (low washout stress and low  $v_{ww}$  during PF feeding) in comparison to other studies [de Kreuk and van Loosdrecht, 2006, Derlon

et al., 2016]. But operating the system at low  $v_{ww}$  also helped maintaining high substrate and nutrient removal rates during the entire experimental phase, as observed in the present study and reported in literature Derlon et al. [2016], Lochmatter and Holliger [2014]. Applying a higher selective pressure by gradually increasing the  $v_{ww}$  during PF feeding would accelerate the formation of granules. But based on our experience, increasing the selective pressure applied via  $v_{crit}$  is coupled with an increase risk of biomass loss. Due to the sensitivity of AGS systems fed by low-strength municipal WW, a fine balance between maintaining high SRT conditions for forming granules and applying a selective sludge removal of slow settling biomass must be found. In terms of nutrient removal, partial denitrification due to poor SND proved to be representative of AGS systems fed with non-diffusible  $X_B$ . An increased TN-removal in those systems was achieved by implementing an additional mixed phase following the PF feeding. Denitrification during this phase likely benefited  $PO_4$ -P removal as observed for R3 and R4.

## III.6 Conclusions

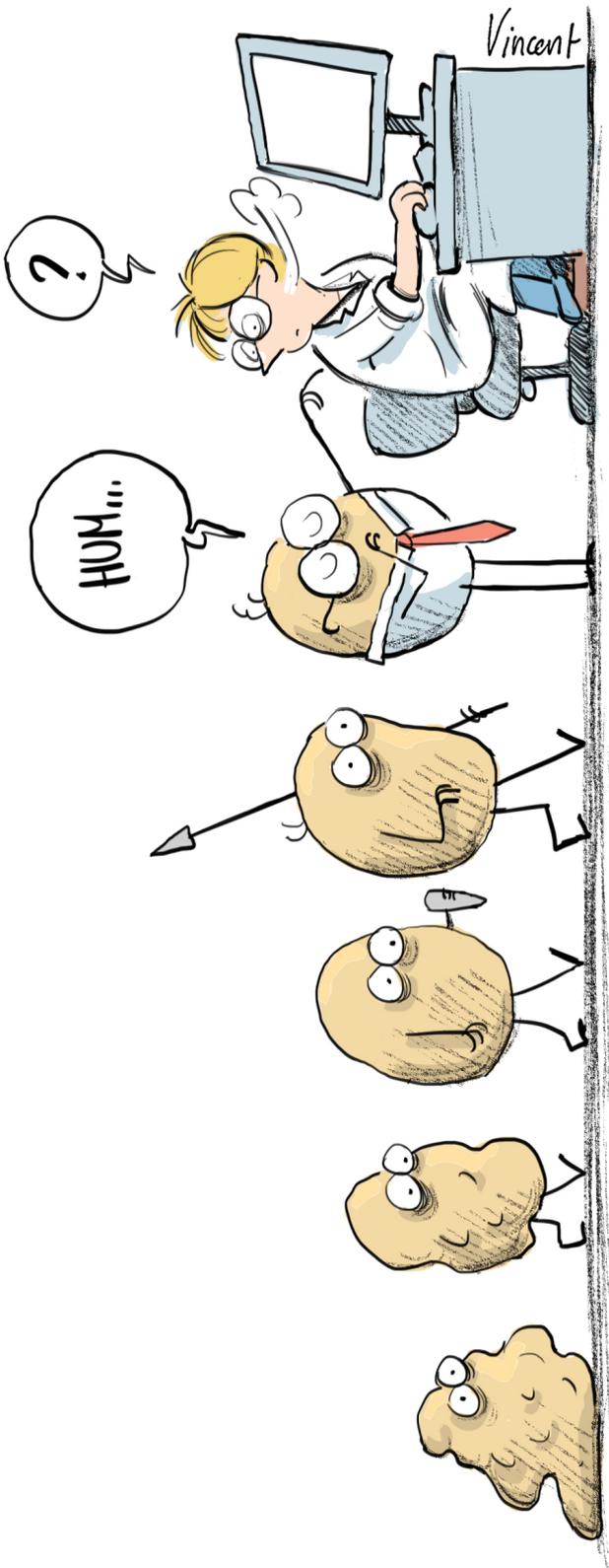
The main conclusions of this study are:

1. The wastewater composition in terms of diffusible and non-diffusible organic substrates governs both the microbial community composition, granulation kinetics, settling properties, and nutrient removal of AGS. High fractions of diffusible organic substrates results in fast granulation and excellent settleability of AGS, whereas presence of non-diffusible  $X_s$  in the influent hampers granulation, reduces settleability and results in the presence of substantial fractions (20-40 %) of flocs. The bacterial communities of flocs and granules were globally very similar within the same reactor, but several taxa were enriched in flocs or granules, respectively.
2. AGS fed by VFA based synthetic WW, resulted - as expected - in a specialized bacterial community containing classical PAO and GAO (e.g., *Ca. Accumulibacter*, *Ca. Competibacter* or *CPB CSI*) that led to fast granulation, excellent settling performance, stable nutrient removal, large granules and a quasi-absence of flocs.
3. AGS fed by complex substrates, containing non-diffusible  $X_B$ , revealed bacterial communities characterized by a high abundance of fermenting bacteria, including fermentative PAO and GAO. High amounts of diffusible organic substrate and total organic load were key factors to enhance the settleability and granulation kinetics of the sludge as well as the stability of nutrient-removal performances.
4. An increased floc fraction was constitutive in AGS reactors fed with complex WW and was attributed to the presence of non-diffusible  $X_B$  in the influent. It is neither possible nor desirable to wash out all flocs in AGS systems fed with complex WW.
5. Complex synthetic WW led to AGS with characteristics resembling those treating raw

### **Chapter III. Transforming activated sludge into granular sludge in lab-scale reactors fed with synthetic and real wastewater**

---

municipal WW. Hence, the often applied 100%-VFA synthetic WW should be replaced by complex synthetic WW as a surrogate of municipal WW in future lab-scale experiments studying AGS.





# **IV Transforming acetate-propionate-grown granules into AGS treating fermentable and polymeric organic substrates**

## **IV.1 Introduction**

### **IV.1.1 Effect of polymeric compounds on AGS**

In the experiment presented in chapter III, activated sludge from a wastewater treatment plant performing enhanced biological phosphorus removal (EBPR), were transformed into aerobic granular sludge with different wastewater types. The proportion of diffusible chemical oxygen demand (COD) in the substrate was linked to fast granulation and good settling properties. The experiments presented in the current chapter aimed at studying the effect of the COD diffusibility on the microbial communities, the settling performance and the nutrient removal of a sludge which is already in the form of granules. The aerobic granular sludge (AGS) with which the experiment started was adapted to a simple monomeric wastewater containing acetate and propionate as carbon source which favors the growth of classical PAO and GOA with anaerobic feast and aerobic famine mode of reactor operation. It was not clear if the stability of the granular biofilm would be maintained as the carbon content of the wastewater would be progressively changed. In the experiment of chapter III, the inoculum activated sludge was adapted to municipal wastewater, whereas in the present study, the sludge was acclimated to simple monomeric synthetic wastewater. It is expected that a transition from simple to complex wastewater is more susceptible to disrupt the settling properties and nutrient removal performance of the sludge than the opposite change. A sludge treating a complex wastewater such as the ones used in these studies, will contain microorganism able to take up volatile fatty acids (e.g. acetate and propionate). It is not certain that a sludge treating a simple wastewater containing only acetate and propionate will contain a sufficient diversity of microorganisms able to ferment and hydrolyze the carbon content of a complex wastewater, so that the microbial community can adapt to this type of substrate.

## **Chapter IV. Transforming acetate-propionate-grown granules into AGS treating fermentable and polymeric organic substrates**

---

Certain bacteria have a versatile metabolism that allows them to adapt to different conditions [Jefferson, 2004]. It is therefore possible that bacteria having the capability to use fermentable and/or polymeric compounds persist in a system with VFA as only carbon source. These bacteria could also feed on sugars and amino acids contained in the EPS matrix secreted by VFA consumers. Moreover, the water used as basis of the wastewater was industrial water pumped from the lake. This water may contain various microorganism, thus allowing the immigration of these microorganisms in the system under favorable conditions.

### **IV.1.2 Resilience, resistance and multistability**

Three concepts of microbial ecology are particularly interesting when studying the response of microbial communities to changes of the 'environment' : resilience, resistance and stability. These concepts can be applied to the composition of the community and/or to its functioning [Holling, 1996]. From an engineer's point of view, it is important to know if the functions of a process will be maintained or recovered quickly after a perturbation or a change of the environment. For the researcher, the link between microbial and functional stability is important, and it is key to understand how a microbial community changes in response to variations in an environmental parameter, and how these changes impact the global function of this community. The sensitivity of a process efficiency to a change of microbial community composition will determine the importance of the stability of the community [Allison and Martiny, 2008].

The resilience was defined by Holling [1973] as "the ability of a system to absorb change and disturbance, and still maintain the same relationships between populations or state variables". The resistance is the level to which a microbial community or its functioning remains constant when facing a perturbation [Allison and Martiny, 2008, Botton et al., 2006]. The stability of a system is its ability to recover a state close to its former equilibrium state after a temporary perturbation [Holling, 1973, Walter, 1980]. Although these concepts are closely related, they are not equivalent. For example, the instability of a community can favor its resilience [Holling, 1973].

Microbial communities are generally highly sensitive to external changes [Allison and Martiny, 2008], some can recover quickly [Shade et al., 2012], whereas others not [Dethlefsen and Relman, 2011]. The non-reversibility of a stable state after a perturbation can be due to stochastic effects, or other factors such as the priority effect, which describes the exclusion of new arrivals by the bacteria already present in a specific niche [Andersson et al., 2014]. This can lead to the existence of multiple stable states for a same ecosystem. Multistability is observed when a small perturbation can lead to a different equilibrium state that will persist even when the external conditions will be back to normal, *as discussed in [Petraitis and Dudgeon, 2004]*. It is difficult to authenticate multistability since small changes in the environment can remain unnoticed while having an impact on the microbial community [Faust et al., 2015, Gonze et al., 2017].

The functional stability of a microbial ecosystem often correlates with its biodiversity [Siripong and Rittmann, 2007, Wittebolle et al., 2009, 2008]. More specifically, this stability is thought to be mainly the result of functional diversity [Briones and Raskin, 2003, Hulot et al., 2000, McCann, 2000, von Canstein et al., 2002]. Yet, the functional redundancy of a microbial community is often difficult to measure because the metabolism(s) of an important part of a microbial ecosystem is generally unknown and often difficult to assess [Allison and Martiny, 2008, Botton et al., 2006], in particular because the functions of microorganisms are not homogeneously distributed across the phylogeny [Achenbach and Coates, 2000]. It is therefore important to increase our knowledge on the functioning of the different taxa composing a microbial community to predict its resilience, resistance and stability.

### IV.1.3 The role of flocs in AGS

AGS treating complex polymeric wastewater generally contain a non negligible proportion of flocs (around 20 %) [Derlon et al., 2016, Pronk et al., 2015b, van Dijk et al., 2018a]. The question whether flocs and granules play a different role in granules have been raised (Chapter III). A role could arise from different physical structures providing different capabilities to capture particles [Liu and Tay, 2012], and/or from different microbial communities. No clear consensus exists on the difference of microbial communities between granules and flocs. Liu et al. [2017] found the microbial communities to be dramatically different between granules and flocs, with only 817 OTUs in common on a total of 2241 in granules and 2009 in flocs, whereas Zhou et al. [2014] found a great similarity between the microbial communities of the two fractions during steady state. These two results are not necessarily contradictory since it was shown that there are different coexisting granulation mechanisms (eg. flocs densification, aggregation of granular biomass and/or floccular biomass, microcolony outgrowth. [Barr et al., 2010a, Cetin et al., 2018, Gonzalez-Gil and Holliger, 2014, Zhou et al., 2014]. Yet the predominant mechanism can vary according to the microbial community and the operational conditions (eg. shear stress) [Weissbrodt et al., 2013]. The comparison of the microbial communities of granules and flocs across multiple samples collected from AGS treating different types of wastewater may provides elements of answer concerning a potentially specific role of flocs in AGS.

### IV.1.4 Hydrolysis in AGS

Previous studies on AGS have highlighted the importance of complete COD-removal during the anaerobic phase to favor the growth of slow growing organisms such as PAO and GAO to the detriment of fast growing ordinary heterotrophs thus promoting granulation and biological P-removal [Weissbrodt et al., 2012]. Yet, it is not clear if a total COD removal is possible during a typical lab-scale reactor anaerobic phase (60 to 90 min) when a significant part of the COD is in a polymeric form [Morgenroth et al., 2002]. Indeed, hydrolysis was identified as the rate limiting step for the biological degradation of organic carbon during wastewater treatment [Dueholm et al., 2001, Morgenroth et al., 2002, San Pedro et al., 1994]. Moreover, depending on

## **Chapter IV. Transforming acetate-propionate-grown granules into AGS treating fermentable and polymeric organic substrates**

---

their sizes, the polymeric compounds cannot diffuse, or diffuse more slowly than monomeric compounds into the granules and must therefore be hydrolyzed at the surface [Larsen and Harremoës, 1994, Mosquera-Corral et al., 2003]. Small, fluffy granules and flocs thus gain a competitive advantage over big smooth granules with the presence of polymeric compounds [de Kreuk et al., 2010, Martins et al., 2004, Xia et al., 2007]. It is therefore expected that the introduction of polymeric compounds in the wastewater will have a negative influence on the granulation and nutrient removal efficiency.

In wastewater treatment plants, hydrolysis is performed by various microorganisms including bacteria by means of exoenzymes [Confer and Logan, 1998, Goel et al., 1998a]. Several studies have investigated the identity of the bacteria capable of excreting these exoenzymes [Xia et al., 2007, 2008a,b]. They found that members of *Tetrasphaera*, have the capability to hydrolyze starch, while protein hydrolyzers have been detected in the genus *Ca. Epiflobacter*, the phyla of Chloroflexi, Betaproteobacteriales and Saccharibacteria.

Big differences in the physical characteristics (size, density), the settling properties and the nutrient removal efficiency of AGS grown in lab-scale reactors treating simple monomeric wastewater and full-scale tank treating real wastewater have been reported [Wagner et al., 2015b, Wang et al., 2018, Winkler et al., 2013]. Yet it is not simple to untangle the effects of the different variables on the AGS characteristics. Moreover, the microbial communities of activated sludge performing EBPR with municipal wastewater have been extensively studied, and they turned out to be significantly different from the ones treating simple monomeric synthetic wastewater. To date, only few studies investigated the microbial communities of AGS treating complex wastewater [Cetin et al., 2018, Kang et al., 2018, Swiatczak and Cydzik-Kwiatkowska, 2018, Wang et al., 2018]. As for EBPR activated sludge, they are likely different from the ones treating simple wastewater, which contain mainly classical PAO and GAO from the class Gammaproteobacteria known to favor granulation and good nutrient removal [de Kreuk and van Loosdrecht, 2004].

### **IV.1.5 Objectives**

The experiments presented in the current chapter investigated separately the impact of fermentable and polymeric compounds on the bacterial communities of an AGS grown with simple monomeric wastewater. The changes in wastewater composition were performed progressively over several weeks in order to promote at the most the maintenance of granulation and nutrient removal capability of the sludge. The effect of these transitions on the AGS were investigated by monitoring the bacterial communities, the nutrient removal performance and the sludge volume index of the sludge. Statistical analysis were used to investigate the links between wastewater composition, bacterial communities, settling properties and nutrient removal performance in AGS.

## IV.2 Material and methods

### IV.2.1 Reactors operation

The AGS was studied in bubble column SBR reactors with a diameter of 6.2 cm and a working volume of 2.4 L similar to the ones described in [Lochmatter and Holliger, 2014, Weissbrodt et al., 2012]. Two reactors, RA and RB, were operated in parallel, sometimes under the same operation conditions, sometimes under different ones. The reactor RB had been initially inoculated in spring 2014 with activated sludge collected in a wastewater treatment plant performing enhanced biological phosphorus removal (ARA Thunersee, Switzerland). The synthetic wastewater used until the beginning of the experiment was composed of acetate and propionate as carbon source in equal proportions for a total COD of  $400 \text{ mg O}_2 \text{ l}^{-1}$ ,  $22 \text{ mg l}^{-1}$  of  $\text{P-PO}_4$ ,  $56 \text{ mg l}^{-1}$  of  $\text{N-NH}_4^+$ , and trace elements.

During experiment 1 (IV.1), the composition of the wastewater was progressively changed from simple monomeric to complex monomeric : on day 77, 96 and 109, approximately 20%, 40% and 66% of the COD, respectively, was replaced by glucose and amino acids in equivalent COD proportions. From day 121 on, the ammonium concentration was adapted to counterbalance the ammonium produced by amino acids degradation. The TN concentration of the medium was fixed to  $56 \text{ mg l}^{-1}$ .

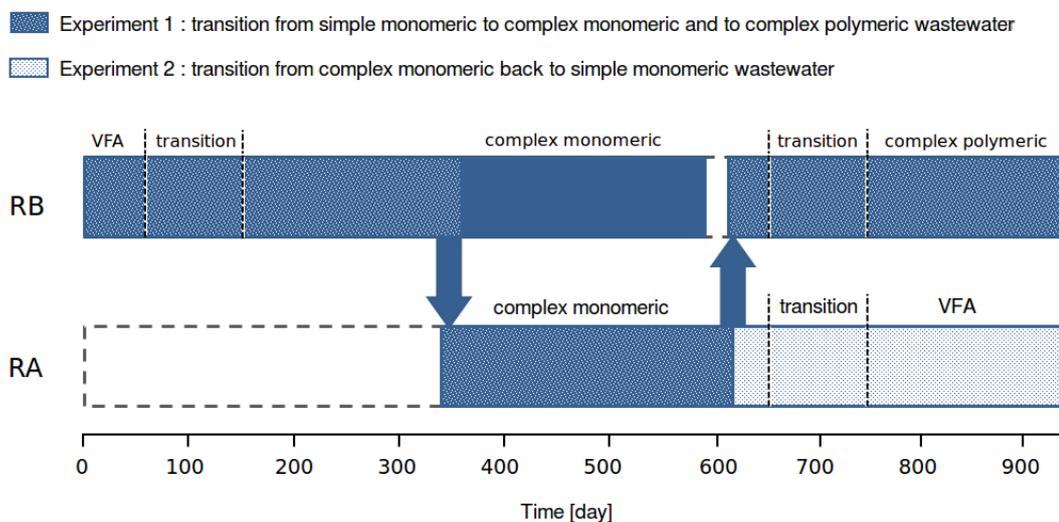


Figure IV.1 – Summarizing diagram of the reactor operation and the two experiments

The reactor was duplicated on day 333 by distributing equally the biomass of RB in the two reactors RA and RB, and they were run in parallel with the complex monomeric wastewater. Several incidents occurred during the following 300 days. The presence of nematodes was detected during this period in AGS of the two reactors and the whole biomass of the two reactors was crushed in a glass homogenizer (see IV.2.5) on day 539 in RA and on days 560 and 798 in RB. Problems with the pH regulation caused two pH incidents, characterized

#### **Chapter IV. Transforming acetate-propionate-grown granules into AGS treating fermentable and polymeric organic substrates**

---

by a pH above 11 in the bulk water, on day 422 and 529 in RA. The influent wastewater lacked trace element solution from day 564 to 582 in RB and from day 578 to day 582 in RA. Following this, the decision was taken to stop RB and restart it with half of the biomass from RA. The composition of the wastewater was then changed from complex monomeric to complex polymeric, in RB : on day 639, 680 and 729, 10%, 20% and 33% of the COD in the form of glucose and amino acids was replaced by starch, a polymer of glucose (solubility: 20 mgml<sup>-1</sup>, Sigma-Aldrich), and peptone, a mixture of peptides of different length and amino acids (<10KD, BD Peptone Bacto, Fisher scientific), in equivalent COD proportions. During experiment 2 ( IV.1), the composition of the wastewater was progressively changed from complex monomeric back to simple monomeric, in RA : on day 638, 684, and 732, the proportion of COD in the form of glucose and amino acids was reduced to 40%, 20% and 0%, by replacing it with acetate and propionate.

The temperature was kept constant by recirculation of 18 °C water in the double wall of the reactor. The pH of the bulk water was monitored and regulated with an ISFET probe (Endress+Hauser, Switzerland) using a PID process controlling the injection of NaOH or HCl solutions (1%) during the periods of mixing. The pH was maintained between 6.5 and 8 from day 0 to day 119 and between 7.2 and 8.5 from day 120. The pO<sub>2</sub> and the conductivity were monitored by two ISFET probes (Endress+Hauser, Switzerland).

The monitoring of information collected by the reactor probes and the control of the different pumps was relayed to/from a computer through relay modules (WAGO, Switzerland) and processed using the software DAQFactory (Windows version 5.86, 2011, AzeoTech, Inc.)

A typical cycle consisted of four phases: an anaerobic feeding phase and an aeration phase, followed by a settling and withdrawal phase. During the anaerobic phase, the synthetic wastewater was injected from the bottom during 12 minutes. Then the reactor was mixed by injecting 2 L/min of N<sub>2</sub> during 60 to 90 minutes. During the aerobic phase, compressed air was injected with a flow rate of 2 L/min. From day 48 on, periods of 10 minutes of aeration were alternated with idle periods of 10 minutes in order to optimize denitrification [Lochmatter et al., 2014]. The aeration phase stopped when a sharp decrease of O<sub>2</sub> consumption was detected, with a maximum time of five hours. Half of the reactor content was withdrawn after 2 to 5 minutes of settling and N<sub>2</sub> was injected in the reactor to restore anaerobic/anoxic conditions for the following cycle. On day 157, headspace gas recirculation was introduced with a PID controlling the inflow of incoming gases, in order to save nitrogen and compressed air consumption.

The length of the different phases were adapted to allow an optimal functioning of the AGS reactors. For example, the length of the anaerobic phase was increased from 72 up to 102 minutes mostly during changes of wastewater composition in order to obtain a COD in the bulk water below 20 mg O<sub>2</sub>/L before start of aeration. The settling time was adapted to the settling properties of the biomass in order to select fast settling granules while maintaining sufficient biomass in the reactors.

## IV.2. Material and methods

Table IV.1 – Theoretical characteristics of the different influent synthetic wastewater in  $\text{mg l}^{-1}$ . The theoretical concentration of phosphate was  $22 \text{ mg l}^{-1}$  of  $\text{P-PO}_4$  in all the wastewaters.

denomination	medium name	COD [ $\text{mg O}_2\text{l}^{-1}$ ]	VFA [% of COD]	glucose	starch	amino acids	peptone	TKN [ $\text{mg N l}^{-1}$ ]
simple	1	456	100	-	-	-	-	56
transition from simple to	2	448	82	9	-	9	-	63
	3	439	62	19	-	19	-	69
	4A	427	36	32	-	32	-	78
complex monomeric	4B	600	33	33	-	33	-	89
	4C	600	33	33	-	33	-	56
transition from complex monomeric to	5	600	33	28	5	28	5	56
	6	600	33	23	10	23	10	56
complex polymeric	7	600	33	17	17	17	17	56
transition from complex monomeric to	3B	600	60	20	-	20	-	56
	2B	600	80	10	-	10	-	56
simple	1B	600	100	-	-	-	-	56

### IV.2.2 Wastewater composition

The influent wastewater was a mixture of 0.135 l of concentrated autoclaved C-medium, 0.135 l of concentrated NP-medium and 0.93 L of industrial water filtered at  $100 \mu\text{m}$  from day 0 to 515, and at  $5 \mu\text{m}$  from day 515 to day 925.

The composition of the different C- and NP-media are presented in supplementary material (Table A.13).

An aliquot of 50 ml trace element solution containing  $16.22 \text{ g l}^{-1}$  of  $\text{C}_{10}\text{H}_{14}\text{N}_2\text{Na}_2\text{O}_8 \cdot 2\text{H}_2\text{O}$ ,  $0.44 \text{ g l}^{-1}$  of  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ ,  $1.012 \text{ g l}^{-1}$  of  $\text{MnCl}_2 \cdot 6\text{H}_2\text{O}$ ,  $7.049 \text{ g l}^{-1}$  of  $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ ,  $0.328 \text{ g l}^{-1}$  of  $(\text{NH}_4)_6\text{Mo}_7\text{O}_{24} \cdot 4\text{H}_2\text{O}$ ,  $0.314 \text{ g l}^{-1}$  of  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  and  $0.322 \text{ g l}^{-1}$  of  $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$  was added to the NP-medium for a final volume of 10 l.

The resulting characteristics of the different influent synthetic wastewaters are presented in table IV.1.

### IV.2.3 Nutrient-removal performance monitoring

The nutrient removal performance of the reactors was measured on a bi-weekly basis on the influent wastewater (I), the water collected from the reactor at the end of the anaerobic phase (AN) and three minutes before the withdrawal, considered as similar to the effluent of the reactor (S). Water samples were filtered at  $45 \mu\text{m}$  and stored at  $4 \text{ }^\circ\text{C}$  before further analysis. VFA concentrations were measured by high performance liquid chromatography (HPLC, Jasco Co-2060 Plus, refraction index detector, Omnilab, Germany) with organic acid ion-exclusion column (ORH-801, Transgenomics, United Kingdom). Glucose concentrations were assessed by spectrophotometry using the kit GlucoQuant (Roche/Hitachi, Germany) according to the manufacturer instructions. Ammonium concentrations were measured by spectrophotometry, using two different kits; the ammonia cuvette test ( $0.015\text{-}2.0 \text{ mg l}^{-1} \text{ N-NH}_4^+$ , LCK304, Hach,

## Chapter IV. Transforming acetate-propionate-grown granules into AGS treating fermentable and polymeric organic substrates

---

USA) for the samples with an ammonia concentration estimated below  $3 \text{ mg l}^{-1}$  of  $\text{N-NH}_4^+$ , and the Spectroquant ammonium test (photometric  $0.010\text{-}3.00 \text{ mg l}^{-1}$   $\text{N-NH}_4^+$ , Merck, Germany) for the other samples<sup>1</sup>. The concentrations of anions ( $\text{P-PO}_4^-$ ,  $\text{N-NO}_3^-$  and  $\text{N-NO}_2^-$ ) were measured by ionic chromatography (IC, ICS-90, IonPacAS14A column) with an electrical conductivity detector (Dionex, Switzerland).

### IV.2.4 Measures of SVI, TS and VS

The sludge volume index (SVI), the volatile solids (VS), and the total solids (TS) were measured according to the standard method [APHA, 2012]. Approximately 500 ml of mixed liquor were collected from the reactor during mixing. The volume occupied by the sludge bed was measured in a graduated cone after 3, 5, 8, 10 and 30 minutes. This mixed liquor was mixed again and three aliquotes of approximately 20 ml were collected in falcon tubes and centrifuged at  $8392 \times g$  during five minutes. For each aliquote, the mass of the dried pellet after 12h of drying at  $105^\circ\text{C}$  gave the TS and the mass loss after two hours of calcining at  $550^\circ\text{C}$  gave the VS. The SVI corresponding to the different times were calculated as the ratio between the volume of the sludge bed ( $V$ ) and the total volume ( $V_{\text{tot}}$ ), divided by the mean TS of the triplicates IV.1.

$$\text{SVI}(\text{ml g}^{-1}) = \frac{V(\text{ml})}{\text{TS}(\text{g ml}^{-1}) \cdot V_{\text{tot}}(\text{ml})} \quad (\text{IV.1})$$

### IV.2.5 Nematodes monitoring and controlling

The presence and abundance of nematodes in the sludge was monitored by visual inspection with light microscopy (Eclipse NI-U Microscope, Nikon). In case of too high abundance of nematodes, their number was reduced by centrifuging the biomass with an Avanti J-26 XPJ (Beckmann Coulter) with a JLA rotor ( $8.1000$ , Beckmann Coulter) at  $15000 \times g$  and then crushing this biomass with a glass homogenizer having a distance between the pestle and the tube between  $0.15$  and  $0.25 \text{ mm}$  ( $50 \text{ cm}^3$ , Carl Roth, Germany). The homogenization lyzed and hence killed the nematodes.

### IV.2.6 Biomass sampling

AGS biomass was collected from the reactor(s) on a weekly basis according to the following procedure. Mixed liquor containing between 1 and 2 ml of wet biomass was centrifuged during five minutes at  $8392 \times g$  (Nuair Awel CF-48R centrifuge, USA) and washed twice with 5 ml of phosphate buffer saline (PBS,  $8 \text{ g l}^{-1}$  of NaOH,  $0.2 \text{ g l}^{-1}$  of KCl,  $1.44 \text{ g l}^{-1}$  of  $\text{Na}_2\text{HPO}_4$ ,

---

<sup>1</sup>An interference of amino acids with the compounds of the Spectroquant kit lowering the ammonium concentrations detected was noted on non-diluted samples.

0.24 g l<sup>-1</sup> of KH<sub>2</sub>PO<sub>4</sub>) at 4 °C followed by 5 minutes of centrifugation at 8392 x g. The pellet was resuspended in 3 ml of PBS, homogenized with a glass homogenizer having a distance between the pestle and the tube between 0.15 and 0.25 mm (5cm<sup>3</sup>, Carl Roth, Germany) and distributed in 5 cryotubes. The samples were stored at -20 °C until DNA extraction. When a sufficient proportion of the biomass was in the form of flocs, granules and flocs, separated samples were collected by using a 250 µm sieve before washing and homogenization. In this case total biomass was also collected.

### IV.2.7 DNA extraction and 16S rRNA gene sequencing

For each sample, 200 µl of homogenized biomass was mixed with 400 µl of elution buffer (T<sub>10</sub>E<sub>0.1</sub>) and 100 µl of lysozyme solution (25 mg ml<sup>-1</sup>). After one hour at 37 °C, DNA was extracted by using the automatic robot 16 DNA purification system (Maxwell, Promega Corporation, Switzerland). The DNA concentration of each DNA extraction was determined by spectrophotometry with a NanoDrop (ND1000, Witec AG, Switzerland).

For reasons out of our control, two different Illumina MiSeq sequencing platforms, and therefore two slightly different protocols were used for 16S rRNA gene amplicon sequencing. In both cases, the hypervariable regions V1-V2 of the bacterial 16S rRNA gene were amplified by polymerase chain reaction (PCR) using the universal bacterial primers 27F and 338R (see sequence below). The amplified and barcoded DNA was sequenced in multiplexed by groups of 60 (run1) or 96 samples (other runs), in paired-end mode (2x250 bp).

#### Protocol n°1

Runs 1, 2 and 3 (corresponding to samples from days 0 to 385) were performed at the Laboratory of Intestinal Immunology (UPHARRIS, EPFL). The indexing of the sample was performed during the PCR by using a 27F primer and 60 (run1) or 96 (other runs) different barcoded 338R primers (in bold) with overhang containing Illumina adapters. 5'GAGATCTACACTATGGTAATTCC -**AGMGTTYGATYMTGGCTCAG**3' and 5'CAAGCAGAAGACGGCATAACGAGAT -barcode -AGTCAGTCAGAA -**GCTGCCTCCCGTAGGAGT**3', respectively. The DNA was amplified according to the procedure described in section III.3.8. The amplified DNA quantified on a labCHip GX (Perkinelmer), with 1 to 2 µl of each PCR product and the kit DNA 5k (CLS760675). The samples were then pooled to a final concentration of 4 nM, purified by using Agencourt AMPure XP beads (Beckman Coulter, A63880) and sequenced.

#### Protocol n°2

Runs 7, 8 and 11 (corresponding to samples from days 392 to 925 and some replicates of samples corresponding to day 0 to 385) were performed at the Genomic Technologies Facility (GTF, University of Lausanne, Switzerland). A first PCR was performed with the 27F and 338R primers (in bold) with overhang adapter sequences 5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG

## **Chapter IV. Transforming acetate-propionate-grown granules into AGS treating fermentable and polymeric organic substrates**

---

-AGMGTTYGATYMTGGCTCAG3' and 5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG - GCTGCCTCCCGTAGGAGT3', respectively. The same procedure as in the protocol n°1, and described in section XX was followed. The PCR products were purified by using Agencourt AMPure XP beads (Beckman Coulter, A63880), quantified with a fragment analyzer using the DNF-473 standard sensitivity NGS fragment analysis kit (Advanced analytical Technologies Inc.). These amplified samples were transmitted to the GTF for a secondary indexing PCR and multiplex sequencing.

### **IV.2.8 Taxonomic affiliation of 16S rRNA gene sequences**

The taxonomic affiliation of the 16S rRNA gene sequences was performed according to the procedure described in section III.3.8.

### **IV.2.9 Statistical analysis**

The statistical analysis were performed as described in section III.3.8.

### **IV.2.10 Determination of stable states and discriminant taxa**

The Bray-Curtis matrix of the bacterial OTU (97%) relative abundance (supplementary material table A.14) was used to define the bacterial "stable states" of these experiments. A stable state was defined as the maximal cluster of successive samples having a pairwise distance lower than 0.5 (the maximum pairwise distance between two samples was 0.94). The seven stable states obtained were used in some of the statistical analyses. After Hellinger transformation, the mean relative abundance of each taxa (at the genus level) was compared, using t-tests, between the stable states grouped by influent wastewater types (simple monomeric, complex monomeric and complex polymeric). The taxa were considered "divergent" if their mean was significantly different (p-value 0.01 with Bonferroni correction for multiple testing,  $p\text{-value} = 0.01/292 = 3.45 \times 10^{-4}$ ) between at least two influent wastewater types. These taxa were considered "abundant" if their average abundance was higher than 1 % in the stable states of at least one wastewater type. The taxa being divergent and abundant were considered as "discriminant" in the following analysis.

### **IV.2.11 Comparison of the bacterial communities in flocs and granules**

The average proportions of the abundant taxa were compared by using t-tests within the stable state for which granules and flocs were collected separately for at least 4 samples. The difference was considered significant if the p-value was lower than the Bonferroni corrected p-value of 0.01 ( $p\text{-value} = 0.01/29 = 3.45 \times 10^{-5}$ ).

## **IV.3 Results**

### **IV.3.1 Nutrient-removal performance**

The overall nutrient removal performances were maintained despite the changes of wastewater types (Figure IV.2). Several operational incidents had major, but only temporary impact on these performances. The food to microorganism ratio was not adapted during the first duplication of the reactor, leading to high COD concentrations at the end of the anaerobic phase during 4-5 weeks. On day 422 and 529 important losses of the biomass occurred, but functional AGS unexpectedly recovered within the next 10-15 days. Problems in the pumping of NP-medium resulted in too low (day 377) or too high (day 592 to 594) ammonium and phosphate concentrations in the influent wastewater.

The  $\text{NH}_4$ -removal was complete during most of the experiment. However temporary decreases of this efficiency were noticed and this happened almost always due to the operational difficulties mentioned above and listed in the legend of Figures IV.2 and IV.3.

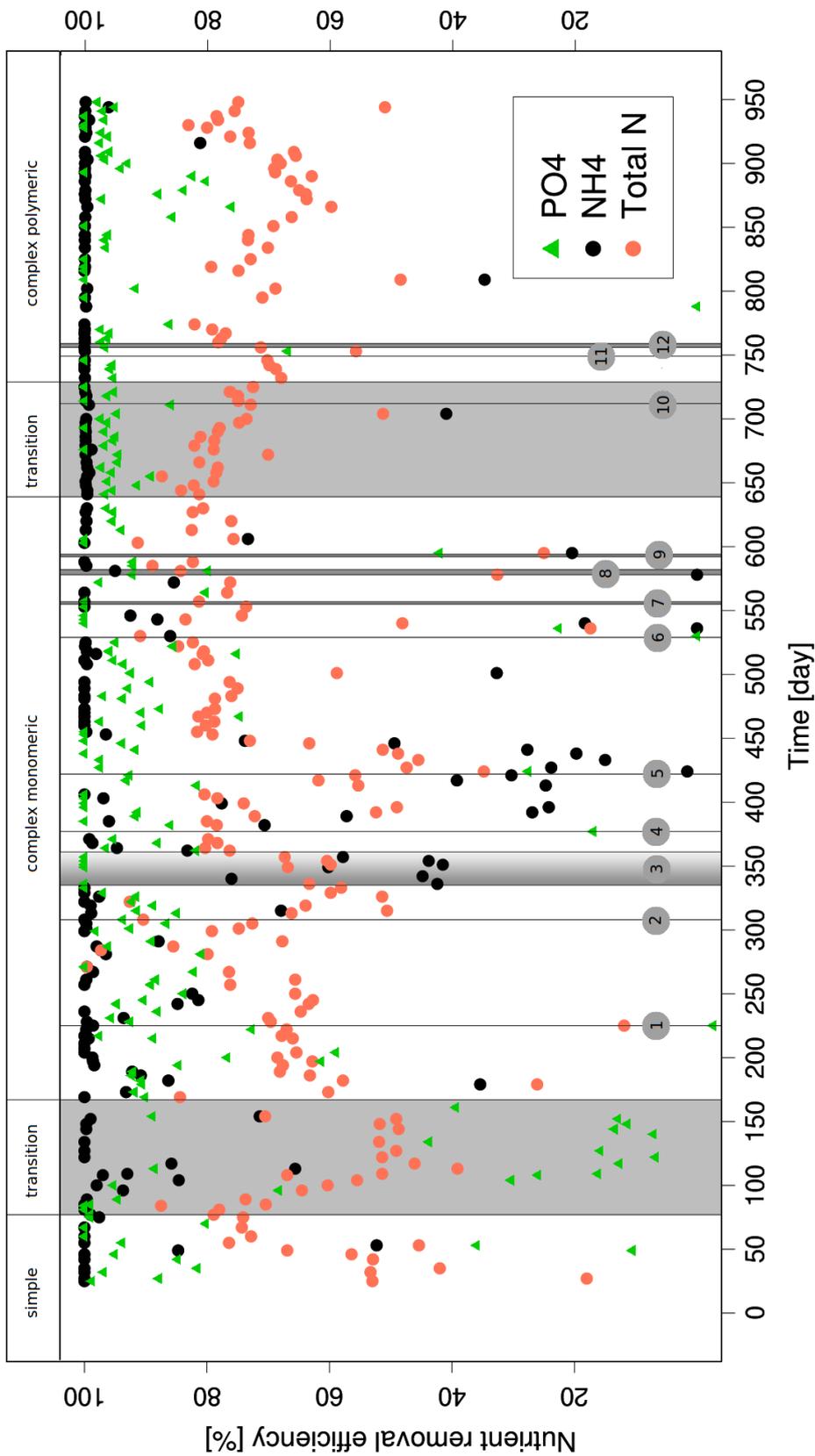


Figure IV.2 –  $\text{PO}_4^-$ ,  $\text{NH}_4^-$  and TN-removal performance of the AGS during experiment 1. The process events susceptible to having had a negative impact on nutrient removal performance are numbered from 1 to 12. (1) No C-medium in the influent wastewater during 8 cycles. (2) The leak of  $\text{O}_2$  detected since several days is corrected. (3) Few biomass in the reactors due to the duplication of the reactors and the feeding was not adapted. (4) Problem with pumping of NP-medium. (5) The pH of the bulk water was higher than 11 during at least two hours. (6) The pH of the bulk water was higher than 11 during a period of time estimated to 24h. (7) Problems with the recirculation pump possibly leading to lower shear stress during two days. (8) No trace elements in the influent during four days. (9) The influent was composed of twice the NP-medium and  $200 \text{ mg O}_2 \text{ ml}^{-1}$  of glucose during two days. (10) No feeding during the night. (11) Loss of biomass due to accidental change from 5 to 3 minutes of settling. (12) Lower concentrations of C-medium in the influent due to malfunctioning of the pumping.

The TN-removal was enhanced after the introduction of the alternated aeration during the aerobic phase at day 48. It decreased during the first transition due to the supplementary nitrogen input due to amino acids degradation. The transition from complex monomeric to polymeric wastewater went with a slight decrease of denitrification.

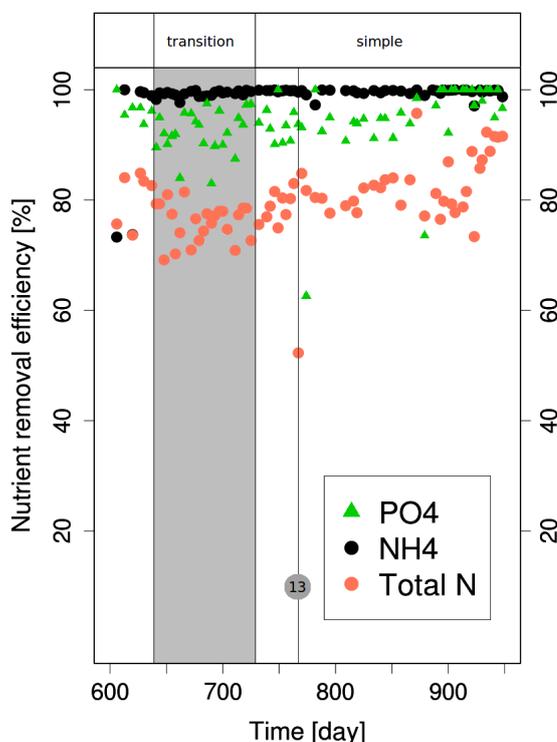


Figure IV.3 –  $\text{PO}_4^-$ ,  $\text{NH}_4$ - and TN-removal performances during the transition back from complex monomeric to simple monomeric influent wastewater. The vertical bar with the number (13) corresponds to the day with no feeding during several cycles.

The P-removal performances decreased to less than 20% during the transition from simple to complex monomeric influent wastewater. The increase of the total COD of the wastewater from 427 to 600 on day 145 was followed by a recovery of the P-removal efficiency (higher than 80 %). Aside from the operational difficulties mentioned above, the incidents impacting the P-removal were linked to the presence of  $\text{O}_2$  in the anaerobic phase (on day 308) or to incidents causing a decreased COD concentration in the wastewater (on day 313 and 751).

### IV.3.2 Settling properties

The  $\text{SVI}_3$  measures the volume occupied by 1 g of sludge after 3 minutes of settling and the ratio  $\text{SVI}_{30}/\text{SVI}_{10}$  indicates how much the biomass changed of volume between 10 and 30 minutes of settling. They were measured on the AGS during the transition from simple to complex monomeric wastewater and to complex polymeric wastewater (Figure IV.4) in order to assess the evolution of its settling properties during these two changes of wastewater

#### **Chapter IV. Transforming acetate-propionate-grown granules into AGS treating fermentable and polymeric organic substrates**

---

composition. Before the beginning of experiment 1, the AGS was fed during 18 months with a simple synthetic wastewater. At the beginning of the experiment, the AGS was composed almost exclusively of granules and had excellent settling properties with a sludge volume index below  $25 \text{ ml g}^{-1}$  and a SVI ratio of 1. The introduction of amino acids and glucose in the wastewater did not have a big impact on the settling properties, the  $\text{SVI}_3$  was still below  $25 \text{ ml g}^{-1}$  on day 124, and the SVI ratio was still 1. However, the big proportion of large granules became slightly lower between day 384 and 482 and the AGS did not recover its initial size distribution during these two experiments (Figure IV.5). Moreover, the AGS seemed to be less dense and tended to aggregate more during centrifugation after this first transition from simple to complex monomeric wastewater and during approximately one year, but no specific analyses have been carried out to characterize this phenomenon in more detail.

The homogenization of the whole biomass for nematode control affected the settling properties of the biomass. Indeed, the granules were destroyed by this operation and new granules were forming again after one month. On day 614, the settling properties were slightly lower, mainly due to the homogenization of the biomass two months earlier. The introduction of polymeric compounds was followed by a big increase of the sludge volume index. At first, there was no production of new granules and the biomass became floccular (pictures C and D, Figure IV.6). In the following, small granules began to form (on day 788, and again on day 848 after the second homogenization of the biomass), but a rather high proportion of flocs around 20 % remained, (picture E, Figure IV.6).

During the transition from complex back to simple monomeric wastewater, the settling properties of the AGS decreased slightly but the  $\text{SVI}_3$  remained around  $50 \text{ ml/g}$  and the SVI ratio was above 0.9.

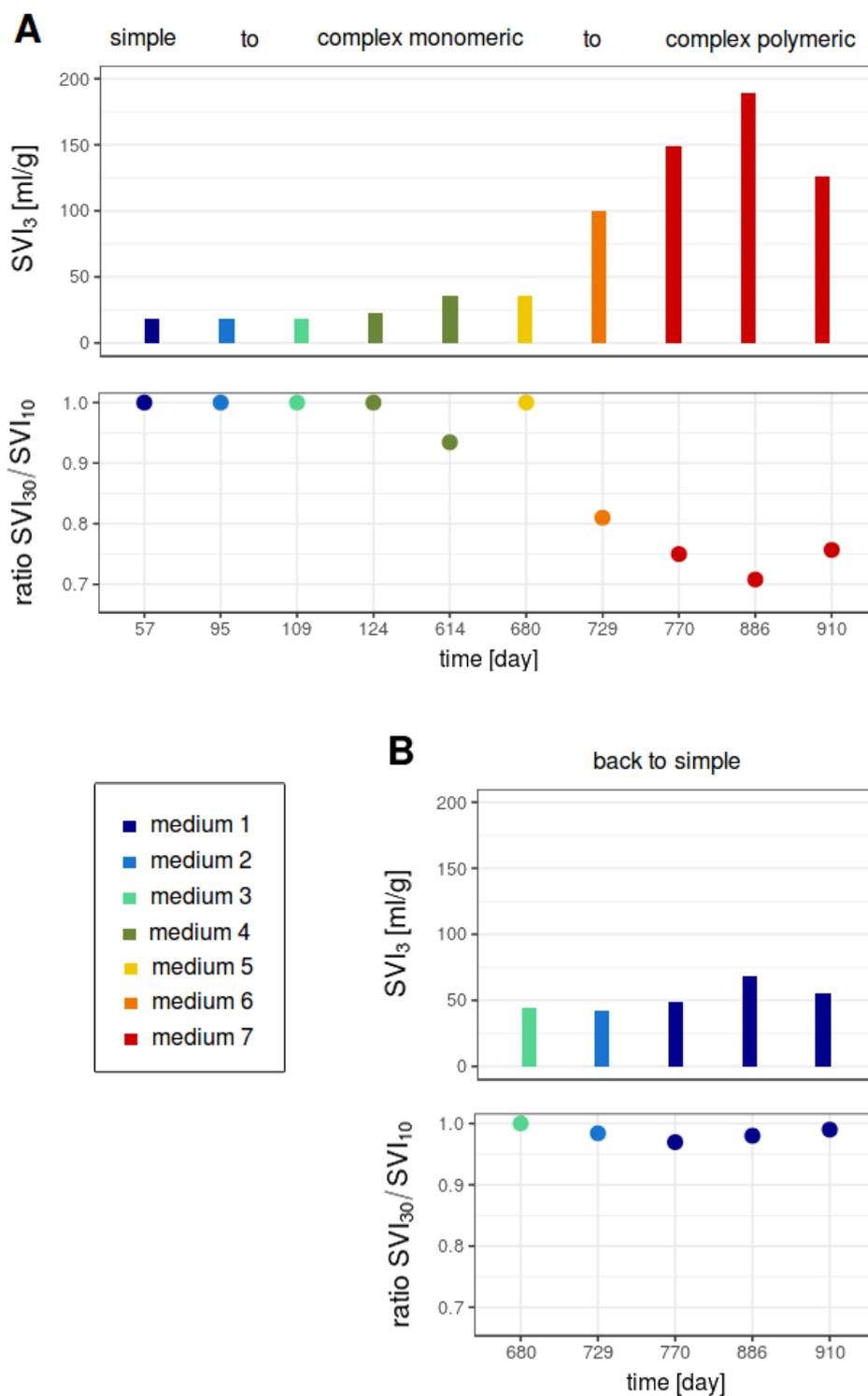


Figure IV.4 – Evolution of the settling characteristics of the sludge during the transition from simple to complex monomeric wastewater and to complex polymeric wastewater (A) and during the transition from complex monomeric back to simple wastewater (B). In both figures, the upper graph shows the SVI after 3 minutes, the lower graph shows the ratio of the SVI after 30 and 10 minutes. The colors indicate the medium feeding the sludge (IV.1).

**Chapter IV. Transforming acetate-propionate-grown granules into AGS treating fermentable and polymeric organic substrates**

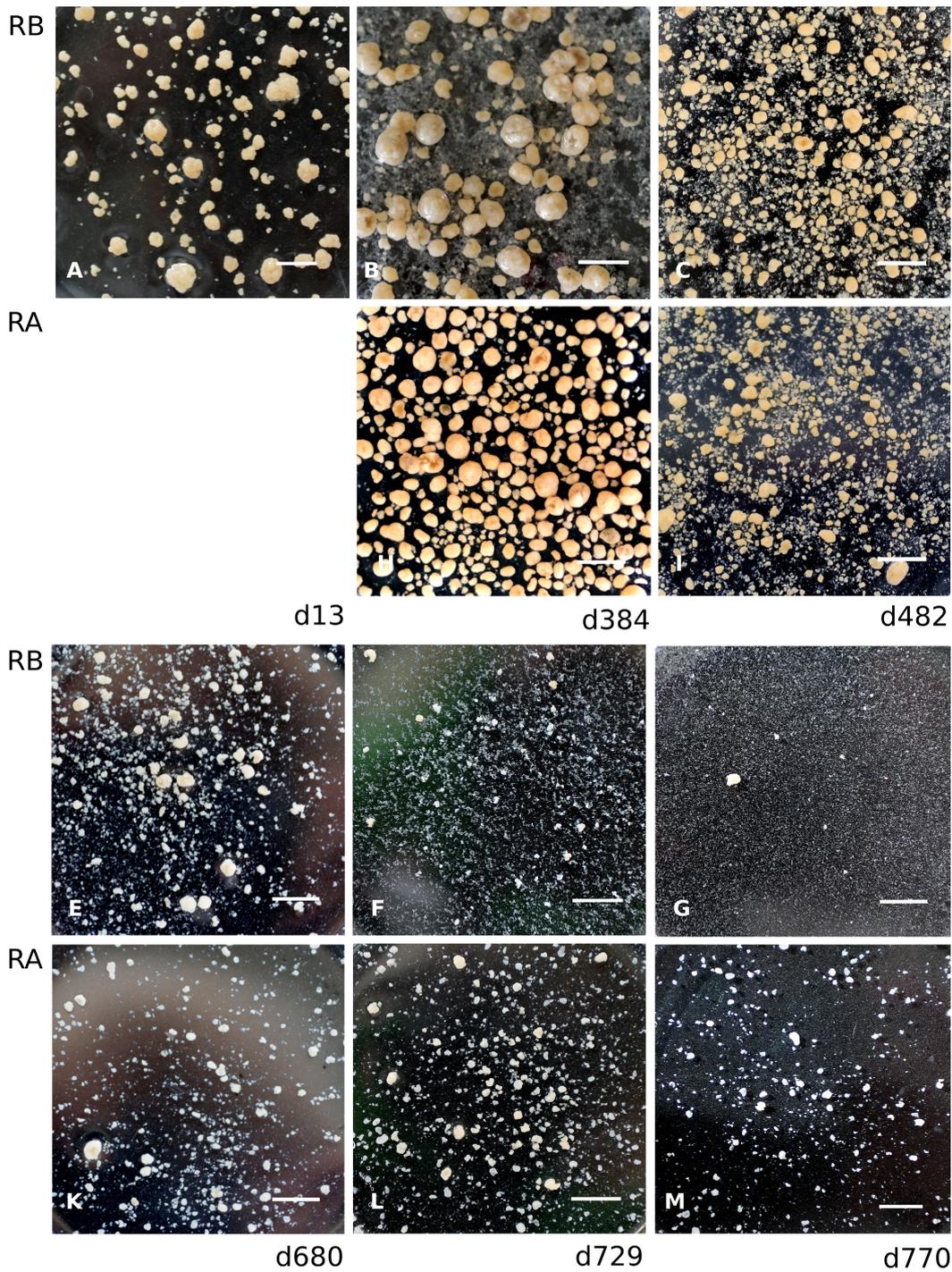


Figure IV.5 – Pictures of the biomass sampled during experiment 1 and experiment 2. Pictures of the biomass sampled in RB are on the first line (A-G) and the biomass sampled in RA is on the second line (H to M). The white bars represent 1 cm. The first column corresponds to biomass sampled on day 13 with medium 1 (simple wastewater). The three following columns corresponds to days 384, 482 and 606, when the reactors RA and RB were operated with medium 4 (complex monomeric wastewater). On the three last columns are the pictures of the biomass collected during the transition to polymeric wastewater (RB, days 680, 729 and 770) and the transition back to simple wastewater (RA, days 680, 729 and 770).

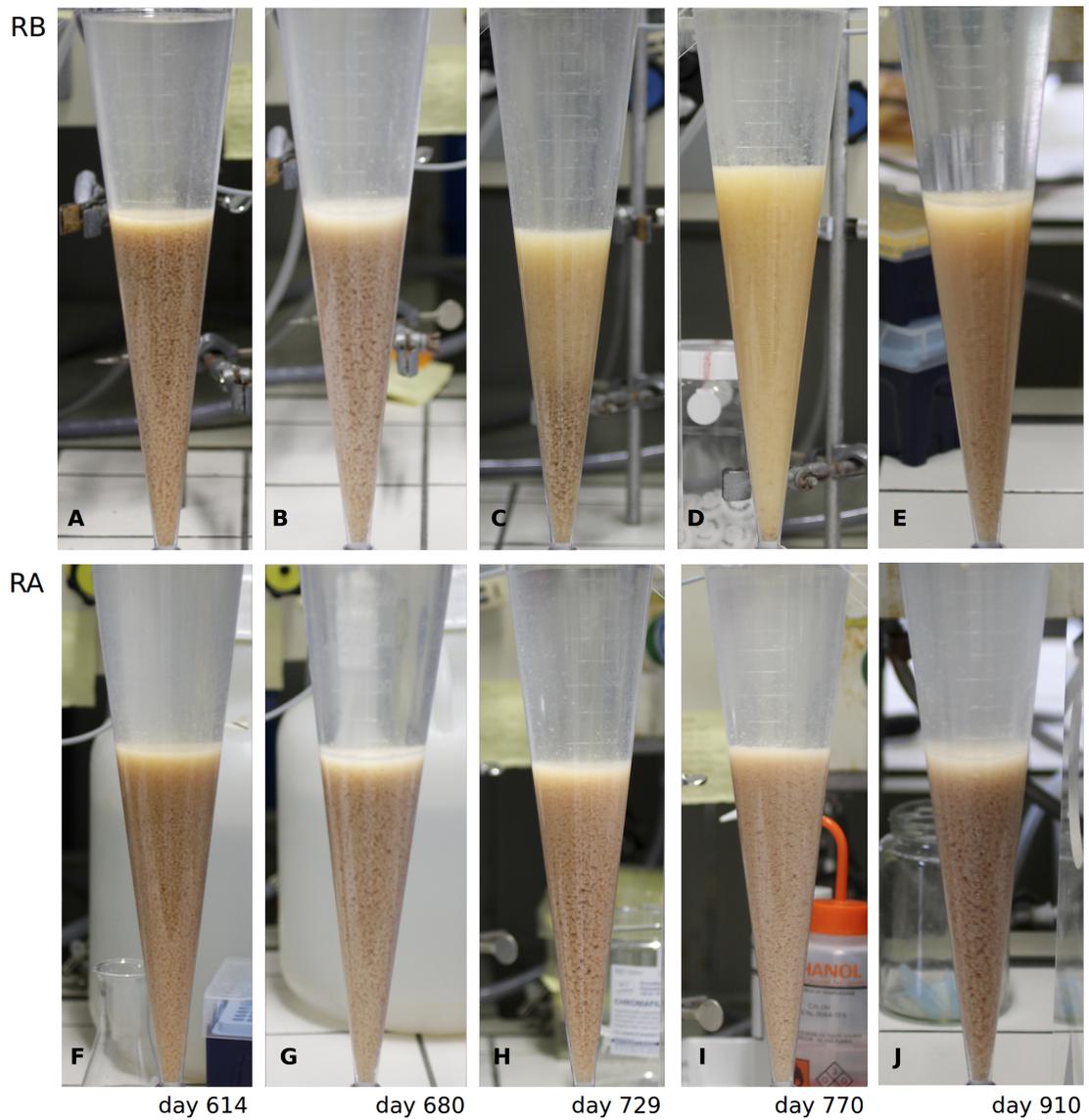


Figure IV.6 – Pictures taken after 10 minutes of settling, with the exception of the picture of day 770 which was taken after 3 minutes of settling, with biomass borrowed from reactor RB (A-E) and RA (F-J) during SVI measurements.

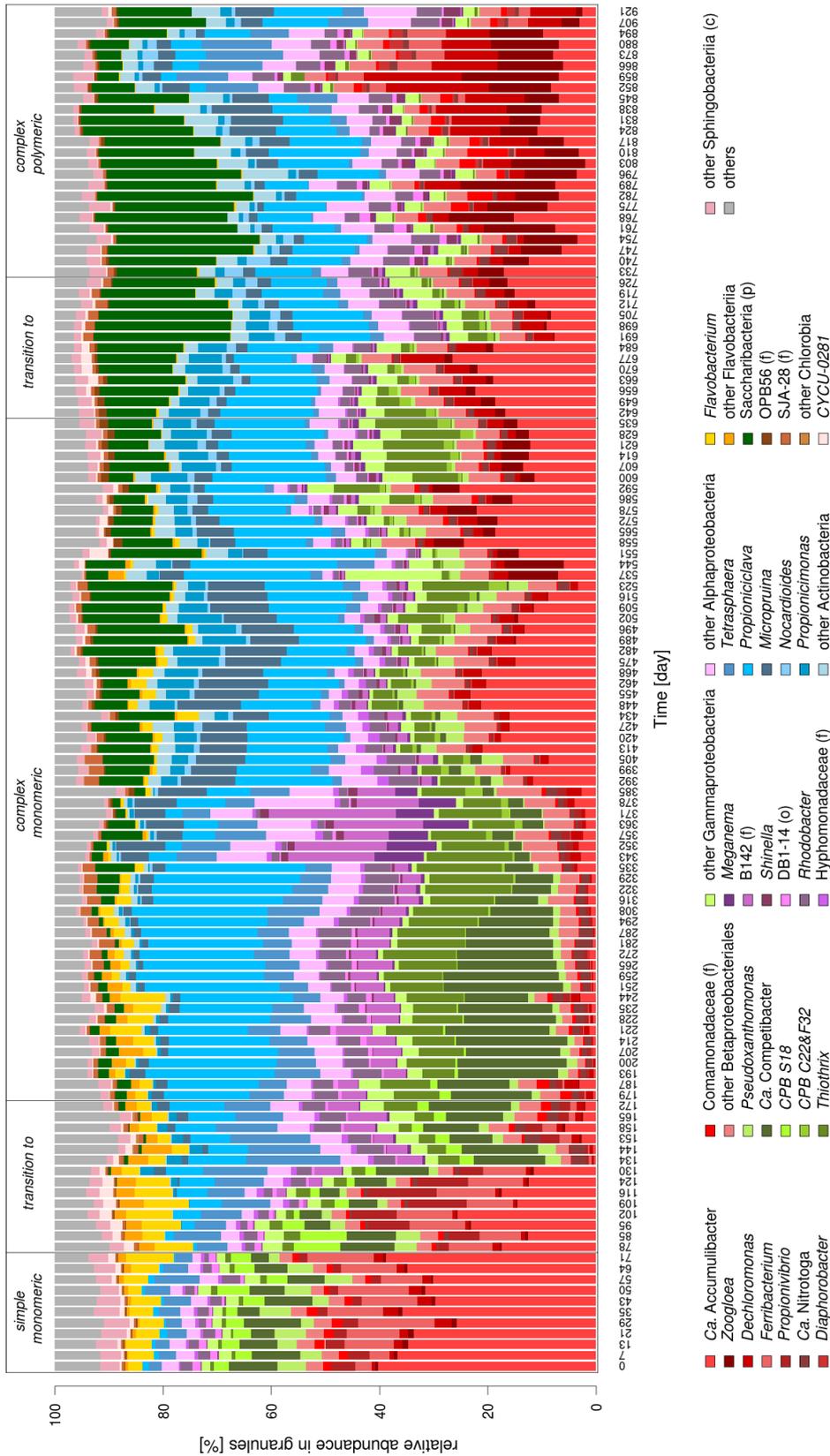


Figure IV.7 – Evolution of the bacterial community composition during the transition from simple to complex monomeric wastewater and to complex polymeric wastewater. The genera are colored according to the class they belong to, with the exception of the Betaproteobacteriales, previously the class of Betaproteobacteria, colored in red. They were recently merged with the Gammaproteobacteria [Parks et al., 2018], here colored in light green.

### IV.3.3 Evolution of the bacterial community composition during the transition from simple to complex monomeric and complex polymeric wastewater

The bacterial community composition, followed during 921 days of experiment 1 and 318 days of experiment 2 by 16S rRNA gene amplicon sequencing are presented in Figures IV.7 and IV.8, respectively. During the first transition from simple to complex monomeric wastewater, the relative abundance of Gammaproteobacteria decreased to the benefit of Actinobacteria. In a first time, the bacteria dominant in the AGS fed with the simple wastewater, *Ca. Accumulibacter*, decreased abruptly and *Tetrasphaera* and *Ca. Competibacter* became abundant. During the weeks following the duplication of the reactor on day 333, the amount of biomass per reactor was decreased by a factor of two and the food to microorganism ratio was higher during this period. The relative abundance of bacteria from the class of Alphaproteobacteria increased from 20 to 40 % for seven weeks. Afterwards, the relative abundance of *Ca. Accumulibacter* increased and remained around 20 %, and those of *Tetrasphaera* and *Ca. Competibacter* decreased again below 2 %. Uncharacterized bacteria from the phylum of Saccharibacteria appeared in the system during the transition to complex monomeric wastewater and progressively reached a relative abundance around 10%. After the second pH incident, on day 529, the relative abundance of *Zoogloea* increased up to 10% and remained around 5% during the following weeks.

The transition from complex monomeric to polymeric wastewater had a smaller effect on the bacterial community composition than the transition from simple to complex monomeric wastewater. The relative abundance of Saccharibacteria was around 20% during the transition and the three following months. In the class of Gammaproteobacteria, the proportion of *Ca. Accumulibacter* decreased and stabilized around 10 %, whereas those of *Dechloromonas* and *Zoogloea* increased. On day 848, corresponding to the formation of new granules, the proportions of *Dechloromonas*, *Zoogloea* and *Tetrasphaera* were around 10 % each. The proportion of *Thiothrix*, which represented around 10 % of the biomass with the complex monomeric wastewater, decreased to less than 1 % with the complex polymeric wastewater.

To assess the reversibility of the changes induced on the bacterial community by the transition from simple to complex monomeric wastewater, the reverse transition in the wastewater composition was performed in reactor RA. The second simple wastewater (medium 1B) had the same proportion of acetate and propionate as the first one (medium 1) (50 % each) but had a COD load of 600 mg O<sub>2</sub> l<sup>-1</sup>, compared to 450 mg O<sub>2</sub> l<sup>-1</sup> in the first one. After this transition from complex back to simple monomeric wastewater, the bacterial community was similar to the one composing the AGS at the beginning of experiment 1, but several differences were observed. Although Gammaproteobacteria had a similar abundance (60 %) as at the beginning of experiment 1 (70 %), the genera *Zoogloea*, *Dechloromonas* and *Thiothrix* were more abundant after the transition back to simple monomeric wastewater whereas *Ca. Competibacter* was less abundant. The Actinobacteria, whose abundance had increased during the change to complex wastewater, became even less abundant after the change back to simple wastewater than they were initially. Also the classes of Alphaproteobacteria, Chlorobia

## Chapter IV. Transforming acetate-propionate-grown granules into AGS treating fermentable and polymeric organic substrates

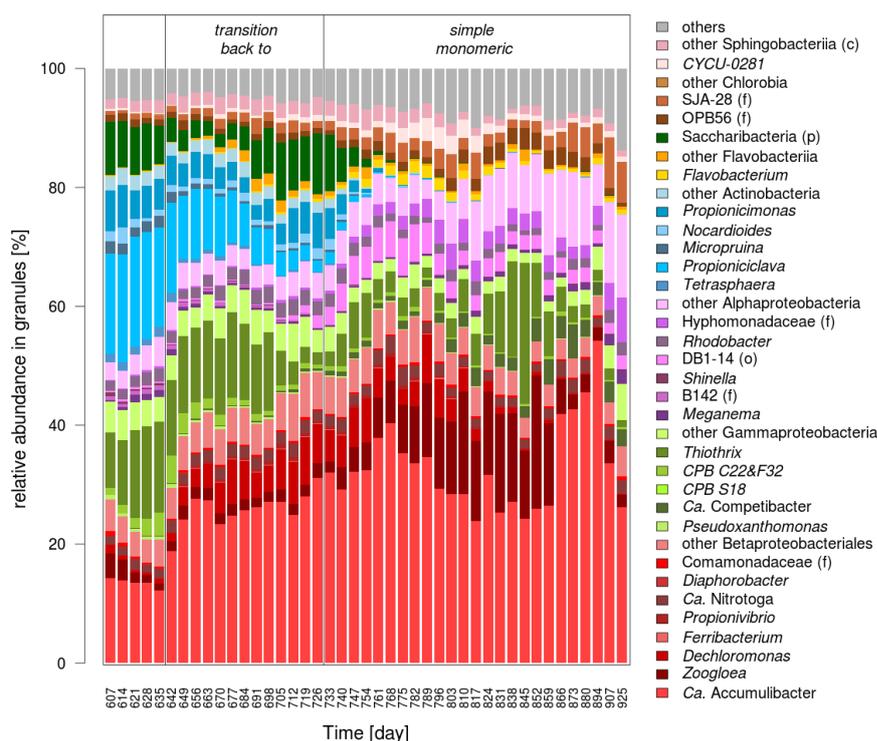


Figure IV.8 – Evolution of the bacterial communities during the transition back from complex monomeric to simple wastewater. The genus are colored according to the class they belong to, with the exception of the Betaproteobacteriales, colored in red, whereas the other Gammaproteobacteria are colored in light green.

and Spirochaetes were more abundant than they were at the beginning. After day 894, new changes were noticed in the bacterial community composition, with in particular, the decrease of the relative abundance of Zoogloea.

The two first coordinates of the Principal coordinate analysis (PCoA) based on the Bray-Curtis distance matrix of the bacterial communities, at genus level (abundant genera only), explain 63 % of the variance (Figure IV.9). During the first transition from simple to complex monomeric wastewater, the bacterial community changed rapidly and stabilized for some time before changing and stabilizing again. The second transition from complex monomeric to complex polymeric induced only small changes in the abundant genera of the microbial communities. On the other hand, during the transition back to simple wastewater, the bacterial community got closer to the one present at the beginning of the experiment, with the same type of wastewater. According to this plot, the AGS bacterial community is influenced by the presence of fermentable compounds in the wastewater, and to a lesser extent by the presence of polymeric compounds. This PCoA plot also highlights that other factors can have a strong effect on the bacterial community of AGS. Indeed, it was subject to important changes aside from any wastewater composition changes.

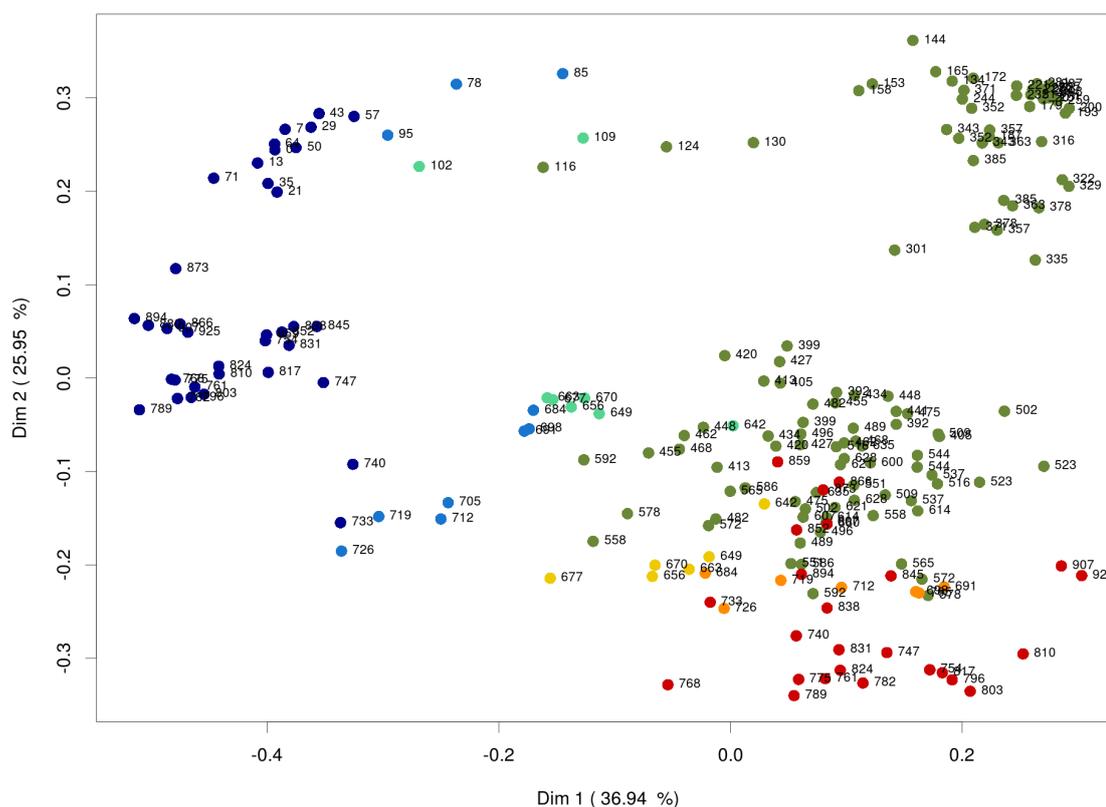


Figure IV.9 – Principal coordinate (PCoA) plot based on the Bray-Curtis distance matrix of the bacterial communities (at genus level). Only the abundant taxa were considered for this analysis. The two first coordinates Dim 1 and Dim2 explain 36.94 % and 25.95 % of the variance, respectively. The plots of the three first coordinates are provided in supplementary material (Figure A.5).

### IV.3.4 Stable states and discriminant taxa

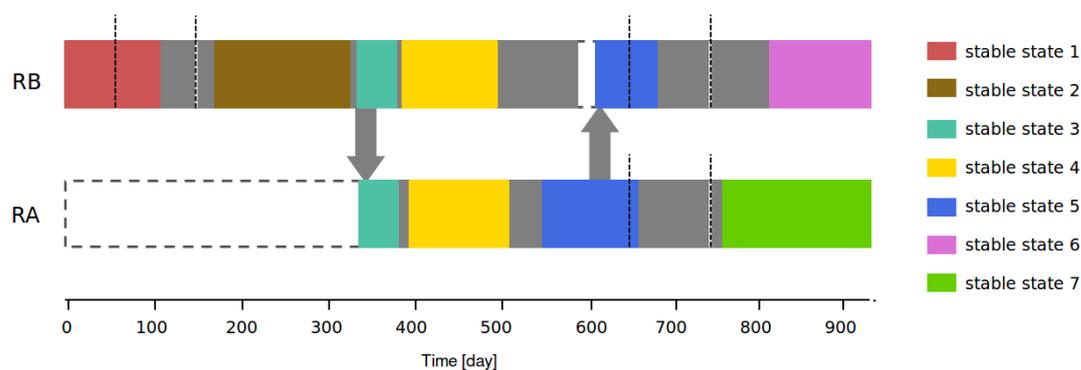


Figure IV.10 – The seven bacterial stable states detected during experiment 1 and 2.

The changes of wastewater composition, along with other factors induced changes in the bacterial community composition of the AGS, generally observed with a lag of some weeks. Seven different 'stable states' were detected in the bacterial communities in the AGS of experiment 1 and 2. The first stable state was observed with the simple monomeric wastewater and persisted during one month of transition. The second stable state begins a week after the transition to monomeric wastewater was complete and persisted until the inoculation of the second reactor (RA). This second state was characterized by a particularly low proportion of *Ca. Accumulibacter* and a high proportion of *Ca. Competibacter* and *Tetrasphaera*. The third stable state also corresponds to complex monomeric wastewater feeding and was detected in the two reactors in parallel. It began after the splitting of the biomass in two reactors and lasted 6 weeks, until the usual amount of biomass was recovered.

Shortly afterwards, a fourth stable state was observed in parallel in the two reactors. The end of this stable state correspond to a period of time when a lot of biomass was sampled from the reactors for different experiments, followed by operation failures such as a pH regulation incidents (RA) or the lack of trace elements in the influent (RA and mainly RB). The fifth stable state started several weeks after the stabilization of the reactor operation. It persisted in RB after inoculation and few weeks during the beginning of the transition to different wastewater types. The sixth stable state was observed three months after the transition to complex polymeric wastewater and the seventh one month after the end of the transition from complex back to simple monomeric wastewater.

In order to determine which taxa were the most impacted by the changes of substrate, t-test analysis were performed between the mean relative abundance of the bacterial communities of the three influent wastewater types (supplementary file Table A.15). There were 29 abundant, 186 divergent taxa (at genus level) on a total of 292.

Among the 29 abundant taxa, *Dechloromonas*, *Ferribacterium*, *Ca. Nitrotoga*, and uncharacterized taxa from the families of Comamonadaceae, OPB56 and SJA-28 were not divergent. The 23 other divergent and therefore discriminant taxa are presented in Table IV.2. *Propionivibrio*,

Table IV.2 – Table of the discriminant taxa of the study. The sign '-' indicates that the taxon was significantly less abundant in the second wastewater type compared to the first one. Conversely, the sign '+' indicates that the taxon was significantly more abundant in the second wastewater type compared to the first one. The average abundances of the taxa according to the three wastewater types are in electronic supplementary material Table A.16

Genus	simple to complex monomeric	simple to complex polymeric	complex monomeric to polymeric
<i>Ca. Accumulibacter</i>	-		
<i>Rhodobacter</i>	+		
<i>Tetrasphaera</i>	+		
DB1-14 (o)	-	-	
Hyphomonadaceae (f)	-	-	
<i>CYCU-0281</i>	-	-	
<i>Nocardioides</i>	+	+	
<i>Propionicalva</i>	+	+	
<i>Micropruina</i>	+	+	
<i>Propionicimonas</i>	+	+	
Saccharibacteria (p)	+	+	
<i>Shinella</i>	+	+	
<i>Diaphorobacter</i>		+	
<i>Propionivibrio</i>		-	-
<i>Ca. Competibacter</i>		-	-
<i>Flavobacterium</i>		-	-
CPB S18			-
<i>Meganema</i>			-
<i>Pseudoxanthomonas</i>			-
B142 (f)	+		-
<i>Thiothrix</i>	+		-
<i>Zoogloea</i>	-		+
CPB C22&F32	+	-	-

#### Chapter IV. Transforming acetate-propionate-grown granules into AGS treating fermentable and polymeric organic substrates

Table IV.3 – Variations in the relative abundance of the discriminant taxa of AGS across the different stable states observed with complex monomeric wastewater (stable states 2, 3, 4 and 5). Only the taxa with significant and non ambiguous differences of relative abundance from a stable state to all the others are shown. These differences are indicated by different signs, '-' being the lowest relative abundance and '++' the highest. Other functionally important taxa are not included in this table if they did not show differences between the stable states. The p-values of the t-tests are in electronic supplementary material Table A.15

function	Taxon	stablestate 2	stablestate 3	stablestate 4	stablestate 5
PAO	<i>Ca. Accumulibacter</i>	-	+	++	++
	<i>Tetrasphaera</i>	+	+	-	-
GAO	<i>Ca. Competibacter</i>	++	+	-	--
	<i>CPB_S18</i>	+	+	-	--
	<i>Propionivibrio</i>	+	+	+	-
	<i>CPB C22&amp;F32</i>	-	+	+	+
putative denitrifier	<i>Rhodobacter</i>	++	--	+	-
	<i>Meganema</i>	-	+	-	-
	<i>Zoogloea</i>	-	-	+	++
	<i>Pseudoxanthomonas</i>	+	--	++	-
fermenter	<i>Propionicalva</i>	++	-	+	+
	<i>Nocardioides</i>	-	+	++	++
putative hydrolyzer	<i>CYCU-0281</i>	+	-	+	++
	<i>Flavobacterium</i>	+	-	-	-
not determined	DB1-14 (o)	+	-	-	-
	<i>Shinella</i>	-	+	-	-
	<i>Diaphorobacter</i>	-	+	+	+
	B142 (f)	+	++	-	-

*Ca. Competibacter* and *Flavobacterium* were enriched in the monomeric wastewaters. *Ca. Accumulibacter* was significantly less present in the AGS fed with the complex monomeric wastewater than in the AGS fed with the simple wastewater, but the difference is not significant between the simple and the polymeric wastewater. During the transition to complex monomeric wastewater, the proportion of *Ca. Accumulibacter* decreased abruptly.

Bacteria from the order DB1-14, the family Hyphomonadaceae and the genus *CYCU-0281* were significantly enriched in the simple wastewater compared to the complex wastewater. The Actinobacteria *Nocardioides*, *Propioniciclava*, *Micropruina* and *Propionicimonas*, able to grow by fermentation, were significantly enriched in the complex wastewaters. This was also the case for *Shinella* and the uncharacterized Saccharibacteria (p). *Flavobacterium*, *CPB\_S18*, *Meganema* and *Pseudoxanthomonas* decreased with the introduction of polymeric compounds in the wastewater.

The mean relative abundances of the abundant taxa in the different stable states observed with complex monomeric wastewater were compared in order to identify important changes inside the main functional groups. Globally, the decrease in the proportion of one or more members of a functional group is concomitant with the increase of one or more other members of this

group. For example, in stable state 2, the PAO *Tetrasphaera* was around 6 % relative abundance and the PAO *Ca. Accumulibacter* was around 1 %. In the following stable states, the proportion of *Ca. Accumulibacter* increased while the proportion of *Tetrasphaera* decreased. The proportion of the GAO *Ca. Competibacter* followed the opposite trend from *Ca. Accumulibacter*, its relative abundance decreased between each stable state 2 to 5. The GAO *CPB\_S18* and *Propionivibrio* followed the same trend, with a lower intensity, whereas the *CPB\_C22&F32* increases between stable state 2 and 3.

#### IV.3.5 Changes in the microbial communities during the transition to a complex polymeric and back to simple wastewater

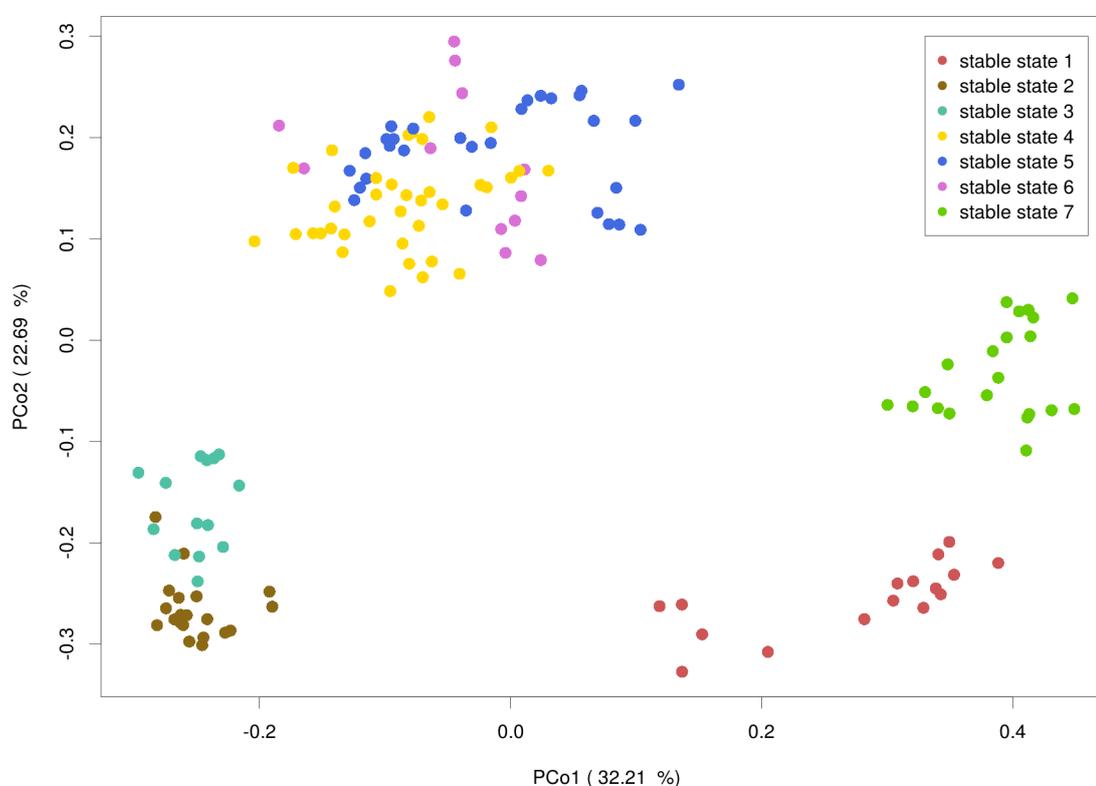


Figure IV.11 – Principal component analysis based on the Bray-Curtis distance of the bacterial communities (genus level) at stable states. The two first coordinates PCo1 and PCo2 explain 32.21 % and 22.69 % of the variance, respectively. The plots of the three first coordinates are provided in supplementary material (Figure A.6).

The main similarities between the bacterial communities are highlighted in the PCoA based on their Bray-Curtis distance matrix of the samples at stable state (Figure IV.11). On the two first axis of this PCoA, the stable states are grouped in three big clusters : stable states 1 and 7 (simple monomeric), stable states 2 and 3 (complex monomeric), and stable states 4, 5 (complex monomeric) and 6 (complex polymeric).

## Chapter IV. Transforming acetate-propionate-grown granules into AGS treating fermentable and polymeric organic substrates

---

The bacterial communities from stable states 1 and 7, corresponding to the simple wastewater before and after the passage to complex monomeric wastewater, are close to each other, but they are not overlaid. This confirms that the changes in the bacterial communities induced by this change of wastewater composition were only partly reverted. This can also imply that the bacterial communities respond to factors other than the composition of the wastewater alone. Stable states 2 and 3 are clearly separated from the rest of the samples on the third axis, although they were fed with the same type of wastewater as the samples from stable states 4 and 5. The difference between the bacterial communities of these two stable states may be primarily due to the low abundance of *Ca. Accumulibacter* characterizing these bacterial communities. The stable states 4, 5 and 6 are grouped together in this analysis. The differences between the samples collected before and after the transition from complex monomeric to complex polymeric wastewater do not show on the two first axis of this PCoA, revealing that the changes of bacterial communities during this transition were small compared to the changes induced by other wastewater composition changes or perturbations.

### IV.3.6 Comparison of the bacterial communities in flocs and granules

The proportions of the abundant taxa in the flocs and granules were compared in order to determine potential enrichment in one fraction or the other (Figure ( IV.12)). For the majority of the taxa, no strong tendencies are detected. Stable states 2 and 4 gather respectively 14 and 19 observations, whereas stable states 5 and 6 gather respectively 8 and 4 observations. The low amount of data can be the cause for the absence of significant results for these two latter states.

In the stable states 2, 4 and 5, *Tetrasphaera*, the GAO *Ca. Competibacter*, *Rhodobacter*, SJA-28 (f), *Nocardioides*, and B142 (f) were mostly enriched in granules, whereas *Thiothrix*, *Zoogloea*, *Propionicimonas* and *Dechloromonas* were slightly enriched in flocs.

The tendencies in the stable state 6 do not correspond with those of the stable state 2, 4 and 5. During this stable state, the formation of new granules was not observed, therefore, the bacteria enriched in the granules during this state probably correspond to the old community which was maybe not active but retained in the system because of the high sludge retention time of the granules. The active communities were, at that time, located in the flocs. Indeed, *Tetrasphaera*, which was detected to be enriched in granules during the other stable states, was mostly in the flocs in the samples from the stable state 6. This indicates that the relative abundance of *Tetrasphaera* was higher in the recently formed biomass.

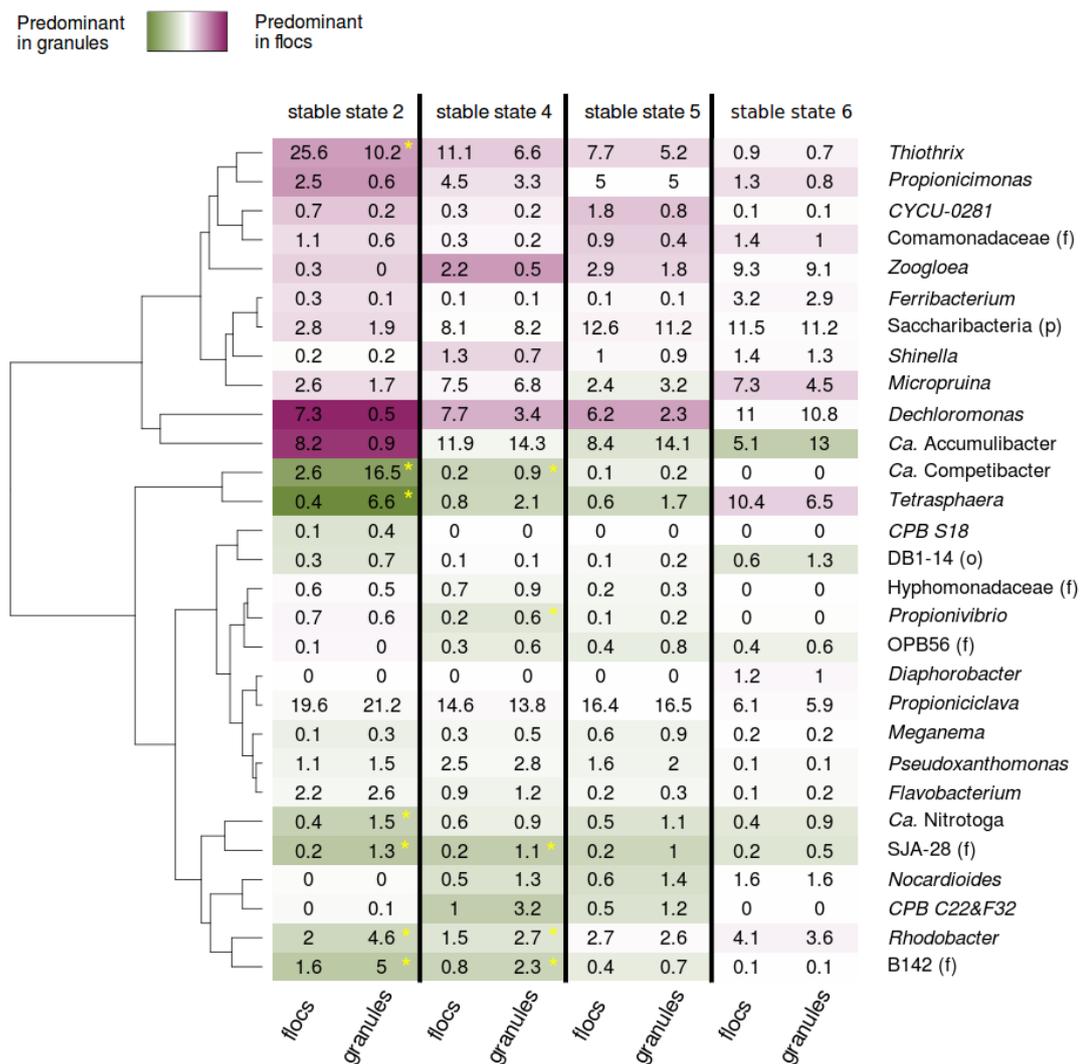


Figure IV.12 – Granules and flocs heatmap showing the average relative abundance of the abundant taxa during stable states 2, 4, 5 and 6. The flocs and granules data of the other stable states were too sparse for this analysis. Green indicates a higher proportion in granules, purple a higher proportion in flocs. In order to lower the possible effect of noise in the very low abundant taxa, a pseudo-count of 0.5 % was added to each abundance for the color code and the significance analysis. A yellow asterisk indicates the differences that are significant with a adjusted p-value of  $3.45 \times 10^{-4}$  (0.01/29).

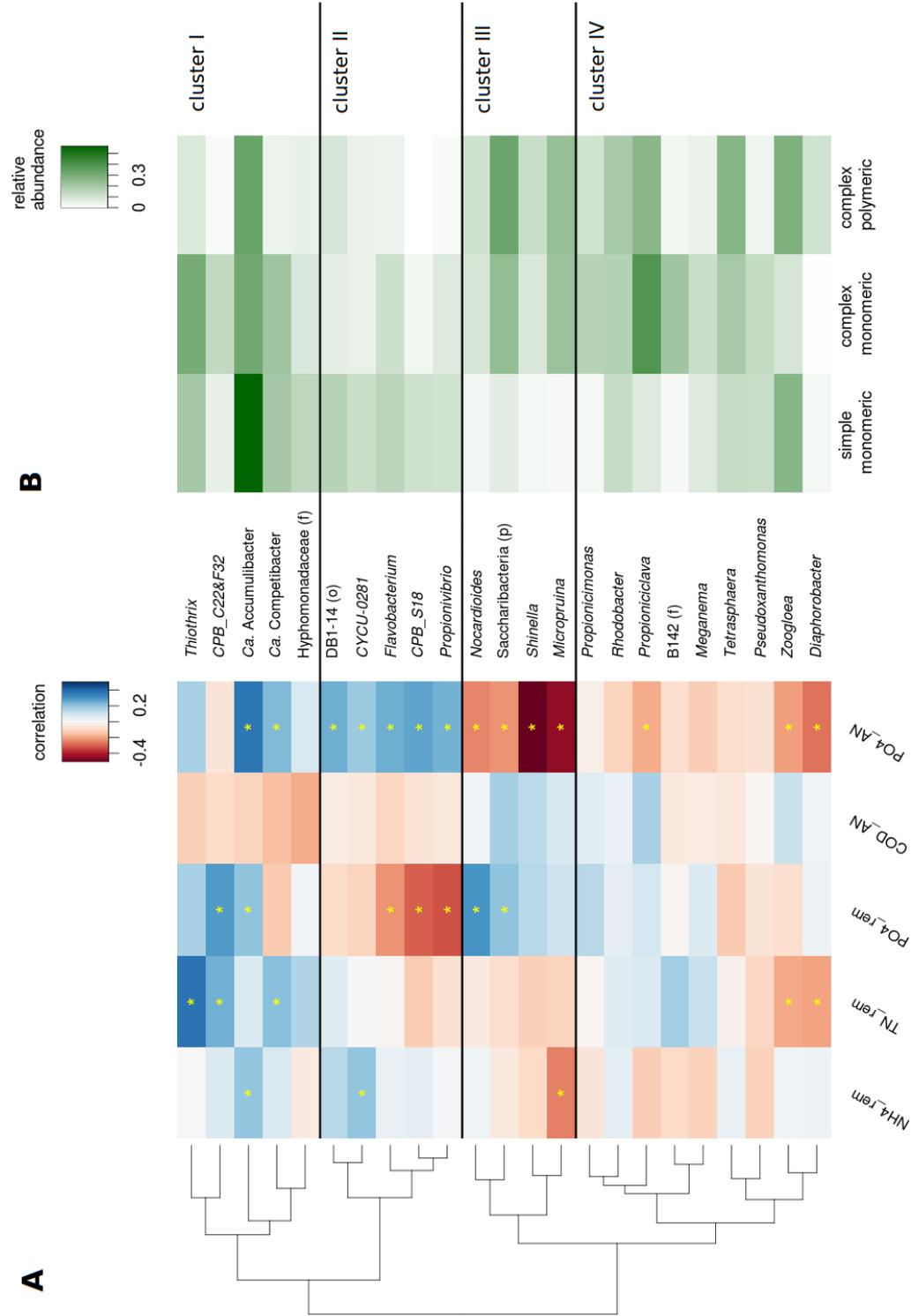


Figure IV.13 – Correlation heatmap between the discriminant taxa and the nutrient removal during experiment 1 and 2 (A) and average Hellingert-transformed relative abundance of these taxa in the stable states corresponding to the different wastewater types(B).

### IV.3.7 Correlation between the bacterial communities and the nutrient removal performances

Correlations between the nutrient removal performance and the discriminant taxa are shown in Figure IV.13(A). Few of these correlations are strong or significant. Probably because the main reasons for nutrient-removal performances decrease were mostly operational dysfunctions. The variable showing the most significant correlations with the relative abundance of the discriminant taxa is the level of phosphate in the bulk water at the end of the anaerobic phase. According to this heatmap, the discriminant taxa were grouped in four clusters. Cluster I is associated to a good TN-removal, medium to good P-removal, low COD and medium to high phosphate concentrations in the bulk water at the end of the anaerobic phase. It is composed of the classical PAO *Ca. Accumulibacter*, two classical GAO *Ca. Competibacter* and *CPB\_C22&F32*, uncharacterized genera from the Hyphomonadaceae and the aerobic bacteria *Thiothrix*. As expected, *Ca. Accumulibacter* is strongly correlated with high P-release during the anaerobic phase and high P-removal. Its most famous competitor, *Ca. Competibacter*, is weakly and not significantly correlated with lower P-removal. An explanation might be that it took advantage of the sudden decrease of *Ca. Accumulibacter* during the transition to complex monomeric wastewater. At that time, the P-removal performances were impacted for several weeks. This cluster is principally composed of bacteria that were more abundant with the monomeric wastewater (Figure IV.13(B)). This can be an explanation for their positive correlations with TN-removal. Indeed, TN-removal was slightly lower with the polymeric wastewater. Another explanation for the positive correlation with good TN-removal can also be that some of them have denitrifying capabilities.

Cluster II is linked with lower P-removal, low concentrations of COD and high concentrations of phosphate at the end of the anaerobic phase. It is composed of bacteria that were most abundant with the simple monomeric wastewater. Therefore their positive correlation with high P-release can be linked with their co-occurrence with *Ca. Accumulibacter*, which was particularly abundant with this wastewater type. On the other hand they are negatively correlated with the P-removal performance, in particular the GAO *CPB\_S18* and *Propionivibrio*.

Cluster III is linked with good P-removal performance, higher concentrations of COD and lower concentrations of phosphate at the end of the anaerobic phase. It is composed of bacteria that were more abundant with the complex wastewater, and especially with the polymeric one. The reasons for these correlations remain unclear but can be related to the wastewater type.

Cluster IV is weakly linked with low amounts of phosphate at the end of the anaerobic phase. *Zoogloea* and *Diaphorobacter* are correlated with low TN removal, possibly because of their higher proportions with the polymeric wastewater.

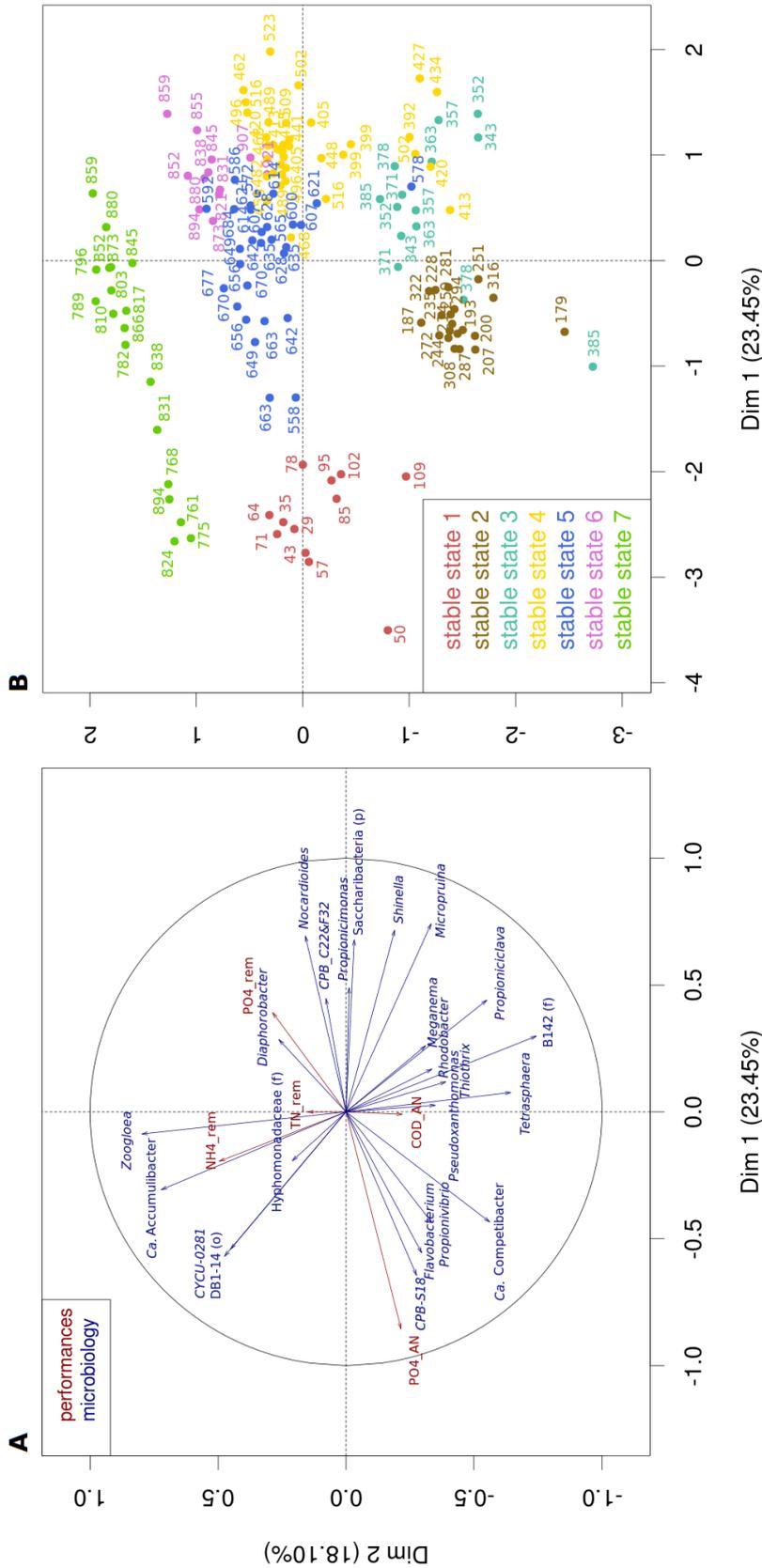


Figure IV.14 – Multiple factor analysis (MFA) performed on the nutrient-removal performance and the bacterial communities of the samples considered at bacterial stable state. The contribution to the composition of the two first axis is shown for the nutrient-removal data and the 23 discriminant taxa (A). PO4\_AN and COD\_AN are the phosphate and the COD concentrations in the bulk water of the reactor measured at the end of the anaerobic phase. NH4\_rem is the ammonium removal efficiency, TN is the total nitrogen removal efficiency and PO4\_rem is the phosphate removal efficiency. The projection of the samples points on the two first axis of the MFA is shown in (B). Stable states 1 and 7 correspond to simple monomeric wastewater, stable states 2 to 5 correspond to complex monomeric wastewater (for 5, a little bit of the transition to simple or polymeric is comprised) and stable state 6 corresponds to complex polymeric wastewater.

### IV.3.8 Multiple factor analysis

The multiple factor analysis (MFA) performed on the nutrient removal performance and the bacterial communities of the samples at bacterial stable state (Figure IV.14) summarizes the main tendencies of these two characteristics of AGS during this experiment. The comparison between the projection of the axes composition and the positions of the samples in the two dimensions highlights correlations between the discriminant taxa and the different stable states, themselves linked to different wastewater types. This figure reflects that *Ca. Accumulibacter* and *CYCU-0281* were particularly abundant with the simple monomeric wastewater. *Diaphorobacter* was the only discriminant taxa to be present in higher proportion with the polymeric wastewater (stable state 6). The bacteria present in higher proportions with the complex monomeric wastewater (stable states 2 to 5) are located around the bottom right corner of the axis projection. They are primarily belonging to Actinobacteria : *Tetrasphaera*, *Propionicyclava*, *Micropruina*, *Propionicimonas*, and other bacteria for which the *in situ* metabolism was not studied in wastewater treatment sludge. The samples from stable state 2 and 3 are separated from the ones from stable states 4, 5 and 6, mostly due to differences in the microbial communities. Most of the samples from stable state 7 are clearly separated from those of stable state 1, while they correspond to the same wastewater type. Their microbial communities are a little different and the wastewater 7 had a higher COD load.

## IV.4 Discussion

### IV.4.1 General observations

The composition of the wastewater was changed from simple to complex monomeric in a first time and to complex polymeric in a second time. The aims were to assess how the mature AGS bacterial communities, developed on a mixture of acetate and propionate, would adapt to the presence of fermentable and polymeric compounds in the influent wastewater, and if the settling and nutrient removal performance of these AGS would be impacted by such changes. Despite the fact that the bacterial communities changed substantially with the introduction of fermentable compounds in the influent, the settling properties and the nutrient removal performances remained globally unchanged. Conversely, the introduction of polymeric compounds in the influent caused smaller changes in the bacterial communities, but significantly altered the settling properties and TN removal performances.

### IV.4.2 Transition from simple to complex monomeric wastewater

The class of Actinobacteria with *Tetrasphaera*, *Propionicyclava*, *Propionicimonas* and *Micropruina*, and the phylum of Saccharibacteria became abundant with the transition from simple to complex monomeric wastewater, while the class of Gammaproteobacteria and in particular *Ca. Accumulibacter* diminished from around 35 % to less than 5 %. The Actinobacteria genera detected with the complex wastewater were previously detected in high abundance in EBPR

#### Chapter IV. Transforming acetate-propionate-grown granules into AGS treating fermentable and polymeric organic substrates

---

activated sludge [Nielsen et al., 2012b, Stokholm-Bjerregaard et al., 2017, Xia et al., 2008a] treating real wastewater. Their presence in AGS treating complex wastewater is therefore not surprising. Yet the influence on the granulation of the sludge and the settling properties was unknown. The results of this study show that a high proportion of Actinobacteria and Saccharibacteria is not detrimental to the granulation and the settling properties of the AGS, since the  $SVI_3$  of the AGS stayed below  $50 \text{ ml g}^{-1}$  with the complex monomeric wastewater.

The members of Saccharibacteria, formerly named Candidate phylum TM7, detected in this study, are only characterized at the phylum level. Their metabolism is therefore unknown. Yet, the metagenome assembled genomes of other members of this phylum (e.g., *Ca. Saccharimonas*) have recently been obtained from EBPR sludge samples [Albertsen et al., 2013]. The reconstructed genome of *Ca. Saccharimonas* suggests an obligate fermentative metabolism, taking up glucose and releasing acetate and lactate, with the ability to survive in aerobic conditions through oxidative stress protective enzymes such as superoxide dismutase and glutathione peroxidase [Albertsen et al., 2013]. The fact that the uncharacterized members of Saccharibacteria were abundant in the system only with complex wastewater, suggests that they may also have the ability to ferment glucose or amino acids.

The dramatic decrease of the relative abundance of *Ca. Accumulibacter* is difficult to explain. First, there doesn't seem to be a limitation in availability of VFA. VFA were still present in the influent although at lower concentrations and the fermentative bacteria produce VFA upon glucose and amino acid fermentation. The fermentative PAO *Tetrasphaera* can ferment various sugars and amino acids anaerobically and release succinate, acetate and other fermentation products [Barnard et al., 2017, Kristiansen et al., 2013]. The fermentative GAO *Micropruina* also releases part of the fermented carbon mainly as acetate and lactic acid [McIlroy et al., 2018]. *Propioniciclava* and *Propionicimonas* are known to ferment glucose and other carbon source and release propionic acid [Bae et al., 2006, Zhang et al., 2017b]. Hence, the fermentation products released by these fermentative bacteria should allow classical PAO and GAO to remain in the system. Second, the 'disappearance' of *Ca. Accumulibacter* from the AGS was temporary and this PAO was around 20 % 8 months later.

The most likely hypothesis for the sudden decrease of *Ca. Accumulibacter* is the infection by a bacteriophage. Indeed it is common that the dominant bacterial population of a microbial community is attacked by a specific phage [Rodriguez-Brito et al., 2010]. More specifically, sudden decreases of *Ca. Accumulibacter*, detected in EBPR activated sludge by using quantitative FISH, were attributed to a lytic *Ca. Accumulibacter*-associated bacteriophage [Barr et al., 2010b]. As in the present study, this event was associated with an increase of the proportion of *Ca. Competibacter* and a deterioration of the P-removal efficiency. The P-removal performance was impaired from day 103 on. Significant proportions DNA sequences corresponding to *Ca. Accumulibacter* were detected until day 130. Therefore, either the phage attack occurred around day 102 and the DNA of the lysed *Ca. Accumulibacter* cells remained in the granules for 20 days, or the decrease of P-removal performances were in this case not related to a phage attack but rather to the decrease of VFA and total COD content in the influent wastewater.

Whether it was linked or not to the decrease of P-removal efficiency, the phage attack is supported by previous analysis performed on the AGS present in the reactor before the beginning of the experiment (chapter II). Indeed the genomes of the phages EBPR podovirus 1 (EPV1) was also detected in the DNA sequences of the granules sequenced separately [Leventhal et al., 2018]. This phages was previously detected in the metagenomes from an EBPR reactors and both was proposed to be able to infect *Ca. Accumulibacter* [Skennerton et al., 2011].

### IV.4.3 Transition from complex monomeric to polymeric wastewater

The changes in the bacterial communities after the transition from complex monomeric to polymeric wastewater were more subtle than the bacterial community changes during the first transition from simple to complex monomeric influent. The samples of the stable state 6 corresponding to polymeric wastewater clustered with stable states 4 and 5, corresponding to complex monomeric wastewater, on the first two components of the PCoA of the bacterial communities during stable states. Likely some bacteria able to ferment sugars and amino acids also have the capability to hydrolyze polymeric compounds [Xia et al., 2008a]. One of the major changes observe during this transition was the increase of *Zoogloea*. In AGS, this heterotrophic EPS producer [Larsen et al., 2008] is likely a marker of VFA leakage in the aerobic phase [Weissbrodt et al., 2012]. It had already began to thrive after the second pH incident. At that time, the biomass was lower and even if the volume exchange ratio was adapted to the amount of biomass, the COD concentrations at the end of the anaerobic phase were slightly higher than the usual ones (data not shown).

The transition from complex monomeric to polymeric wastewater was followed by a significant decrease of the settling properties. This link between polymeric compounds and lower settling properties of AGS was reported in previous studies [de Kreuk et al., 2010, Martins et al., 2011]. With AGS, high amounts of COD at the beginning of the aerobic phase are believed to be responsible for the development of floccular or less dense structures due to the overgrowth of aerobic heterotrophs [Pronk et al., 2015a]. Yet, hydrolysis being a slow process and the rate limiting step for the COD uptake by the bacterial communities [Morgenroth et al., 2002], it is likely that part of the COD is not hydrolyzed during anaerobic phases shorter than two hours. After the addition of polymeric compounds in the influent, the COD at the end of the anaerobic phase was higher than it was with the monomeric wastewaters. Moreover it is possible that part of the polymers were not detected in the bulk water because they were adsorbed at the surface of the granules [de Kreuk et al., 2010]. In this case, our analysis of soluble COD could have under-estimated the amount of COD available during the aerobic phase. It is therefore likely that an increased amount of COD available at the beginning of the aerobic phase was an important factor of the deterioration of the settling properties of the sludge observed with polymeric wastewater. Another explanation for the presence of flocs and small granules with the polymeric substrate is that big polymers cannot diffuse into the granules as easily as monomers. Even though the polymeric compounds in the wastewater of this study had a rather small size, it may not diffuse into the granules as quickly as monomeric

## **Chapter IV. Transforming acetate-propionate-grown granules into AGS treating fermentable and polymeric organic substrates**

---

compounds, thus giving a competitive advantage to flocs and small or porous granules.

The transition to polymeric wastewater was first followed by a progressive replacement of the granules by floccular biomass. A first apparition of new granules, announced by a contraction of the sludge, was detected two months after the end of the transition to polymeric wastewater (day 788). Yet, no significant change in the bacterial community was noted around that day, therefore one can only speculate as to the reasons for the temporary loss of granulation. First, the fact that no changes were detected at the genus level in the bacterial population does not guaranty that the community did not change. Bacteria from a same taxon can have different metabolic capabilities. It is therefore possible that bacteria able to produce hydrolyzing enzymes were not present in sufficient abundance in the sludge before the transition to polymeric wastewater and took some time to thrive. It would also be possible that this capability to hydrolyze starch and peptone became abundant in the microbial population by horizontal gene transfer [Soucy et al., 2015]. A second potential reason for the impaired granulation resides in biofilm regulation possibly via quorum sensing [Tan et al., 2014]. The eukaryotic population of AGS can play a role in the degradation of polymeric COD from the wastewater. Unlike prokaryotes, eukaryotes are able to uptake macro particles before their hydrolysis. A part of the polymeric COD was possibly utilized by the eukaryotes present in the sludge. A lower availability of COD content of the wastewater by the bacteria implied in biofilm formation, can inducing the bacterial communities to temporarily invest less in biofilm formation [Burne, 1998, Jefferson, 2004, Jefferson et al., 2004].

Even after granulation, the settling properties of the sludge were not as good as they were with the monomeric wastewaters, in particular because of a part of the biomass remained floccular. This effect of polymeric compounds on the settling properties were previously observed with AGS treating low to medium-strength wastewater containing an important proportion of polymers (typically municipal wastewaters) [Derlon et al., 2016, Pronk et al., 2015b, Wagner et al., 2015b] including the study presented in chapter III. A slight decrease in the TN-removal, due to higher nitrate concentrations in the effluent, was noticed with the polymeric wastewater. This negative correlation between the proportion of polymeric COD in the influent and the nitrate concentration in the effluent was previously noticed [Wagner et al., 2015b, Wang et al., 2018] and will be discussed further in chapter VI.

### **IV.4.4 Reversibility of the changes induced in the bacterial communities**

The changes in the bacterial community noticed after the change from simple to complex monomeric wastewater were partly reversed after the transition back to simple wastewater (Figure IV.11). The total proportion of bacteria from the class of Actinobacteria and the phylum of Saccharibacteria, that likely have a fermentative metabolism and benefited from the introduction of glucose and amino-acids in the influent, decreased again below 0.5 %. It is not surprising that fermentative bacteria decrease drastically if the fermentable compounds are not provided anymore. Yet, *Tetrasphaera* was among the abundant bacteria during the

initial phase with simple monomeric wastewater. Members of this genus have the capability to take up acetate under anaerobic conditions (without P-cycling) [Kristiansen et al., 2013], allowing them to maintain themselves in AGS fed with simple wastewater. Yet this did not allow *Tetrasphaera* to stay abundant in the AGS after the transition back to simple monomeric wastewater.

The class of Gammaproteobacteria almost returned to the proportions that were usually measured before the change to complex monomeric wastewater (around 70 %). Conversely, the classes of Alphaproteobacteria and Chlorobia were measured at higher proportions than they were at the beginning of this study.

The difference in the bacterial communities of AGS fed with the simple wastewater before and after the changes of wastewater composition can be explained by the existence of multiple equilibria or also by the difference of the total COD concentration between the two simple wastewaters.

### IV.4.5 Multiple stable states as the result of multistability ?

The AGS reactors were operated during 15 and 10 months with complex monomeric wastewater. During that time, four different stable states were observed. Perturbations such as too high pH, change in food over microorganism ratio were detected as potential trigger factors for the switch to a new equilibrium in the bacterial community. It is difficult to state if the transient stable states observed were a case of multiple stable states, since each one of them seemed to result of a perturbation. The monitoring of AGS bacterial community composition and functioning fed with the same wastewater composition and without perturbations over a long period of time would have been necessary to assess the existence of multiple stable states.

These four stable states were performing similarly in terms of settling and nutrient removal efficiency, even though some differences were noted regarding the amount of P-release during the anaerobic phase. Indeed, most of the losses in nutrient removal corresponded to operation failures, and no strong correlation could be establish between particular taxa and nutrient removal performance. Similarly, microbial communities from conventional or EBPR activated sludge or methanogenic bioreactors were also reported as extremely dynamic in terms of composition without that the performance were affected [Fernandez et al., 2000, Wang et al., 2010, Wells et al., 2011] .

This functional stability noticed in the present study was likely possible through a sufficient functional redundancy present in the bacterial community before each perturbation [Botton et al., 2006]. At the beginning of the experiment, *Ca. Accumulibacter* was the dominant PAO, but *Tetrasphaera* was present in the system, probably through its metabolic versatility allowing it to take up acetate [Nguyen et al., 2011]. Therefore, after the sudden decrease of *Ca. Accumulibacter*, *Tetrasphaera* took over and the P-removal was maintained. The co-existence

## Chapter IV. Transforming acetate-propionate-grown granules into AGS treating fermentable and polymeric organic substrates

---

of these two PAO occupying different niches was previously reported as a factor of P-removal stability in the EBPR process [Nguyen et al., 2011].

### IV.4.6 The microbial communities in flocs and granules

A dynamic equilibrium tends to establish between flocs and granules, with the bacteria detaching from the surface of the granules forming new flocs, and parts of broken granules and flocs aggregating or serving as seed for new granules [Barr et al., 2010a, Gonzalez-Gil and Holliger, 2014, Szabo et al., 2017a, Zhou et al., 2014]. Several studies report big differences between the microbial communities in flocs and granules [Liu et al., 2017], but generally they are more similar between flocs and granules fractions in a single reactor than between the the same fractions in different reactors [Zhou et al., 2014]. The data collected during this experiment confirms that even though some taxa were slightly enriched in flocs or in granules, the overall bacterial community was globally similar in the two fractions.

Yet, several factors can favor the enrichment of specific taxa in one or the other fraction. The sludge retention time is higher in granules due to their excellent settling properties [Ju and Zhang, 2015]. Therefore slow growing bacteria can be preferentially enriched in granules. The nitrite oxidizing bacteria *Ca. Nitrotoga*, the GAO *Ca. Competibacter* and the fermentative PAO *Tetrasphaera* were maybe slightly more abundant in the granular fraction for this reason. Depending on the stability of the granules, the microbial populations located far from the surface of the granule are more or less exposed to erosion [Szabo et al., 2017a]. In the case of very resistant granules, those microbial populations will be sparse in the flocs fraction and the flocs will mainly contain bacteria that are located at the surface of the granules [Weissbrodt et al., 2013]. Indeed, in the present study aerobic heterotrophic bacteria such as *Thiothrix* and *Zoogloea* were enriched in flocs. Conversely to other bacteria, the ones having the capability to use nitrate as electron acceptor may have to possibility to grow in the regions located under the surface of the granules. These bacteria will be less susceptible to erosion and therefore enriched in granules. This could be the reason why *Rhodobacter*, *Ca. Competibacter* and *Tetrasphaera* were generally more abundant in the granules.

## IV.5 Conclusion

This study was designed to assess the impact of the introduction of fermentable and polymeric compounds in the influent wastewater on the bacterial community, the settling properties and the nutrient removal performance of an AGS sequencing batch reactor operated previously with a simple monomeric influent containing acetate and propionate.

The transition from simple to complex monomeric synthetic wastewater induced a change in the microbial communities with the increase of Actinobacteria and Saccharibacteria to the detriment of Betaproteobacteriales. During the transition to the complex monomeric wastewater, the proportion of the classical PAO *Ca. Accumulibacter* decreased suddenly, to

the benefit of the fermentative PAO *Tetrasphaera* and the classical GAO *Ca. Competibacter*. However the attack of a phage on *Ca. Accumulibacter* may have increased the changes induced by the wastewater changes and the classical PAO reestablished in the system after eight months, as the AGS was still fed with complex monomeric wastewater. However the microbial communities treating simple wastewater remained significantly different from the one treating complex wastewater, especially with important groups of fermenting bacteria being absent from the first one. This points out that important taxa are missing when working only with simple wastewater.

This drastic change of the microbial community would have remained unnoticed without the monitoring of the bacterial communities by amplicon sequencing, as no changes in the settling properties and nutrient removal efficiency was measured after the transition to complex monomeric wastewater.

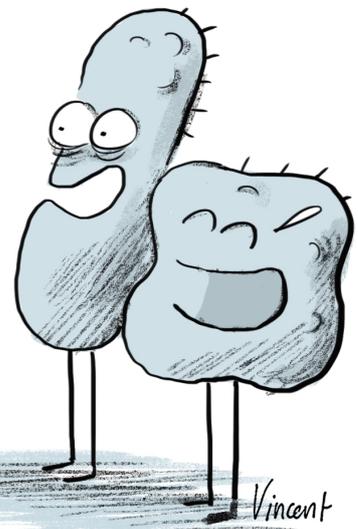
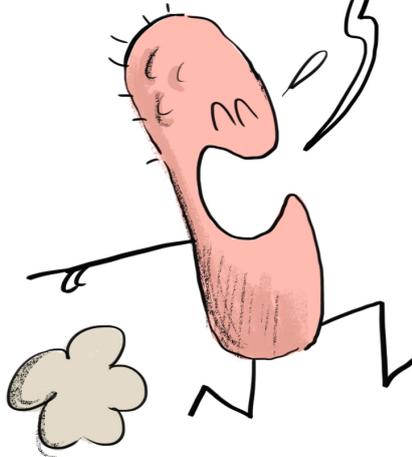
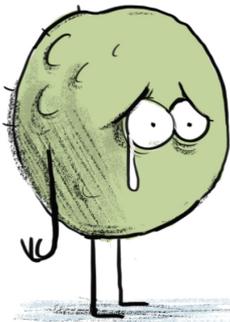
Conversely, the changes in the bacterial community were less spectacular with the transition from complex monomeric to polymeric wastewater. Yet the granulation of the sludge was lost during the transition and recovered only three months after. The TN-removal performances were also slightly lower after the introduction of polymeric compounds in the wastewater.

The modifications in the bacterial community induced by the change of substrate were mostly reversible, the few remaining differences can be stochastic, due to other factors like the priority effect, or to the difference in the COD concentrations between the simple wastewater used at the beginning and at the end of this study.

These results confirm that the microbial community of AGS is very dynamic and other factors than wastewater composition can trigger important changes in its composition. Yet, this dynamism does not necessarily alter the stability of the settling properties or the nutrient removal performance of this AGS, which can be assured through functional redundancies as indicated by multiple stable states with complex monomeric wastewater as influent.



HA, HA, HA,  
IL A UN TOUT  
PETIT GÉNOME!



Vincent



# V Metagenomic analysis of different AGS samples

## V.1 Introduction

### V.1.1 The use of molecular techniques to study microbial communities

Numerous bacteria found in aerobic granular sludge (AGS) belong to genera that have not been isolated in pure culture (Chapter II, III and IV). If amplicon sequencing or quantitative fluorescence *in situ* hybridization (qFISH) allow to assess the microbial composition of a system, the individual metabolism of the different microorganisms is often more difficult to investigate. Molecular methods can help to gain information about those metabolisms [Cetin et al., 2018, McIlroy et al., 2014]. DNA sequencing technologies, starting with Maxam and Gilbert's [1977] and Sanger's in [1977], evolved to next-generation sequencing technologies producing high quantities of short-reads genomic data, at a price constantly decreasing [van Dijk et al., 2018b]. The knowledge obtained from the genomic information collected can allow to estimate the composition of a microbial community or enable inferences on the metabolism encoded by the DNA sequences, possibly being the starting point of transcriptomic or/and proteomic studies [He and McMahon, 2011, Kunin et al., 2008, McIlroy et al., 2017]. The knowledge acquired on the DNA of the bacteria is also used to design probes for *in situ* hybridization and visualization (FISH) or primers for quantitative PCR. Microautoradiography-(MAR-)FISH has for example been used to investigate the metabolic activity of bacteria in activated sludge performing biological phosphorus removal [Kong et al., 2005, McIlroy et al., 2014, Okabe et al., 2004, Xia et al., 2008a]. Little by little, the information collected with those molecular methods provides precious understanding of the function of populations within microbial communities in systems that were before considered as black boxes [Kristiansen et al., 2013, Lawson et al., 2017, Luecker et al., 2010, Nielsen et al., 2012b, Weissbrodt et al., 2014b, Wilmes et al., 2008, Xia et al., 2014].

### V.1.2 Grouping DNA sequences according to their phylotype remains challenging

The separation of the short DNA reads coming from high-throughput sequencing into distinct phylogenetic bins (binning) remains challenging and needs to be adapted on a case-by-case basis [Sloan et al., 2013]. Reference-based binning sorts the reads according to their homology with sequences from reference genomes. The major drawback of this approach is the incompleteness of genomic databases which often simply do not contain the reference genomes corresponding to part of the taxa present in the community to study [Leventhal et al., 2018, Wang et al., 2014b]. Unsupervised binning methods are thus often preferred. They are generally based on oligonucleotide frequencies [Vinh et al., 2015, Wrighton et al., 2012] and differential coverage from different extractions/samples [Nielsen et al., 2014], or both [Albertsen et al., 2013, Alneberg et al., 2014, Skennerton et al., 2015].

The tetranucleotide frequency is phylotype-specific, but DNA sequences shorter than 2 kbp are too small to provide a reliable information [McHardy et al., 2007, Pride et al., 2003, Teeling et al., 2004]. Le Van et al. [2015] bypassed this difficulty by grouping the reads and performing the binning based on tetranucleotide frequency on these groups of reads. Another drawback of this method is that the tetranucleotide frequency is not even on the whole genome. For example, sequences acquired through horizontal gene transfer can have a tetranucleotide frequency differing a lot from the rest of the genome [Deschavanne et al., 2000, Dick et al., 2009]. Therefore, this binning method is only rarely used alone for complex metagenomes. If different samples with differential abundance of the different taxa can be obtained, the binning based on differential abundance is generally preferred [Albertsen et al., 2013, Nielsen et al., 2014, Skennerton et al., 2015]. Albertsen et al. [2013] used two different extraction methods on the same biomass sample to obtain this differential abundance and thus recovered nearly complete genomes including some from taxa representing less than 1 % of the total read abundance.

Most of these binning methods rely on an assembly common to all the samples, requiring that the initial reads are pooled and assembled. This can increase the polymorphism of the sequences to assemble and thus the computational cost of the initial assembly [Sloan et al., 2013]. Moreover, obtaining good quality assemblies from short reads of complex metagenomes is not an asset [Schmid et al., 2018]. The presence of closely related species, polymorphisms or sequencing errors create ambiguities that are sometimes impossible to resolve [Olson et al., 2017]. Large intragenomic or intergenomic duplicated regions or repeats that are bigger than the region spanned by a pair of reads lead to fragmented or miss-assembled contigs [McIlroy et al., 2014, Olson et al., 2017, Wooley et al., 2010]. Metagenome assembled genomes are therefore often incomplete and/or contaminated.

Clustered regularly interspaced short palindromic repeats (CRISPR) represent another challenge for genome assembly. CRISPR are part of an adaptative antiphage-defense system which is present in about 50 % of the sequenced bacteria [Hille et al., 2018]. The genomes of bacteria endowed with a CRISPR-Cas system will obviously be very difficult to reconstruct

from metagenomes due to the high polymorphism of the CRISPR regions. Dedicated tools exist to reconstruct CRISPR loci [Skennerton et al., 2013].

### V.1.3 Long-read sequencing technologies facilitates the assembly of complete genomes

Long-read sequencing technology became available in 2011 with the single molecule real time (SMRT) sequencing from Pacific Bioscience (PacBio) [Eid et al., 2009] and was followed in 2014 by nanopore sequencing from Oxford nanopore technologies (ONT) [Jain et al., 2015]. These technologies allow the sequencing of fragments of several kilobase pairs (kbp), with an error rate of around 15 % [Quail et al., 2012]. The errors in PacBio reads being randomly distributed, they can be corrected by the superposition of multiple sequences of the same clone or multiple sequencing of the same fragment (CSS) [Koren et al., 2012, Ono et al., 2013, van Dijk et al., 2018b]. PacBio sequencing has been widely used to close high-quality eukaryotic and prokaryotic genomes, through the advantage of reads longer than most repeat regions. The errors are corrected with the information contained in the read themselves or with Illumina short-reads from the same DNA [Koren and Phillippy, 2015, Ricker et al., 2016].

The use of PacBio in metagenomic studies only began to become common practice since 2016. The high error rate of long reads, the lack of adapted bioinformatic tools and the limited number of PacBio sequencing devices has limited, for a time, their use for metagenomes sequencing [English et al., 2012]. Nowadays, numerous metagenomic studies use long-reads in complement of short-reads [Daims et al., 2015, McIlroy et al., 2018, Schmid et al., 2018, Slaby et al., 2017]. Those comparing the hybrid assembly of PacBio/ONT and Illumina sequences with the assembly of Illumina sequences revealed that the introduction of long-reads significantly improved the quality of the assembly [Frank et al., 2016, Tsai et al., 2016]. PacBio sequences (long-reads) have been assembled without using short-reads for an enriched microbial community [Ahlgren et al., 2017], but to the knowledge of the candidate, this was not tried with more complex metagenomes.

### V.1.4 Denitrification is performed by microorganism of various phylotypes

Nitrogen usually enters the wastewater treatment plants in the form of ammonium (linked or not to organic matter). It is oxydized aerobically into nitrite and nitrate by ammonium oxydizing organisms [von Sperling, 2007]. Several steps are then required to reduce nitrate into dinitrogen gas ( $N_2$ ). The classical denitrification pathway, which can be performed by a single or a combination of different bacteria, includes dissimilatory nitrate, nitrite, nitric oxide, and nitrous oxide reduction [Kuypers et al., 2018].

Two different type of dissimilatory nitrate reductases can reduce nitrate into nitrite, a membrane-associated respiratory nitrate reductase (Nar) encoded by the operon (narGHIJ) and a periplasmic nitrate reductase (Nap), which is found in gram-negative bacteria [Moreno-Vivian et al., 1999]. Alternatively, nitrate can be assimilated in the bacterial biomass through an assimilatory

pathway which will not be discussed further here [Kuypers et al., 2018, Moreno-Vivian et al., 1999]. Nitrite reduction can be carried out by different nitrite reductases: a cytochrome c carrying multiple heme (operon *nrf*), a cytochrome cd1 nitrate reductase (NirS) or a copper-containing nitrite reductase (NirK) [Kuypers et al., 2018, Moreno-Vivian et al., 1999]. A nitric oxide reductase (Nor) reduces nitric oxide into nitrous oxide, which can be further reduced into dinitrogen gas by a nitrous oxide oxydase (Nos) [Kuypers et al., 2018]. If this last enzyme is not present, nitrous oxide, which is a greenhouse gas, will be the end product of the denitrification. It is not clear if the microorganism performing denitrification in AGS all possess a functional Nos. Indeed, the different parts of this denitrification pathway have been observed in multiple bacterial taxa and can vary across different strains belonging to a single genus [Barr et al., 2016, Kim et al., 2013, Wang et al., 2014b].

### V.1.5 Phosphorus cycling related genes

The genes involved in the polyphosphate accumulating organism (PAO) metabolism include polyphosphate kinases (*ppk*), which catalyzes the elongation of polyphosphate chains, an exopolyphosphatase (*ppx*), catalyzing the degradation of polyphosphate chains and phosphate transporters [Akiyama et al., 1992, Rao et al., 2009]. All these genes are commonly found in PAO and non-PAO bacteria and can therefore not be used as markers of PAO-metabolism.

Sequences encoding for two different types of phosphate transporters, the high-affinity phosphate ABC transporter system (*Pst*) and the low-affinity inorganic phosphate transporter (*Pit*) were found in the genomes of *Ca. Accumulibacter* and *Tetrasphaera* PAO [Kornberg et al., 1999, McIlroy et al., 2014, Vanveen et al., 1994]. Homologous sequences of the *pit* genes were not found in the GAO *Ca. Competibacter* making this gene a good candidate to discriminate between PAO and GAO metabolism [McIlroy et al., 2014]. Yet sequences coding for *Pit* system were also detected in non-PAO and even in GAO organisms such as *Defluviicoccus* [Wang et al., 2014a]. A genomic differences allowing to differentiate PAO from non-PAO metabolism has not been found yet.

### V.1.6 *Ca. Accumulibacter* is a highly diversified genus

*Ca. Accumulibacter* strains were clustered in 14 clades, distributed in two types (I and II), based on their polyphosphate kinase (*ppk1*) gene [McMahon et al., 2002]. Activated sludge performing enhanced biological phosphorus removal (EBPR), often harbour multiples *Ca. Accumulibacter* clades whose proportions can change between different reactors and in time [Barr et al., 2016, Flowers et al., 2013, He et al., 2007, Kim et al., 2013, Martin et al., 2006].

The genomic and *in situ* denitrification potential of the different clades has been studied and seems to differ between clades. Metagenomic evidence supports AGS sequencing batch reactor (SBR) observations showing that some *Ca. Accumulibacter* populations can denitrify from nitrite [Kim et al., 2013]. A complete genome of 'Candidatus *Accumulibacter* clade IIA' has

been assembled from an highly enriched sludge [Martin et al., 2014] and several metagenome-assembled-genomes (MAG) were constructed from samples collected in multiple reactors with an abundance-based binning method [Skenner et al., 2015]. The genes encoding the respiratory nitrate reductase Nar were not found in the genome of *Ca. Accumulibacter* IIA [Martin et al., 2006]. Flowers et al. [2009] confirmed through batch experiments that a sludge enriched with clade IIA could not perform phosphorus removal in presence of nitrate, while a sludge enriched with clade IA could. The rest of the denitrifying pathway was found in the genome of *Ca. Accumulibacter* IA and IIA, but parts or full Nos operon seemed to be missing from the genomes of clades IIC, IIF and IC [Barr et al., 2016]. Given the different metabolism harboured by the different *Ca. Accumulibacter* clades, the genetic potential and the dynamics of these clades are worth to be studied.

### V.1.7 Objectives

The experiments presented in the previous chapters allowed to collect useful information about the dynamics of bacterial communities in relation to organic fermentable and polymeric compounds in the influent wastewater and the resulting physical and nutrient removal characteristics of the AGS. Based on previous studies that have investigated the metabolism of bacteria from wastewater treatment systems, the metabolic properties of bacterial populations can be hypothesized based on their 'identity'. However, and as mentioned above, the functional traits are not evenly distributed across a particular genus. Moreover, some of the bacteria detected in these experiments have no taxonomically classified homologues in public 16S rRNA genes databases. Therefore, the whole metagenomes of the sludge collected when the reactor was operated with different wastewaters was sequenced, assembled and examined in order to investigate the putative functions of the bacteria present in the biomass with a focus on genes involved in phosphate removal, denitrification, biofilm formation and hydrolysis.

## V.2 Material and methods

### V.2.1 Biomass sample collection

Four samples from EPFL experiment 1 described in chapter IV were chosen for full DNA sequencing: samples collected on day 71 (d71) when the reactor was treating simple wastewater, on day 322 and 427 (d322 and d427) when the reactor was treating complex monomeric wastewater, and on day 740 (d740) when the reactor was treating complex polymeric wastewater.

### V.2.2 Metagenomic DNA extraction from AGS biomass

Initially, four different protocols were tested for the extraction of metagenomic DNA: a modified version of the bacterial genomic DNA isolation CTAB protocol (DOE Joint Genome Institute), the Maxwell® 16 Tissue DNA Purification Kit (Promega), the PowerBiofilm® DNA Isolation Kit (MOBIO Laboratories) and the FastDNA® SPIN Kit for Soil (MPBIO). When using DNA extraction kits, aliquots of 125 µl of AGS biomass suspension in TE buffer (10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA) normalized at an OD<sub>600 nm</sub> value of 1, was applied for metagenomic DNA extraction. In the case of the CTAB protocol, 740 µl of the same biomass suspension was used.

The bacterial genomic DNA isolation CTAB protocol (hereafter referred to as extraction A) was applied as follows. To 740 µl of cell suspension, 20 µl of 100 mgml<sup>-1</sup> lysozyme solution was added and the sample was incubated for 5 min at room temperature. After addition of 40 µl of 10 % SDS and 4 µl of Proteinase K solution (23 mgml<sup>-1</sup>, Sigma-Aldrich), the sample was incubated for 1 h at 37 °C. Then, 100 µl of 5 M NaCl and 100 µl of a 65 °C pre-warmed CTAB solution (4 % NaCl, 10 % hexadecyltrimethyl ammonium bromide) were added and the reaction was incubated for 10 min at 65 °C. Extraction of DNA was obtained by adding 0.5 ml of a chloroform:isoamyl alcohol (24:1) solution and by centrifugation for 10 min at 16'000 × g. The upper aqueous phase was transferred to a new tube and mixed with 0.6 volume of cold isopropanol. After 30 min incubation at room temperature, the DNA was collected by 15 min centrifugation at 16'000 × g. The pellet was washed with 70 % ethanol and centrifuged again. The pellet was air-dried for 5 min and resuspended in 97.5 µl of ddH<sub>2</sub>O. A volume of 2.5 µl of RNase (4 mgml<sup>-1</sup>) was added and the sample was incubated for 20 min at 37 °C. DNA was precipitated by addition of potassium acetate and ethanol as presented above. Finally, the pellet was resuspended in 100 µl of 10 mM Tris-HCl buffer (pH 8.0).

All three commercial kits were applied according to the manufacturer's instructions with the following modifications.

Maxwell® 16 Tissue DNA Purification Kit (hereafter referred to as extraction B): Elution of the DNA from the magnetic beads was performed in TE buffer. Eluate was treated with RNase as follows. To 150 µl of eluate, 50 µl of ddH<sub>2</sub>O and 5 µl of RNase A solution (4 mgml<sup>-1</sup>, Promega) were added and the reaction was incubated for 20 min at 37 °C. The DNA was then precipitated by adding 20 µl of 3 M potassium acetate solution (pH 5.2) and 400 µl of 100 % ethanol for 15 min at -80 °C. The DNA was recovered centrifugation at 4 °C and full speed in a tabletop centrifuge (16'000 × g) for 15 min. The pellet was washed three times in 250 µl of 70 % ethanol and centrifuged again. The DNA pellet was air-dried for 5 min and resuspended in 50 µl of 10 mM Tris-HCl buffer (pH 8.0).

FastDNA® SPIN Kit for Soil (hereafter referred to as extraction C): Cell lysis was performed as presented above, followed by a 10 min centrifugation step (steps 4-5 in the instruction protocol).

PowerBiofilm® DNA Isolation Kit (hereafter referred to as extraction D): Cell lysis (step 5 in the instruction protocol) was performed by bead-beating using the Precellys 24 Tissue Homogenizer (Bertin) at 6'000 rpm for two times 20 s with 30 s break.

All DNA samples were stored at -80 °C until further use.

### V.2.3 DNA sequencing

The DNA samples obtained with extractions A and B were transmitted to the Lausanne Genomic Technologies Facility (University of Lausanne, Switzerland) for PacBio (extraction A) and Illumina (extraction A and B) sequencing.

PacBio sequencing : 4.5 µg of the DNA from sample d71 and 4 µg of the DNA from samples d322, d427 and d740 were used to prepare a SMRTbell library with the PacBio SMRTbell Template Prep Kit 1 (Pacific Biosciences, Menlo Park, CA, USA) according to the manufacturer's recommendations. DNA fragments were selected by size on a BluePippin system (Sage Science, Inc. Beverly, MA, USA) for molecules larger than 7 kb for sample d71, 8 kb for sample d322, 7 kb for sample d427 and 10 kb for sample d740. The resulting libraries were sequenced with v2/v2.1 chemistry and diffusion loading on a PacBio Sequel instrument (Pacific Biosciences, Menlo Park, CA, USA) at 600 min movie length using SMRT cells v2.

Illumina sequencing : DNA extractions A and B from samples d71, d322, d427 and d740 were sequenced in multiplex on an Illumina HiSeq platform in paired-end mode (2x100 bp).

### V.2.4 Assemblies of PacBio long-reads

Bamtools 2.4.1 [Barnett et al., 2011] was used to convert the long-reads from 'bam' to 'fasta' format and keep only the sequences longer than 500 bp. For each data-set, an approximate mapping of the long-reads against themselves was performed with minimap2 v. 2.12 [Li, 2018] and the option *-x ava-pb*. A de novo assembly of the long-reads and an assembly graph were created by using miniasm v 0.2.r159 [Li, 2016] on the minimap2 output. Several combinations of options were tested in order to maximize the contigs length but keeping the probability of creating misassemblies as low as possible. The quality of the assemblies was assessed by mapping the original raw PacBio long-reads on the assembly with minimap2 and the option *-ax map-pb* and samtools 1.8 [Li et al., 2009] to convert, sort and index the mapping. The visual inspection of the mapping was performed on the genome viewer IGV v. 2.4.16 [Robinson et al., 2011].

After comparison of the assembly statistics and the quality of the mapping, the combination of options *-h 700 -s 3000 -g 500 -r 0.75 -n 5* in miniasm was chosen, where *-h* corresponds to the maximal overhang (unmapped region that should be mapped if the reads were really overlapping) length (default 1000), *-s* is the minimum mapping length (default 1000), *-g* is the maximal gap differences in a mapping (default is 1000), *-r* is the maximal overlap drop ratio

defining the threshold for considering an overlap as short compared to the maximal overlap involving a sequence (default is 0.7),  $-n$  is the number of rounds for progressive short overlap removal.

To remove the errors still present in the assembly, seven rounds of polishing were applied. After each round of polishing, the quality of the assembly was assessed by mapping the raw PacBio long-reads on the assembled reads as described above. A round of polishing was performed as follows : the raw PacBio long-reads were mapped on the current assembly with minimap2 and the wrapper especially designed for PacBio data: pbmm2. Arrow<sup>®</sup> (2011-2018, Pacific Bioscience of California, Inc.) was used to establish consensus sequences based on this indexed mapping and the current assembly.

A co-assembly of all the long-reads of the four samples was performed with the idea of serving as a basis for the binning. The procedure used for the individual assembly of the PacBio long-reads was applied on the concatenated individual input files.

### V.2.5 Assemblies of Illumina reads and hybrid assemblies

The trimming and quality filtering of Illumina sequences was performed with trimmomatic v. 0.36 [Bolger et al., 2014] with a sliding window of 10 base pairs (bp) and a minimal quality score threshold (phred33) of 15 and a minimal sequence length of 50 bp.

A first de novo assembly of each data-set was performed with the trimmed sequences using Spades v. 3.12.0 [Nurk et al., 2013] and the script *metaspades.py* with increasing kmer sizes of 21, 33, 55, 77. A second de novo assembly was performed with Megahit v. 1.1.4 [Li et al., 2015], a minimal kmer size ( $-k-min$ ) of 27, a maximal kmer size ( $-k-max$ ) of 97 and 10 steps ( $-k-step$ ).

Two co-assemblies of all the Illumina reads were performed following the procedure used for the assemblies of individual data-sets, by using the concatenated file of individual input files as input.

### V.2.6 Hybrid assemblies of Illumina reads and PacBio long-reads

An hybrid assembly of each Illumina data-set and the PacBio of the corresponding sample was performed using Spades and the command *metaspades.py* with kmer lengths 21, 33, 55 and 77 and the option  $-pacbio$ .

An hybrid co-assembly of all the Illumina reads with all the PacBio long-reads was tried with the same parameters, but it was not achieved since it got over the 500 GB of RAM available on the server.

### V.2.7 In silico 16S rRNA gene extraction

Potential 16S rRNA genes were extracted from the assemblies by using *infern* v. 1.1.2 [Nawrocki and Eddy, 2013]. The assembly sequences were compared to a consensus RNA profile of bacterial 16S rRNA genes (RF00177.cm), provided by Rfam [Bateman et al., 2017] by using the *infern* command *cmsearch*. This alignment tool is for this particular purpose more reliable than standard mapping softwares since it takes into account the coherence of the secondary structure of the coded rRNA. The DNA sequences matching the 16S rRNA gene consensus were collected using *samtools* with the option *faidx* and compared using *blast* against the MiDAS 16S rRNA gene database v. S123\_2.1.3 [McIlroy et al., 2017].

A set of complete reference genomes was chosen based on the results of the *in silico* 16S rRNA gene extraction from the PacBio and the hybrid assemblies taking into account the lengths of the corresponding contigs.

### V.2.8 Estimation of completeness and contamination

The completeness and contamination of potential genomes was estimated by using two different softwares: *CheckM* v. 1.0.12 [Parks et al., 2015] with the options *lineage\_wf - -reduced\_tree* and benchmarking universal single-copy orthologs (BUSCO) v. 3 [Kriventseva et al., 2015].

### V.2.9 Binning of Pacbio contigs

The PacBio contigs were grouped into metagenome assembled genomes (MAG) with a home made script inspired from binning techniques used with Illumina data [Albertsen et al., 2013, Alneberg et al., 2014]. Three different and independent distances were averaged to obtain a global distance between contigs : a distance based on the tetranucleotides signatures of the DNA sequences, a distance based on the coverage of the Illumina read from the same sample and a distance based on the overlapping of BUSCO v. 3.

The pairwise distance based on tetranucleotide characteristic frequencies vector (hereafter named 'tetranucleotide frequency distance') between two DNA sequences was computed as follows : the tetranucleotide frequency vector was computed on sliding windows of 2000 bp and a resolution of 200 bp over the whole DNA sequences. The tetranucleotide characteristic of each DNA sequence was computed by iteratively removing the sliding window for which the euclidean distance between the corresponding tetranucleotide frequency vector and the mean frequency vector on all the remaining sliding windows was maximal. The iteration was repeated until the maximal distance between a single window and the mean frequency vector was below  $1 \times 10^{-5}$ . The maximum number of iterations was limited so that at least 90 % of the DNA sequence was taken into account to compute the characteristic tetranucleotide frequency vector. An absolute euclidean distance was computed between the resulting frequency vector of two DNA sequences. This absolute distance was computed for reference

genomes. A normalized distance was also computed according to the Formula V.1 (where  $x_i$  is the frequency of the tetranucleotide number  $i$  in the first contig,  $y_i$  is the frequency of the tetranucleotide number  $i$  in the second contig, and  $m_i$  is the mean frequency of the tetranucleotide number  $i$  computed on all the contigs of the corresponding sample), in order to be averaged with the two other distances.

$$\text{Dist}_{\text{norm}} = 0.5 \cdot \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n ((x_i - m_i)^2 + (y_i - m_i)^2)} \quad (\text{V.1})$$

The presence of BUSCO was assessed by BUSCO v. 3 [Kriventseva et al., 2015] with the bacteria BUSCO set 'bacteria\_odb9' containing consensus hidden Markov model (HMM) profiles of 148 BUSCO and the option *-m geno*. The pairwise distance based on BUSCO was computed as the number of common BUSCO (complete or duplicated) over their total in the two sequences.

The pairwise distance based on the coverage between two contigs was computed as follows: the two Illumina data-sets corresponding to two different extraction methods (A and B) were aligned separately on the PacBio contigs from the same sample with bowtie v. 2.3.4.1 [Langmead and Salzberg, 2012] and the options *-k 1 -p 8 -R 3 -D 20 -N 1 -X 1000 -q -phred33*. The output sam file was converted into a sorted bam file with samtools v.1.8 and the commands *view -b -S* and *sort*. The coverage on each nucleotide was computed from the mapping by using BEDtools v. 2.26.0 [Quinlan and Hall, 2010] and the options *genomecov -d*. The average coverage on each contig was computed with a home made script and the distance between two contigs was computed as the normalized distance (V.1) of their two-dimension coverage vectors (coverage with extraction A, coverage with extraction B).

These three distances were averaged and an home made script was used to group the contigs having a distance below 0.022 in a single MAG. The grouping was manually inspected and few (less than 10) contigs were removed or put in another bin manually.

### V.2.10 Annotation and search of homologous proteins

The MAG and the unbinned contigs were annotated by using prokka v. 1.13 [Seemann, 2014] and the option *-metagenome*. The reference genomes were re-annotated in the same way.

Prokka uses several dependencies. Blast+ [Camacho et al., 2009] searches for similarities in the database of prokaryotic proteins sequences, including the clusters of orthologous groups of proteins (COGs) database [Galperin et al., 2015]. Prodigal detects protein-coding sequences (CDS). Hmmer v. 3 searches similarities with protein family profiles of the database HAMAP [Auchincloss et al., 2014], a manually curated database of protein family profiles.

The KEGG database [Kanehisa et al., 2019] was used to identify putative affiliations of protein sequences to metabolic pathways.

Sequences homologous to specific proteins families involved in P-metabolism and denitrification were searched with hmmer v. 3.1b2 comparing hidden Markov models (HMM) of these proteins with the coding sequences annotated by prokka in the MAG, the unbinned contigs and the reference genomes. The gather score thresholds provided with the HMM were used to decide if the proteins were similar enough to the HMM.

The HMM family profiles for this analysis were selected via InterPro v. 73 [Mitchell et al., 2019] a database regrouping various protein sequence databases. The HMM used in this study were selected from the databases TIGRFAMs v.15 [Haft et al., 2001] and Pfam v. 32. [Bateman et al., 2018]. A home made hmm model was built from sequences of the low-affinity transporter Pit from the genomes of *Ca. Accumulibacter* and *Tetrasphaera* (electronic supplementary material A.4.1 File A.1) hmmer v. 3 and the command *hmmbuild -amino*.

### V.2.11 Repartition of the COG in the main MAG

The COG detected during the annotation with prokka were compared for the MAG having an estimated completeness higher than 50 % (further referred to as main MAG). A principal component analysis (PCA) was performed on the COG presence (1)/absence (0) in the main MAG, in R, with the package *vegan*.

Hierarchical clusterings and the corresponding heatmaps were constructed in R with the function *heatmap.2* from the package *gplots* [Warnes et al., 2016] on the sets of COG linked to "Inorganic ion transport and metabolism [P]" and "Signal transduction mechanisms [T]".

### V.2.12 Phylogenetic trees

For the construction of phylogenetic trees of 16S rRNA gene sequences, the sequences were first aligned together to the consensus bacterial 16S rRNA gene sequence by using *infernal* v. 1.1.2 and the option *-A*. The resulting alignment was visually inspected and rare evident inconsistencies were manually corrected. The alignment was then used as input file in *Raxml* v. 8.2.10 [Stamatakis, 2014] with the general time reversible model of nucleotide substitution under the Gamma model of rate heterogeneity (*-m GRTGAMMA*) to construct a tree. The corresponding bootstrap values were calculated based on 100 inferences using the model *CAT* which also allow the integration of rate heterogeneity but has a lower computational cost (*-m GRTCAT*).

The amino acid sequences were aligned together to a hidden Markov model (HMM) of a group of defined homologous protein sequences, using *hmmer* v. 3.1b2 [Institute, 2018] and the option *-A*. The alignment was then used as input file in *Raxml* with the *-m PROT* option, under the Gamma model (*GAMMA*) and the amino acids substitution models *JTT* [Jones et al., 1992] and the optimization of substitution rates *F* to construct a tree. The corresponding bootstrap values were calculated based on 100 inferences using the model *CAT* which also allow the integration of rate heterogeneity but has a lower computational cost (*-m PROTCATJTTF*).

All the phylogenetic trees edited in R with the packages *ggtree* [Yu et al., 2017] and *treeio* [Yu, 2018].

### V.2.13 Coverages of the enzymes in the different samples

The enzyme commission (EC) numbers corresponding to enzymes referenced in the KEGG database [Kanehisa et al., 2019] were extracted from the *prokka* annotation of the PacBio contigs. For each sample, the mean coverages of the contigs on which an EC number was detected were added to estimate the coverage of homologues of the corresponding enzymes. The mean coverage of the EC number in each sample was then normalized by dividing it by the sum of the mean coverages of the EC numbers in the sample. Hierarchical clustering and the corresponding heatmaps were constructed in R with the function *heatmap.2* and the color package *viridis* [Garnier, 2018] on the matrix of EC number coverages.

## V.3 Results

### V.3.1 Sequencing of metagenomic DNA

The total DNA of four biomass samples collected in the AGS sequencing batch reactor described in chapter IV was sequenced with two different sequencing technologies : Illumina and PacBio. Four extraction methods were initially tested in an attempt to obtain different relative abundance of the DNA from the different bacteria. Amplicon sequencing of the 16S rRNA genes revealed that the differences between the different extractions of a sample were not as high as expected (Figure V.1). The extraction method giving the longest DNA fragments (A) (supplementary information A.4.2) and the one previously used for the extraction of all the amplicon sequencing presented in this thesis (B) were used to extract the DNA that was sequenced with Illumina HiSeq technology. This produced 39.6 Gbp of paired-end short reads data (Table V.1), with an estimated error rate lower than 0.3 %.

The DNA extracted with the method A was also sequenced with PacBio, producing 26.5 Gbp of long-reads data with an estimated error rate of around 13 % [Quail et al., 2012].

The filtering of Illumina reads based on the quality scores removed less than 2 % of the total base pairs. The filtering based on the length of the PacBio long-reads removed less than 1 % of the total base pairs. The statistics for the length of the PacBio long-reads (N50 and maximal length) seem to be linked with the distribution of the size of the extracted DNA. The DNA extracted with method A from sample d427 had the smallest average size (6650 bp, see section A.4.2). The length of the longest subread in this sample is shorter than in the other ones. For the sample d322, the average size of the extracted DNA is larger than for sample d427 (10236 bp). Despite that, the N50 for the long-reads of this sample is the smallest. With an average size of 8958 bp, the DNA extracted from sample d71 allowed to recover long-reads with a N50 similar to the ones of d322 and d427, but also the longest subread of this study (116102

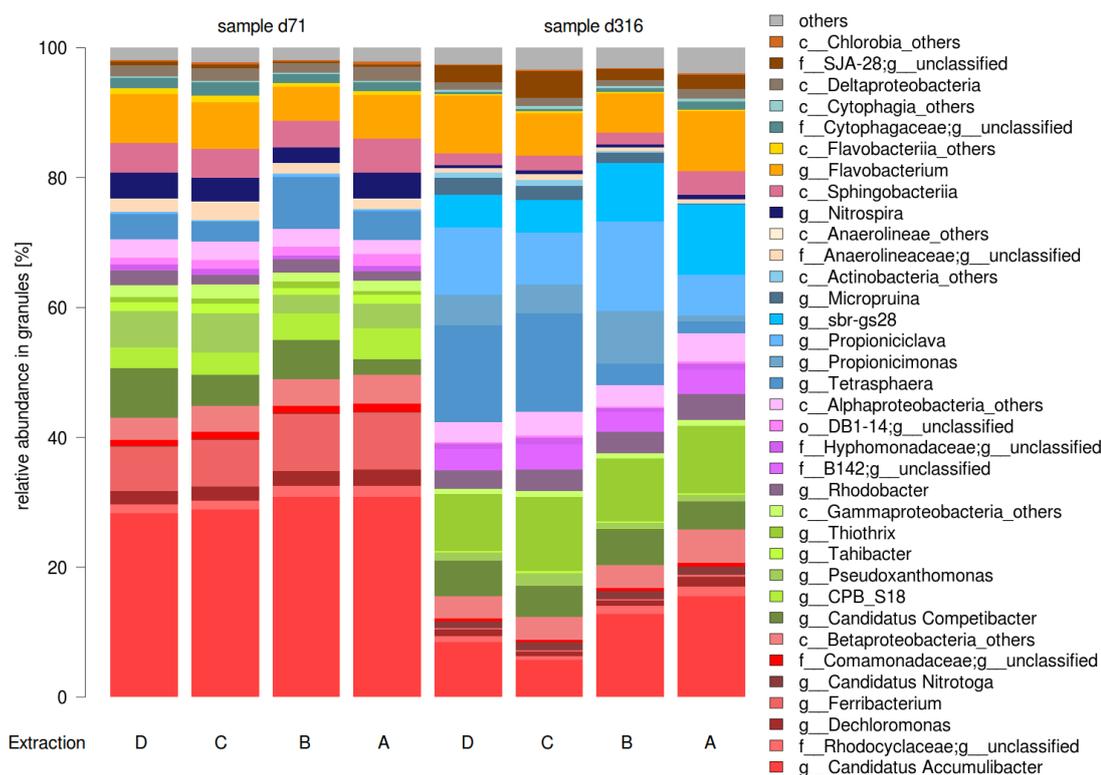


Figure V.1 – Relative abundance of the main bacterial genera estimated by the abundance of the 16S rRNA gene amplicon sequencing of the DNA extracted from sample from day 71 and 316 with four different extraction methods.

bp). The DNA extracted from sample d740 have a very high average size of 20618 bp and the long-reads obtained by sequencing this DNA produced the highest N50 (13024 bp).

### V.3.2 De novo assemblies

Different assembly softwares were tested on PacBio long-reads (assembly Pacbio), Illumina reads (metaspades, megahit) or both (metaspades hybrid). These three types of *de novo* assembly produced very different results (Table V.2). The PacBio assemblies produced the longest contigs with a N50 ranging from 55 511 bp for d427 to 211 404 bp for d740. The maximal length of the contigs was above 3.4 Mbp for sample d71, d427 and d740, which could already be complete bacterial genomes. The N50 is here correlated with the length distribution of the long-reads used for the assembly (Table V.1). The number of Pacbio contigs is rather small ranging from 548 (d322) to 1997 (d427).

The assembly of the Illumina reads produced much smaller contigs with a N50 ranging from 8482 bp for d322\_B\_megahit to 19743 bp for d740\_A\_metaspades. Interestingly, this sample corresponds to the one with the highest average size of the extracted DNA. Since the Illumina reads all have the same size, it was not expected that the length of the starting material

## Chapter V. Metagenomic analysis of different AGS samples

Table V.1 – Characteristics of the PacBio and Illumina sequences before and after the filtering.

Sample name	before filtering		after filtering		N50	max. length
	nb. reads (Mbp)	total length [Gbp]	nb. reads [Mbp]	total length [Gbp]		
d71_A_longreads	1.2	7.1	1.1	7.1	8671	116102
d322_A_longreads	0.8	3.3	0.7	3.3	7644	66269
d427_A_longreads	1.3	7.0	1.2	7.0	8043	59540
d740_A_longreads	1.1	9.1	1.1	9.0	13024	85655
d71_A_reads	6.2	61.6	6.1	60.9	101	101
d71_B_reads	5.2	51.2	5.1	50.6	101	101
d322_A_reads	6.2	60.9	6.1	60.2	101	101
d322_B_reads	5.0	49.7	5.0	49.2	101	101
d427_A_reads	5.7	56.5	5.6	55.9	101	101
d427_B_reads	5.0	49.9	5.0	49.4	101	101
d740_A_reads	6.3	62.2	6.2	61.5	101	101
d740_B_reads	5.8	57.1	5.7	56.4	101	101

influences the length of the Illumina assemblies. A possible explanation is that the DNA extracted with the method A from sample d740 was less degraded than for the other samples and extractions, and that this influenced positively the assembly process. The maximal contig length for Illumina assemblies was obtained for the sequencing of DNA extraction A of sample d427 with Metaspades (2.17 Mbp).

Compared with the Illumina assemblies, the hybrid assembly produced longer contigs with a N50 ranging from 15754 bp for d427\_A\_metaspades\_hybrid to 25358 bp for d71\_B\_metaspades\_hybrid, which is still at least two orders of magnitude below the assembly of PacBio reads alone.

The number of complete 16S rRNA genes recovered in the contigs can depend on the quality of the assembly. Among the 16S rRNA genes extracted from the PacBio assembly after only one step of polishing, around one third were incomplete (data not shown), whereas after seven steps of polishing, the proportion of complete 16S rRNA genes is equal or close to 1. With the assembly of Illumina reads, this proportion is below 0.5, with the exception of the assembly d427\_B\_metaspades and d427\_B\_megahit which was around 0.65 but with few 16S rRNA genes detected in total (around 20, Table V.2). The proportion of complete 16S rRNA genes is in general a little higher with the hybrid assemblies than with the Illumina only assemblies, but the total of 16S rRNA genes detected is high in comparison with the non hybrid assemblies.

Based on these statistics, the Pacbio assemblies and the metaspades hybrid assemblies were considered as the most promising for the recovery of interesting information. For logistic reasons, it was decided to present the results of one assembly type in the present thesis. The PacBio assembly was chosen because it produced the longest contigs, the highest proportion

of complete 16S rRNA genes sequences and because it is innovative for the study of complex metagenomes.

The assemblies with all the sequences from Pacbio or Illumina sequencing were realized with the aim of performing the binning based on differential abundance. The quality of the Illumina assemblies was judge too low for this purpose and for the PacBio assembly, a personalized binning method, taking as input the individual Pacbio assembly was preferred to minimize the risk of chimeras.

A phylogenetic tree of the 16S rRNA genes extracted from the PacBio (in green) and the hybrid (in purple) assemblies was constructed in order to evaluate if the different taxa were equally represented in these two assemblies (Figure V.2). The taxa related to Actinobacteria were unfortunately less represented in the 16S rRNA gene sequences found in the PacBio contigs. In particular 16S rRNA gene sequences related to *Tetrasphaera* or *Micropruina* were only found in the hybrid assembly, which suggests that sequences from this genus were not very abundant in the PacBio long-reads. Indeed most of these sequences come from the assemblies performed on the sequences from extraction B. The 16S rRNA gene sequences related to the phylum Armatimonadetes were found in the contigs of the hybrid assembly but not in the PacBio assembly. In this case, the 16S rRNA gene sequences are equally represented in the assembly of the sequences from extraction A and B.

The 16S rRNA gene sequences corresponding to *Ca. Accumulibacter* were very diverse. The comparison with the 16S rRNA gene sequences of the two reference genomes for which 16S rRNA gene sequences were found (*Ca. Accumulibacter* IIA and *Ca. Accumulibacter* IA) suggest that the type I and II were present in the assembly (16S rRNA gene sequences of *Accumulibacter* IA BA-93 and ACC005 are exactly identical).

### V.3.3 Evaluation of the quality of PacBio assembly with different number of polishing rounds

Mapping of the raw Pacbio long-reads and hybrid contigs against the PacBio contigs were used to evaluate the quality of the assembly. The PacBio assembly likely produce consensus genomes, but it was noticed that in some parts of the assembly, the path chosen did not correspond to the majority of the long-reads. Each polishing steps was removing several of these inconsistencies (Figure A.15, A.16, A.17).

### V.3.4 Grouping of MAG

In order to test the pertinence of grouping contigs based on their characteristic tetranucleotide frequency vector, the distance between the characteristic frequency vectors of selected reference genomes were computed (electronic supplementary material Table A.20). Globally, the more the genomes are phylogenetically related, the shorter is the distance.

## Chapter V. Metagenomic analysis of different AGS samples

---

The distances between genomes from *Ca. Accumulibacter* reaches a maximum of 0.0247 for the genomes *Ca. Accumulibacter* IIC SK02 and *Ca. Accumulibacter* IIF SK12, without taking the plasmids into account. The distance between the plasmids and the principal chromosome is between 0.0155 and 0.0255 for *Ca. Accumulibacter* IIA. The distance between the two *Dechloromonas* genomes is 0.0046 and the one between the two *Tetrasphaera* genomes is 0.0112. The maximal distance was 0.1441 between *Flavobacterium* sp. and *Dokdonella* Koreensis. According to the tests performed on the reference genomes, the criterion of tetranucleotide frequency was judged relevant to group contigs together. However, if a contig is small and contain very specific DNA regions such as CRISPR regions, it may have a too different tetranucleotide frequency compared to the rest of the genome and be rejected.

A set of basic information was collected from the reference genomes in order to have a point of comparison for the analysis performed on the MAG constructed in this study (Table V.3). The research of BUSCO was performed on a set of 148 genes conserved in the bacterial kingdom. Yet, the complete set of genes was found in none of the reference genomes. It is possible that some of these genomes are incomplete but it is more likely that this set of 148 BUSCO is not completely universal among bacteria. Duplicated BUSCO were found in nine reference genomes, in particular in six of the nine *Ca. Accumulibacter* genomes included in this analysis.

Table V.2 – Main characteristics of the assemblies of the PacBio and/or Illumina sequencing data. The 16S rRNA gene sequences were searched only on the assemblies of individual samples.

Day	sample name	Total length [Mbp]	number of contigs	N50	max contig length [Mbp]	16S detected	proportion of complete 16S
71	d71_A_assembly_PacBio	69.3	1 485	65 019	3.46	24	0.96
	d71_A_metaspades_hybrid	196.6	19277	17044	1.62	53	0.45
	d71_A_metaspades	101.2	16031	9056	1.40	34	0.41
	d71_A_megahit	87.2	13515	9364	0.76	31	0.39
	d71_B_metaspades_hybrid	198.2	16098	25358	1.65	47	0.32
	d71_B_metaspades	133.2	18514	11346	0.83	32	0.31
	d71_B_megahit	111.6	16402	10353	0.53	27	0.33
322	d322_A_assembly_PacBio	29.8	548	76707	0.83	12	1
	d322_A_metaspades_hybrid	199.4	21309	16668	1.85	59	0.46
	d322_A_metaspades	135.2	19901	10910	1.40	35	0.34
	d322_A_megahit	117.1	17919	9223	0.96	32	0.34
	d322_B_metaspades_hybrid	170.2	18655	16016	1.37	50	0.44
	d322_B_metaspades	110.6	17446	9123	0.64	32	0.22
	d7322_B_megahit	93.9	15277	8482	0.40	18	0.33
427	d427_A_assembly_PacBio	75.6	1997	55511	3.50	29	0.86
	d427_A_metaspades_hybrid	284.4	29679	15754	2.41	65	0.40
	d427_A_metaspades	155.3	24142	10083	2.17	37	0.32
	d427_A_megahit	132.4	20549	9455	0.98	40	0.33
	d427_B_metaspades_hybrid	195.3	19350	16417	1.51	48	0.65
	d427_B_metaspades	102.8	15281	10381	0.64	21	0.71
	d427_B_megahit	90.3	14235	8538	0.40	19	0.63
740	d740_A_assembly_PacBio	123.7	1300	211 404	3.90	63	1
	d740_A_metaspades_hybrid	409.0	34726	21228	1.34	95	0.39
	d740_A_metaspades	197.3	25320	19743	0.77	48	0.44
	d740_A_megahit	167.3	21091	19430	0.73	53	0.19
	d740_B_metaspades_hybrid	325.2	26330	24755	0.91	74	0.35
	d740_B_metaspades	164.0	24953	9628	0.77	52	0.25
	d740_B_megahit	141.4	22170	9096	0.42	41	0.20
all	all_A_assembly_PacBio	284.8	4285	127558	3.96	-	-
	all_AB_metaspades	792.0	107422	12301	1.89	-	-
	all_AB_megahit	718.0	101854	11365	0.98	-	-



Table V.3 – Information collected on the reference genome sequences. The estimated proportions of completeness (compl.) and contamination (cont.), produced by CheckM and BUSCO analysis, the number of 16S rRNA gene sequences (16S) and the number of coding sequences (CDS) where collected following the same procedure as for PacBio contigs.

sample name	reference	total length	nb contigs	est. comp. [%] CheckM	est. cont. [%]	est. comp. [%] BUSCO	est. cont. [%]	16S	CDS
Accumulibacter_Ila_UW_1	[Martin et al., 2014]	5058518	1	99.8	0.2	98.0	0	2	4414
Accumulibacter_Ila_UW1_pAph01 <sup>a</sup>		167595	1	0	0	0	0	-	165
Accumulibacter_Ila_UW1_pAph02 <sup>a</sup>		42325	1	0	0	0	0	-	41
Accumulibacter_Ila_UW1_pAph03 <sup>a</sup>		37695	1	0	0	0	0	-	49
Accumulibacter_IA_BA93	[Skenerton et al., 2015]	4626262	85	100	0.3	96.6	0.7	1	4043
Accumulibacter_IC_BA92		4947936	218	93.3	1.49	90.5	1.4	-	4232
Accumulibacter_IIC_BA91		4322820	737	81.8	3.15	71.6	0.7	-	4371
Accumulibacter_IIC_SK01		4561076	53	93.7	1.2	91.9	0.7	-	4105
Accumulibacter_IIC_SK02		4359420	155	97.6	1.6	98.6	1.4	-	3874
Accumulibacter_IIF_BA94		3097332	299	79.4	8.2	81.8	5.4	-	3031
Accumulibacter_IIF_SK11		4700180	94	76.8	0.1	78.4	0	-	3946
Accumulibacter_IIF_SK12		4412715	61	92.6	0.1	93.9	0	-	3828
Dechloromonas_aromatica	[Coates et al., 2001]	4501104	1	99.6	0.3	98.6	0	4	4243
Dechloromonas_sp	[Parks et al., 2017]	3258300	1	99.8	0.2	98.6	0	3	3115
Zoogloea_sp	[Muller et al., 2017]	5810207	734	99.7	2.6	97.3	0.7	1	5165
Competibacter_denitrificans	[Albertsen et al., 2013]	4084870	120	96.5	2.4	98.0	0	1	3752
Thiothrix_lacustris	[Kyrpides et al., 2014]	3718676	50	99.6	0	98.0	0	1	3530
Dokdonella_koreensis	[Yoon et al., 2006]	4446619	1	99.2	0.6	95.3	0	2	3583
Rhodobacter_sp_LPB0142	[Kim and Yi, 2016]	3462926	1	99.4	0.5	95.3	0	3	3357
Rhodobacter_sp_LPB0142_pEJ01 <sup>a</sup>		264008	1	0	0	0	0	-	294
Rhodobacter_sp_LPB0142_pEJ02 <sup>a</sup>		29119	1	0	0	0	0	-	28
Rhodobacter_sp_LPB0142_pEJ03 <sup>a</sup>		24776	1	0	0	0	0	-	28
Defluviimonas_denitrificans	[Goeker et al., 2018]	4107985	23	99.7	0.4	95.9	0.7	1	4064
Ferruginibacter_sp	[Im, 2018]	4963791	49	99.5	0.5	93.2	0.7	1	4189
Niabella_soli	[Eisen et al., 2014]	4697343	1	100	0	94.6	0	2	3912
Flavobacterium_sp	[Bekkelund et al., 2018]	4935222	1	99.4	0.2	89.1	0	9	4474
Micropruina_glycogenica	[Mcilroy et al., 2018]	3844677	1	97.9	0.7	94.6	0	1	3561
Nitrospira_defluvii	[Luecker et al., 2010]	4317083	1	97.7	2.3	93.9	0	1	4112
Propionicimonas_paludicola	[Klenk et al., 2017]	3262476	1	95.2	0.4	94.6	0	2	2981
Saccharimonas_aalborgensis	[Albertsen et al., 2013]	1013781	1	67.1	0.9	71.6	0	1	1033
Tetrasphaera_elongata	[Kristiansen et al., 2013]	3174691	1	97	0.5	92.6	0	1	2963
Tetrasphaera_australiensis		4264973	1	88.3	0.7	87.2	0	1	4036

<sup>a</sup> plasmid

Table V.4 – Selection of MAG obtained after the grouping of the PacBio contigs with the estimated taxonomy, the coverage of the Illumina reads obtained with extraction A (cov A) and B (cov B), the number of contigs and the total length, the N50 statistic, the number of coding sequences detected, the estimated proportions of completeness (compl.) and contamination (cont.), produced by checkM and BUSCO analysis. The complete table including all the PacBio bins is available in electronic supplementary material (Table A.19)

MAG name	Taxonomy	cov A	cov B	nb contigs	total length	N50	CDS	compl.		cont.	
								[%]	CheckM	[%]	BUSCO
ACC001a_d71	<i>Ca. Accumulibacter</i> IIA	216.27	196.73	15	1779303	146808	1575	26.7	0	50	0
ACC001c_d71		114.76	79.04	5	291613	57656	254	0	0	2.7	0
ACC001_d322		145.99	35.79	5	376254	87279	383	15.5	0	10.1	0
ACC005_d71	<i>Ca. Accumulibacter</i> IA	65.65	85.71	8	3148374	360783	2863	43.1	0	43.2	0
ACC005_d322		74.12	21.75	41	3089875	92089	2812	67.6	0	62.8	1.4
ACC005a_d427		69.7	10.86	10	3279503	442800	2905	80.3	0.5	74.3	0
ACC005b_d427		66.66	9.79	2	280583	169015	238	1.7	0	0.7	0
ACC003_d71	<i>Ca. Accumulibacter</i>	40.93	20.05	34	3305177	125556	2924	61.7	0.2	64.2	0
ACC003_d322		124.98	54.71	12	5351590	637200	4626	98.1	0.4	98.0	0
ACC003_d427		126	22.18	16	5389451	549355	4612	97.6	1.3	82.4	1.4
ACC004_d322		59.22	15.77	4	123731	35245	111	0	0	0	0
ACC004_d427		36.93	18.69	3	4449509	3458617	3908	85.9	1.3	81.8	0.7
ACC007a_d740		98.42	11.16	12	4541703	2398789	3956	86.3	2.5	90.5	0
ACC007b_d740		110.4	20.27	24	5329261	300467	4663	98.3	5.6	97.3	1.4
ACC009_d71		51.95	22	2	79120	56142	87	0	0	0	0
ACC017_d427		68.85	10.59	2	597055	517084	511	0	0	7.4	0
COM001_d71	<i>Ca. Competibacter</i>	39.21	15.93	13	3483968	361028	3148	86.1	2.2	84.5	0.7
CPB001_d427	<i>CPB_C22&amp;32</i>	14.77	17.29	19	1302016	97627	1230	49.7	0.6	53.4	0
CYC001_d740	<i>CYCU-0281</i>	28.63	17.54	2	6279676	3903295	4996	98.8	1.1	80.4	0
CYC002_d740		9.23	3.3	37	2129442	62229	2395	34.5	0	16.2	0
CYT001a_d740	Cytophagaceae (f)	17.52	7.29	19	3104168	262741	3290	56.8	1.5	46.6	0
CYT001b_d740		20.31	8.67	4	1348657	417562	1384	8.3	0	20.9	0
CYT003_d740		24.18	6.09	15	4093089	494452	4051	94.7	0.9	85.1	0

MAG name	Taxonomy	cov A	cov B	nb contigs	total length	N50	CDS	compl. [%]		cont. [%]	
								CheckM	BUSCO		
DBO001_d71	DB1-14 (o)	25.65	9.5	44	1295302	33801	1623	66.8	5.6	45.3	0.7
DBO002_d740		24.25	10.34	4	2076620	1695390	2008	84.1	0	86.5	0.7
DEC001_d71	<i>Dechloromonas</i>	105.42	62.5	1	3463902	3463902	3297	98.5	0	98.6	0
DEF002_d740	<i>Defluviimonas</i>	13.36	15.06	8	3469995	1955590	3609	85.1	0.6	79.7	0
DOK002_d740	<i>Dokdonella</i>	15.49	12.99	1	3630515	3630515	3184	94.3	0.6	85.8	0
FER001_d71	<i>Ferruginibacter</i>	23.26	14.03	90	3497973	45586	4002	51.5	0	45.9	0
FLA001_d740	<i>Flavobacterium</i>	12.62	9.54	39	1631402	51364	2186	44.5	0.6	39.2	0
FLA002_d322		50.93	58.17	58	2236930	47081	2825	54.3	0	37.8	0
FLA003_d71		18.16	18.84	60	1479039	25705	2022	38	1.4	30.4	0
LED001_d740	<i>Leadbetterella</i>	11.15	11.31	3	211030	84008	250	2.5	0	18.9	0
LED002_d740		8.34	13	34	3471226	130401	4302	60.2	1.8	48.0	0
NGA002_d740	<i>Ca. Nitrotoga</i>	14.19	14.48	8	2620310	722071	2532	80.1	0	78.4	0
NIA001a_d740	<i>Niabella</i>	27.72	29.26	3	1390847	377394	1225	0	0	6.8	0
NIA001b_d740		32.6	33.97	7	4813256	1117204	4170	87.1	0	87.8	0
PRO001_d427	<i>Propioniciclava</i>	19.01	56.12	4	2866104	855992	2611	95.7	0.7	92.6	0
PRO010_d740		18.64	87	4	3945632	2205127	3640	91.6	1	80.4	0
PSE003_d427	<i>Pseudoxanthomonas</i>	5.44	3.66	35	2326379	84285	2150	78.3	0.3	71.6	0
QED001_d740	<i>QEDR3BF09</i>	15.7	8.56	10	4627430	1024482	4244	88.8	1.1	72.3	0
ROB004_d427	<i>Rhodobacter</i>	6.37	8.71	6	172983	36510	191	0	0	1.4	0
ROB006_d740		8.86	6.17	37	1838796	55787	2305	33.8	0	37.2	0
ROB007_d740		7.55	7.01	31	1204397	43329	1637	13.9	0	14.9	0
DEC002/ACCO11_d427	Rhodocyclaceae (f)	25.16	7.16	2	135574	100482	136	0	0	1.4	0
ROD002_d71		16.3	10.53	2	61080	39110	79	0	0	0.7	0
ROD005_d740		24.3	7.15	13	2264020	196540	2480	42.1	0	38.5	0
ROD006_d427		17.15	3.53	15	318164	22506	398	14	0	6.8	0
SBR001_d427	<i>Sbr-gs28</i>	26.84	33.35	1	3504986	3504986	3216	96.1	0.7	92.6	0
SBR001_d740		154.29	147.63	2	3468165	1927230	3178	96.6	0.7	93.2	0
SBR003_d322		51.29	99.46	3	149688	47474	146	8.3	0	4.7	0

## Chapter V. Metagenomic analysis of different AGS samples

MAG name	Taxonomy	cov A	cov B	nb contigs	total length	N50	CDS	compl. [%]		cont. [%]		
								CheckM	BUSCO			
SAC001_d740	Saccharibacteria (p)	23.07	23.81	2	819849	577863	872	57.7	0.01	62.2	0.7	
SAC002_d740		41.67	37.32	1	1000428	1000428	1043	67.3	0	68.9	0	
SAC003a_d740		65.94	45.25	3	905399	270444	1011	62.4	0	41.2	0	
SAC003b_d740		39.31	27.43	5	447457	87152	479	21	0.5	21.6	2	
SAC007_d427		38.93	36.35	5	387994	196152	406	22.4	0	20.3	0	
SAC010_d740		30.32	21.26	4	248823	72813	229	13.9	0	15.5	0	
SAC011_d740		8.42	8.71	6	394487	69343	508	17	0	8.8	0	
SAC012_d740		12.32	13.01	10	815914	87647	1103	56	0	26.4	0	
SAC013_d740		31.15	24.89	1	1058942	1058942	1098	63	2.1	68.2	0	
SAC014_d740		10.51	11.51	8	712823	116547	894	53.8	0	26.4	0	
SAC015_d740		32.55	23.82	1	871087	871087	901	65.8	0	73.6	0	
THI001_d322		<i>Thiothrix</i>	55.11	73.1	24	3586748	253133	3663	88.6	0.4	86.5	0
THI001a_d427			44.84	70.55	16	1132813	101029	1128	21.1	0.4	41.2	0
THI001b_d427			42.72	63.47	35	2261600	87700	2253	55.2	0	41.9	0
UNC_bin2_d71		Actinobacteria (c)	12.35	26.3	23	3388035	172733	3374	84.4	0	73.6	0
UNC_bin_21_d322		44.72	70.52	32	2574315	99435	2482	78.1	0.01	74.3	0.7	
ZOO001_d740	<i>Zoogloea</i>	16.76	3.06	8	2359090	375536	2166	46.2	0.01	44.6	0.7	
ZOO002_d740		15.19	3.36	7	1523590	230444	1380	48.3	0	43.2	0	
ZOO003_d740		14.46	2.83	2	320281	244459	317	0	0	2.7	0	

The contigs were grouped in MAG based on the tetranucleotide frequency, the overlapping of BUSCO and the coverage of Illumina reads. The MAG containing a 16S rRNA gene sequence were named according to the taxonomy of the best match for this sequence. A set of MAG selected based mainly on the length of the contigs containing the 16S rRNA gene sequences are presented in Table V.4 with the coverage for extraction A and B, the number of contigs and total length, the number of coding sequences detected by prokka and the estimated completeness and contamination percentages based on checkM and BUSCO analysis.

The completeness estimated by using CheckM was globally consistent with the one estimated by BUSCO. Yet, for 16 of the MAG, these completeness provided by the two softwares had more than 10 % difference. Since the MAG were constructed by using the information provided by BUSCO, the completeness and contamination of these genomes will be based on the CheckM analysis.

Over the four samples, 24 MAG have an estimated completeness above 70%. Six MAG related to the phylum Saccharibacteria have an estimated completeness between 56 and 67 %, this last value corresponds to the estimated completeness of the reference genome of the same phylum, *Saccharimonas aalborgensis*. Moreover, the number of coding sequences is around 1 000 for the reference and the biggest Saccharibacteria MAG. The chosen method may underestimate completeness of the genomes belonging to Saccharibacteria.

MAG corresponding to *Ca. Accumulibacter* IA (ACC005, ACC007 and ACC017), IIA (ACC001) and undetermined clades (ACC003, ACC004 and ACC009) were recovered. Two of them (ACC007b\_d740 and ACC003\_d322) have a length, a number of predicted coding sequences, and a number of complete BUSCO very close to those of the most complete *Ca. Accumulibacter* reference genomes. The estimated contamination percentage is higher for the MAG related to *Ca. Accumulibacter* than for the MAG related to other taxa. The estimated contamination is lower than 5.6 % and it is globally very low (< 1% for 80 % the MAG). This is partly because the overlapping of BUSCO was included as a criterion in the method for grouping the MAG.

Among the MAG for which no reference genomes of the same genus was found in the NCBI database, CYC001\_d740 related to CYCU-0281 is estimated to be 99 % complete with 1 % of contamination, PRO001\_d427 and PRO010\_d740 related to *Propioniciclava* are estimated to be 96 and 92 % complete, respectively, with 1 % of contamination, QED001\_d740 related to *QEDR3BF09* is estimated to be 89 % complete with 1 % of contamination, SBR001\_d427 and SBR001\_d740, related to the genus *Sbr-gs28* are estimated complete at 96 and 97 %, respectively with a contamination percentage of 1 %.

Moreover, several nearly complete MAG related to unclassified taxa were recovered : CYT003\_d740 corresponding to the Cythophagaceae family with a completeness estimated at 95 %, DBO002\_d740 with a completeness estimated at 84 %, several MAG related to Saccharibacteria for which the completeness estimation do not seem to be reliable, and two MAG related to the class of Actinobacteria which did not contain any 16S rRNA gene sequence but are estimated 84 and 78 % complete, respectively.

## Chapter V. Metagenomic analysis of different AGS samples

---

The completeness of the MAG does not seem to depend only on the coverage. For example, ACC001a has an estimated completeness of 27 %, even though its coverage with the extraction A is the highest of the main MAG. In comparison, DEC001 has twice less coverage but an estimated coverage of 99 % with a unique contig. Therefore other factors may influence the assembly of the PacBio long-reads.

The phylogenetic relation between the 16S rRNA gene sequences of the MAG and the reference genomes is presented in Figure V.3. The diversity of the MAG related to *Ca. Accumulibacter* was high with six different MAG from sample d71, collected when the AGS was treating simple synthetic wastewater, five different MAG in each of the sample d322 and d427 collected when the AGS was treating complex monomeric wastewater and only two different MAG from sample d740 collected when the sludge was treating complex polymeric wastewater. According to this phylogenetic tree, *Ca. Accumulibacter* IIA was present in sufficient abundance in the samples d71 and d322 for the recovery of a corresponding MAG but not anymore in sample from day d427 and d740.

A high number of MAG were recovered from the DNA extracted from sample d740. The very long DNA fragments recovered from this sample likely allowed to get numerous long sequences that were easily assembled into good quality contigs.

Several Saccharibacteria related MAG seem to be closely related to the reference genome *Saccharimonas aalborgensis*, in particular the MAG SAC002, SAC007 and SAC014. But without other reference genomes corresponding to this phylum, it is difficult to estimate precisely the phylogenetic relation between these MAG and *Saccharimonas aalborgensis*.

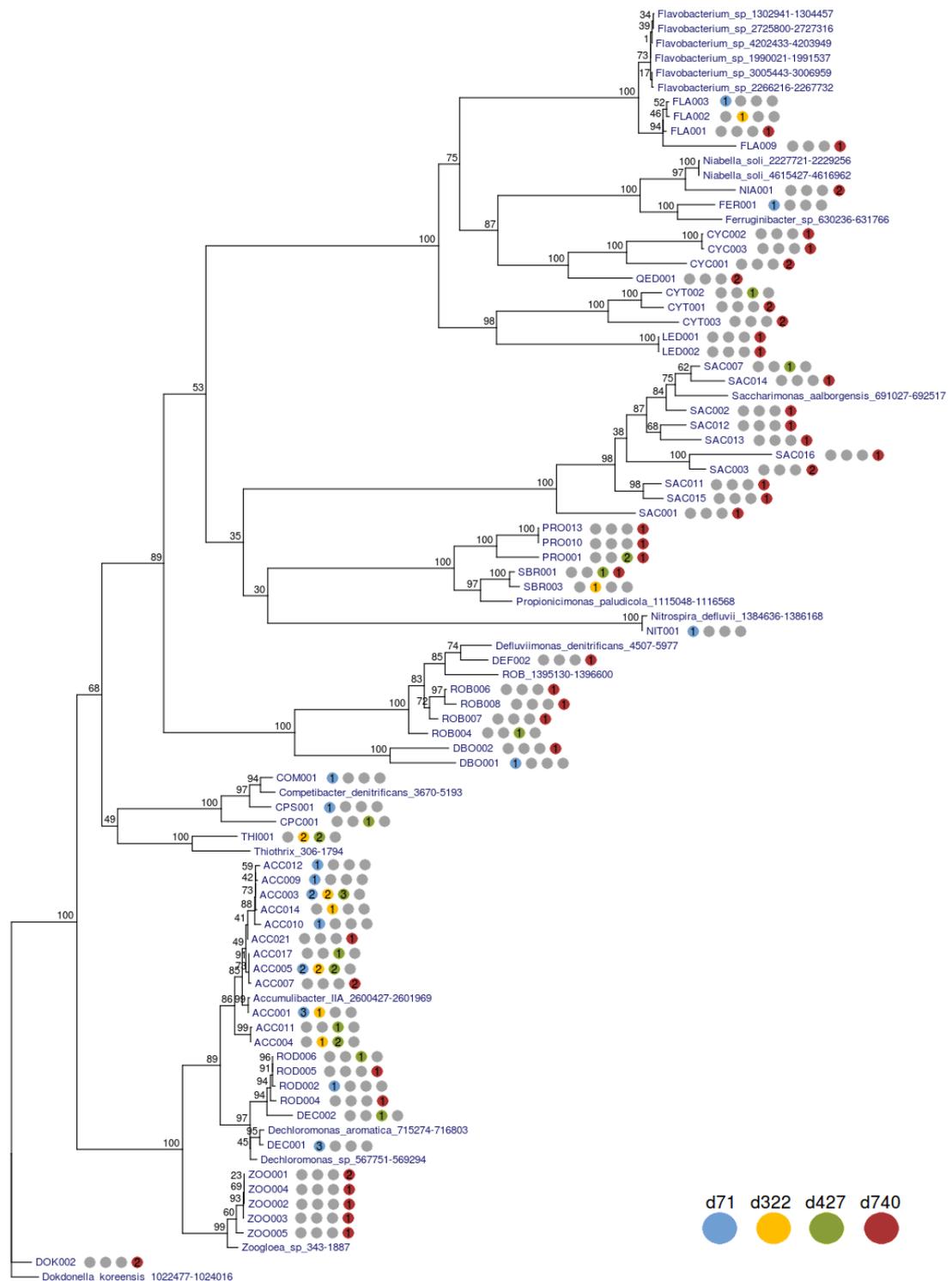


Figure V.3 – Phylogenetic tree of the 16S rRNA gene sequences extracted from PacBio contigs and selected reference genomes. The presence of contigs containing these sequences is indicated aside the MAG name with colored dots. The position of the 16S rRNA gene in the reference genome is indicated after its name.









## Chapter V. Metagenomic analysis of different AGS samples

---

The four assembly graphs give precious information about the relation between the contigs and the potential existence of variants. The assembly graph of the contigs corresponding to sample d71 is shown in Figure V.4. Approximately half of the contigs are part or linked to a *Ca. Accumulibacter* MAG. Bubbles and multiple paths are created in the graph when the assembler cannot decide for a path or another. A clear difference can be noticed between the contigs related to *Ca. Accumulibacter* and the other contigs (eg. *Dechloromonas*, *Ca. Competibacter*). They have more bubbles and multiple connection structures, giving evidence of the extreme microdiversity of the genus *Ca. Accumulibacter* in this sample. The existence of different strains with the same 16S rRNA gene is even possible since ACC001c and ACC001b seem to be two variants of a path. Moreover, ACC001a and ACC001c were not grouped together. With such a diversity, with parts of sequences that are very similar and others that are different from a strain to another, the assembler cannot create a complete genome, even if the coverage is high.

The sequences of the 'complex' structure were blasted on the NCBI online blastn to have a first approximation of the origin of these sequences. Lots of them show a short but 100 % match with a sequence in the *Ca. Accumulibacter* IIA UW1 genome located between bp 2 158 388 and 2 148 350, just before a region coding for a tRNA or between 2 566 281 and 2 566 253, in the repeat region next to the CRISPR-associated protein Cas1.

In comparison to the sample d71, the MAG ACC003 is more complete in the graphs of sample d322 (figure V.5). The contigs from the MAG ACC001 are not predominant in this sample and this may have allowed a easier assembly of the contigs from MAG ACC003.

The graph of assembly d427 (Figure V.6) shows that the sequences from the MAG ACC003 and ACC005 are very similar on numerous places. On the other hand, the MAG ACC004 is clearly separated from these two samples. The phylogenetic tree of the MAG (Figure V.3) placed ACC004 little aside the other *Ca. Accumulibacter*-related MAG. It is maybe different enough so that the assembler can distinguish its sequences from those of the others. The MAG PRO001 and SBR001 related to *Propioniciclava* and *Sbr-gs28* have few or no ambiguities.

In the assembly graph from sample d740 (Figure V.7), the configuration of the contigs related to *Ca. Accumulibacter* suggest that the MAG ACC007a and ACC007b associated with the same 16S rRNA gene sequence are from two (or more) very similar strains. The sequences related to the phylum Saccharibacteria are also intertwined suggesting that some part to the sequences of the different MAG are very similar to one another.

The four assembly graphs show that the majority of the contigs bigger than 100 kbp have been grouped in an MAG and related to a 16S rRNA gene sequence.

### V.3.5 Denitrification-related genes in MAG and reference genomes

The genes involved in denitrification were searched in all the PacBio contigs and the reference genome by using HMM models of these genes. Table V.5 shows the number of positive

matches for the main MAG. The complete table is available in the electronic supplementary material (Table A.22).

Table V.5 – Positive match with the hmm models for genes involved in denitrification found in the main MAG and selected reference genomes. The columns for NirS, NorB and NorC were obtained directly from the prokka annotation of the MAG by searching for the EC numbers 1.7.2.1, 1.7.2.5 and the gene name 'norC' respectively. The abbreviation nit. r. stands for nitrite reductase (copper form).

gene	narG	narH	narI	narJ	ntxB	ntxC	NapA	nit r.	NirS	NirB	NirD	NorB	NorC	NosZ	NosD	NosZ_GP
<i>Ca. Accumulibacter</i> BA-93 (IA)	0	0	0	0	0	0	1	0	4	1	1	1	0	0	1	1
ACC005_d71 (43 %)	0	0	0	0	0	0	0	0	4	1	1	1	0	0	1	1
ACC005a_d427 (74 %)	0	0	0	0	0	0	1	0	4	1	1	0	0	0	0	0
ACC007a_d740 (91 %)	0	0	0	0	0	0	1	0	1	1	1	0	0	0	1	1
ACC007b_d740 (97 %)	0	0	0	0	0	0	1	0	3	1	1	1	0	0	2	1
ACC003_d322, ACC003_d427 (98 and 82 %)	1	1	1	1	0	0	0	0	2	1	1	0	0	0	0	0
<i>Ca. Accumulibacter</i> BA-92 (IC)	0	0	0	0	0	0	1	0	3	1	1	0	0	0	2	1
<i>Ca. Accumulibacter</i> IIA UW1	0	0	0	0	0	0	1	0	2	1	1	0	0	0	2	1
ACC001a_d71 (50%)	0	0	0	0	0	0	0	0	2	0	0	1	0	0	1	1
<i>Ca. Accumulibacter</i> SK-01 (IIC)	1	1	1	1	0	0	0	0	2	1	1	0	0	0	0	0
<i>Ca. Accumulibacter</i> SK-12 (IIF)	0	0	0	0	1	1	0	0	2	0	1	0	0	0	1	0
ACC004_d427 (82 %)	1	1	1	1	0	0	0	0	2	0	0	1	1	0	2	1
<i>Dechloromonas</i> aromatica RCB	0	0	0	0	1	1	1	0	3	1	1	0	0	0	2	2
<i>Dechloromonas</i> sp. HYN0024	0	0	0	0	1	0	1	0	2	2	2	0	0	0	1	1
DEC001_d71 (99 %)	0	0	0	0	1	1	2	0	2	1	1	1	1	0	2	1
DBO002_d740 (86 %)	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
<i>Zoogloea</i> sp. LCSB751	0	0	0	0	1	1	1	0	2	2	2	0	0	1	1	0
ZOO001_d740 (45 %)	0	0	0	0	0	1	1	0	1	0	1	3	1	0	0	0
ZOO002_d740 (43 %)	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0
ROD005_d740 (39 %)	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1
<i>Niabella</i> soli	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1
NIA001b_d740 (88 %)	0	0	0	0	0	0	0	1	1	1	1	0	0	0	1	0
<i>Ca. Competibacter</i> denitrificans	1	1	1	1	0	0	0	0	3	0	0	0	0	1	1	0
COM001_d71 (84 %)	0	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0
<i>Thiothrix</i> lacustris	1	1	1	1	0	1	0	0	0	1	1	0	0	0	0	0
THI001_d322 (86 %)	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
<i>Rhodobacter</i> sp. LPB0142	0	0	1	1	0	1	0	0	0	1	1	0	0	0	1	1
ROB006_d740 (37 %)	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0
<i>Deftuivimonas</i> denitrificans	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	0
DEF002_d740 (80 %)	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	0
<i>Ferruginibacter</i> sp. BO-59	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1
FER001_d71 (46 %)	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0
SBR001_d427, SBR001_d740 (93 %)	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0
PRO010_d740 (80 %)	1	1	1	1	0	0	0	0	1	1	1	0	0	0	0	0
<i>Flavobacterium</i> sp. 140616W15	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
FLA003_d71 (30 %)	0	0	0	0	0	0	0	1	1	0	0	3	2	0	1	0
<i>Dokdonella</i> koreensis	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
DOK002_d740 (86 %)	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
CYC001_d740 (80 %)	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1
LED002_d740 (48 %)	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1

The chosen HMM families corresponding to the nitric oxide reductase (COX1, Pfam) and the nitrite reductase NirS are probably a little too general and match with different cytochrome c oxidases that may not be involved to nitrite reduction. Therefore a research of the NirS, NorB and NorC proteins were searched in the prokka annotations of the reference genomes and the MAG.

The same groups of genes were found in *Ca. Accumulibacter* IA and the related MAG (ACC005 and ACC007), with few exceptions potentially due to the incompleteness of the genomes. These genomes contain the large subunit of the periplasmic nitrate reductase (NapA), several respiratory nitrite reductase (NirS) and a assimilatory nitrite reductase (NirBD), suggesting a putative capability of the corresponding organism to transform nitrate into nitrite and nitrite into nitric oxide. The MAG ACC005\_d71 and ACC007b\_d740 and the reference BA-93 have an homologue of the nitric oxide reductase (NorB) and two members of the nitrous oxide reductase complex potentially conferring the capability to reduce nitric oxide into nitrous

oxide and to gaseous nitrogen.

The complete nitrate reductase complex, and therefore the putative capability to reduce nitrate, was found in the MAG ACC004, ACC003 for d322 and d427 and in the reference genome *Ca. Accumulibacter* IIC. ACC004 and *Ca. Accumulibacter* IIC also contain a putative respiratory nitrite reductase and homologues of the nitric and nitrous oxide reductase. The potential transform nitrate into gaseous nitrogen is therefore present in these genomes.

The putative capability to reduce nitrate was found in other MAG from various phylogeny. For the order DB1-14 (DBO002), *Thiothrix* (THI001), *Deftuviimonas* (DEF002), *Sbr-gs28* (SBR001) and *Propioniciclava* (PRO010) it was present under the form of intracellular nitrate reductase and for *Dechloromonas* (DEC001), *Zoogloea* (ZOO001), *Rhodobacter* (ROD005 and ROB004) under the form of periplasmic nitrate reductase. The putative capability to reduce nitrite was also found in MAG of various taxa, with dissimilatory or assimilatory nitrite reductase homologues: *Dechloromonas* (DEC001), *Spb280* (SPB001), *Niabella* (NIA001), *Ferruginibacter* (FER001) and *Propioniciclava* (PRO010), *Zoogloea* (ZOO001), *Rhodocyclaceae* (ROD005), *Rhodobacter* (ROB006), *Sbr-gs28* (SBR001) and *Flavobacterium* (FLA003).

The NosZ and NOSZ\_GP are two distinct nitrous-oxide reductases, NosZ was first characterized in *Pseudomonas* whereas NosZ\_GP was characterized in *Geobacillus thermodenitrificans*. Homologues of the first type (NosZ) were found in the MAG from various phylogenetic origins: *Ca. Accumulibacter* (ACC017 and ACC004\_d427), in particular the clades IA (ACC005\_d71 and ACC007), and IIA (ACC001a\_d71), *Dechloromonas* (DEC001), *Rhodocyclaceae* (ROD002 and ROD005), CYCU-0281 (CYC001) and *Leadbetterella* (LED002). Homologues of the second type (NosZ\_GP) were found in the MAG of *Zoogloea* (ZOO002), *Rhodobacter* (ROB007) and *Deftuviimonas* (DEF002). The distribution of these two putative enzymes does not seem completely related to the phylogeny, since different types were found in bacteria of the same class (e.g. *Dechloromonas* and *Zoogloea*).

Based on these homologous protein sequences, ACC004, ACC007b\_d740 and DEC001\_d71, are potentially capable of performing complete denitrification from nitrate and ACC005\_d71 and ACC001a\_d71 from nitrite. For other MAG, the genes that would complete the pathway may be missing due to the incompleteness of the genomes (eg. FLA003\_d71).

### V.3.6 Phosphate metabolism-related genes in MAG and reference genomes

HMM models corresponding to gene families involved in phosphate metabolism, in particular the ones involved in the production (ppk1, ppk2) and degradation of poly-P (ppx), in phosphate transport (Pst, Pit, ABC tr.) and regulators (PstB, PhoU), were searched in the PacBio contigs and the reference genomes. These homologues were found in MAG related to various taxa (Table V.6) and no clear pattern seems to distinguish the known PAO from the other bacteria. Even the Pit HMM model was found in numerous bacteria, including the GAO *Micropruina glycogenica* or *Zoogloea*. Yet, this model was maybe a little too general

as it captured a lot of proteins annotated as sulfate permease CysP. Therefore a model was built from Pit sequences from *Ca. Accumulibacter* and *Tetrasphaera*, but even this HMM model matched with several sequences from non-PAO bacteria. The phylogenetic tree of the sequences matching this home-HMM pit model (Figure V.8) shows that two clusters of similar sequences were found mostly in the MAG and reference genomes of PAO (*Ca. Accumulibacter*, *Tetrasphaera*) or putative PAO (*Dechloromonas*). It is therefore surprising that a sequence from an MAG related to the genus *Spb280* (SPB001) was also present in one of these two clusters. This genus belongs to the same order as *Ca. Accumulibacter*, which could explain that their sequences cluster together.

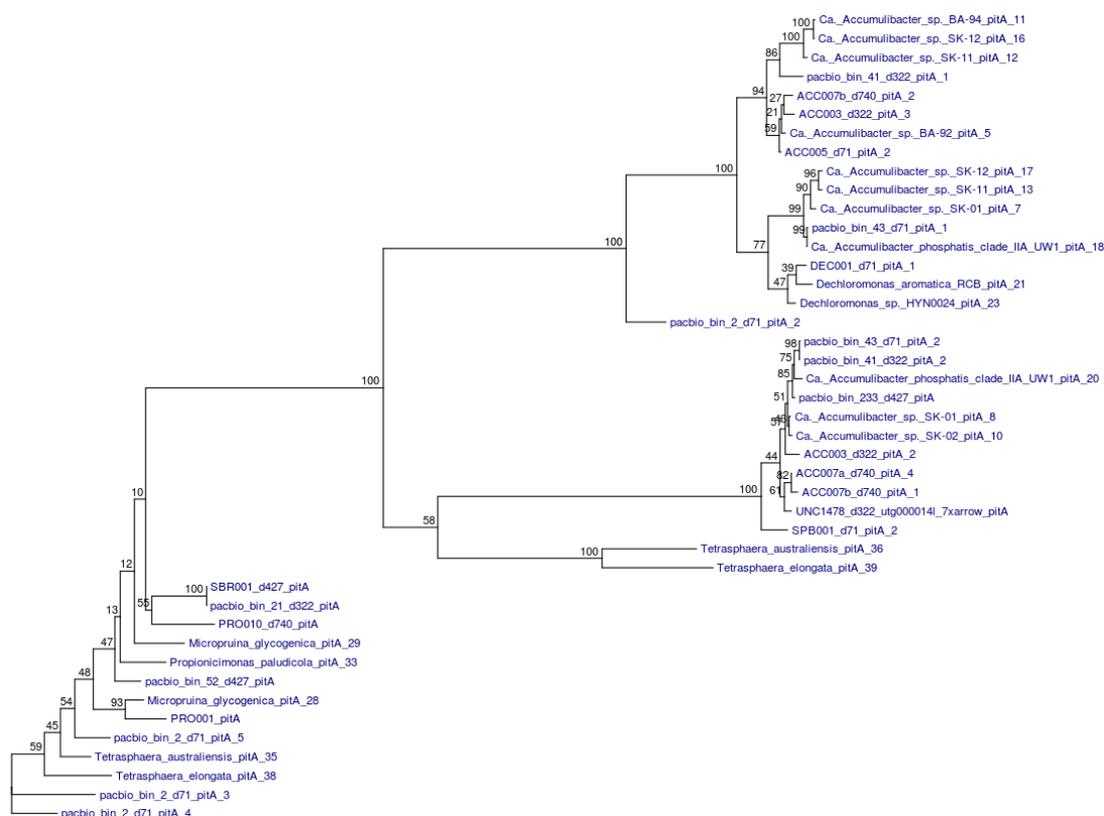


Figure V.8 – Phylogenetic tree of the sequences matching the HMM model constructed from Pit sequences of *Ca. Accumulibacter* and *Tetrasphaera*. The tree is based on an alignment of 369 amino acids.

A tendency emerging from these results is that the MAG related to *Ca. Accumulibacter* have more copies of homologous sequences to phosphate transport system regulatory protein (PhoU). Yet, only one copy was found in each of the *Tetrasphaera* reference genomes, the presence of multiple PhoU homologous genes is therefore not a necessary criterion for PAO metabolism. Yet, among the HMM model tested, it is the only that was found in *Tetrasphaera* but not in *Micropruina* or in the MAG SBR001. They were also not found in the MAG or the reference genomes of the GAO from Competibacteraceae : *Ca. Competibacter*, COM001, CPC001.

## Chapter V. Metagenomic analysis of different AGS samples

---

Phylogenetic trees were constructed with the sequences having a positive match with ppk1 (Figure V.9A) and ppk2 (Figure V.9B) in order to investigate their degree of homology. The sequences corresponding to the ppk1 seem to be distributed according to the phylogeny of the genome they belong to. Moreover, the relation between the ppk1 of the MAG related to *Ca. Accumulibacter* correspond to the classification based on 16S rRNA gene sequences with the ppk1 related sequence from ACC005 and ACC007 being close to the one of *Ca. Accumulibacter*-BA-93 (clade IA). The alignment of the sequences matching ppk2 created three big clusters. The two first clusters correspond to the two parts of a single protein sequence with two homologous domains. The two corresponding subtrees are similar to one another and contain sequences from MAG and reference genomes of *Ca. Accumulibacter* as well as *Competibacteraceae*. The third big cluster groups sequences that are homologous on 260 amino acids, while the whole ppk2 has around 500 amino acids. These sequences may not have the same function as ppk2. As for ppk1, the ppk2 homologues cluster similarly to the 16S rRNA gene sequences, the sequences of the MAG ACC005 are even exactly the same for the two parts of one of the ppk2.

Table V.6 – Positive match with the hmm models for genes involved in the phosphorus metabolism found in the main MAG and selected reference genomes. The complete table can be found in electronic supplementary material (Table A.21)

hmm	ppk1	ppk2	ppx	PstA	PstB	PstC	PstS	PstS_2	Pit	ABC tr.	ABC tr.	PhoU	phoR
<i>Ca. Accumulibacter</i> sp. BA-93 (IA)	1	2	0	1	1	1	0	1	4	1	0	6	1
ACC005_d71 (43 %)	1	1	0	1	1	0	0	1	3	1	0	5	1
ACC005_d322 (63 %)	1	2	0	0	1	0	0	0	3	1	0	4	1
ACC005a_d427 (74 %)	1	2	0	1	1	1	0	1	4	0	0	6	1
ACC007a_d740 (91 %)	1	2	0	2	1	1	1	1	4	1	1	5	1
ACC007b_d740 (97 %)	1	2	0	3	1	3	1	1	6	1	1	7	1
ACC003_d71 (64 %)	0	2	0	2	1	2	1	1	2	0	1	1	1
ACC003_d322, ACC003_d427 (98, 82 %)	1	3	0	2	1	2	1	1	4	1	1	5	1
<i>Ca. Accumulibacter</i> BA-92 (IC)	1	2	0	2	1	2	1	1	3	1	1	3	1
<i>Ca. Accumulibacter</i> clade IIA UW-1	1	2	0	2	1	2	0	3	4	1	1	5	1
pacbio_bin_43_d71 (37 %)	1	2	0	2	1	2	2	2	4	1	2	6	1
ACC001a_d71 (50 %)	0	0	0	0	0	0	0	0	1	1	0	1	0
<i>Ca. Accumulibacter</i> SK-01 (IIC)	1	1	0	1	1	1	1	0	3	2	1	4	1
<i>Ca. Accumulibacter</i> SK-12 (IIF)	1	2	0	1	1	1	0	2	4	0	1	6	1
ACC004_d427 (82 %)	1	0	1	2	1	2	0	2	1	0	2	1	1
<i>Dechloromonas</i> aromatica	1	1	1	1	1	1	1	2	2	2	1	2	1
DEC001_d71 (99 %)	1	1	1	1	1	1	0	1	2	1	1	2	1
<i>Zoogloea</i> sp.	1	1	1	1	1	1	1	0	1	2	1	1	1
ZOO001_d740 (45 %)	0	1	0	0	0	0	0	0	0	1	0	0	1
ZOO002_d740 (43 %)	1	0	1	0	1	1	1	0	1	0	1	1	0
<i>Ca. Competibacter</i> denitrificans	2	2	1	2	2	2	2	0	0	0	1	0	1
COM001_d71 (84 %)	1	1	0	1	2	1	0	0	0	1	1	0	1
CPC001_d427 (53 %)	1	1	0	1	0	1	2	0	0	0	1	0	0
<i>Thiothrix</i> lacustris	1	0	1	1	1	1	0	0	1	0	1	1	1
THI001_d322 (86 %)	1	1	0	1	0	1	0	0	1	0	1	2	1
THI001b_d427 (42 %)	0	1	0	1	1	1	0	0	1	0	1	1	1
<i>Dokdonella</i> koreensis	1	0	1	1	2	1	1	0	1	0	1	1	1
DOK002_d740 (86 %)	1	0	1	1	1	1	1	0	1	0	1	0	1
<i>Defluviimonas</i> denitrificans	1	0	0	1	1	1	0	0	1	1	1	0	0
DEF002_d740 (80 %)	1	0	0	1	1	1	0	0	0	1	1	0	0
<i>Tetrasphaera</i> elongata	1	1	0	1	1	1	1	0	2	0	1	1	0
<i>Tetrasphaera</i> australiensis	1	1	0	1	1	1	0	1	2	0	1	1	0
<i>Micropruina</i> glycogenica	1	1	0	1	1	1	1	0	2	0	1	0	0
SBR001_d427, SBR001_d740 (93 %)	1	1	0	1	1	1	1	0	1	0	1	0	0
PRO001_d427 (93 %)	0	1	0	1	1	1	0	0	1	0	1	0	0
PRO010_d740 (80 %)	1	1	0	1	1	1	1	0	1	0	1	0	0
QED001_d740	1	1	0	1	1	1	0	0	2	0	1	1	0
CYC001_d740 (80 %)	1	1	0	0	0	0	0	0	3	0	0	2	0
CYT001a_d740 (47 %)	1	1	0	0	0	0	0	0	3	0	0	0	0
CYT003_d740 (85 %)	1	1	0	0	0	0	0	0	1	0	0	0	0
LED002_d740 (48 %)	1	0	0	0	0	0	0	0	2	0	0	0	0
<i>Niabella</i> soli	2	1	0	0	0	0	0	0	1	0	0	1	0
NIA001b_d740 (88 %)	1	1	0	0	0	0	0	0	1	0	0	1	0
<i>Ferruginibacter</i> sp. BO-59	1	0	0	1	1	2	1	0	1	0	1	1	0
FER001_d71	1	0	0	0	0	0	0	0	1	0	0	1	0



## V.3.7 Analysis of the most complete MAG

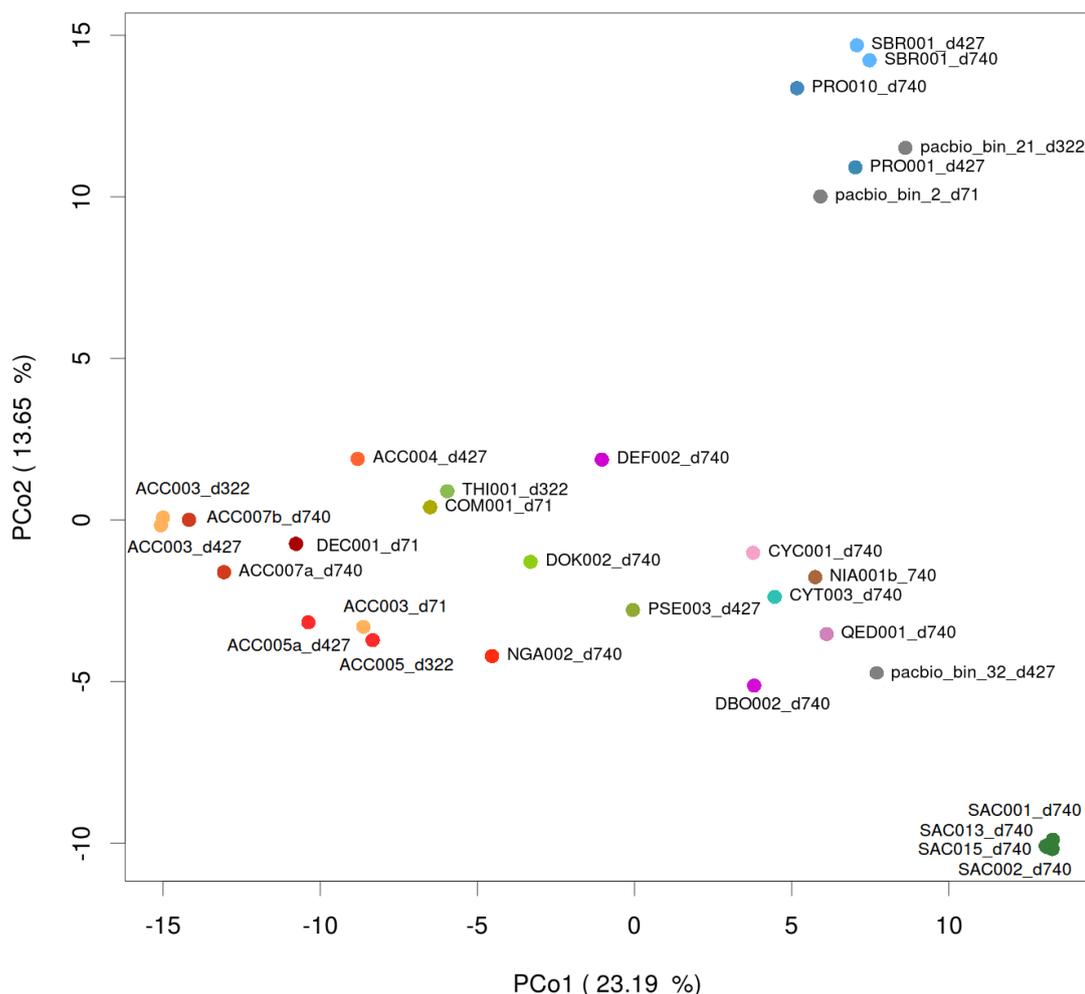


Figure V.10 – Principal component analysis (PCA) of the presence/absence of the COG in the main MAG.

As a first approach of the potential functions contained in the MAG with a completeness higher or equal to 50 %, the presence/absence of the different COG was analyzed through a principal component analysis PCA (Figure V.10). The MAG recovered from different samples but carrying the same 16S rRNA gene sequence (SBR001\_d427 and SBR001\_d740, ACC003\_d322 and ACC003\_d427, or ACC005a\_d427 and ACC005\_d322) are grouped together, meaning that they have very similar COG. The MAG related to Saccharibacteria are closely grouped together, suggesting that they have similar COG, but also that they are very different from the rest of the MAG. The MAG related to Actinobacteria are clearly separated from the others, in particular in the second axis. On the first axis, these MAG are similar to the cluster of Saccharibacteria related MAG. Globally the MAG are clustered according to their phylogeny, with the MAG of the order of Betaproteobacteriales clustering together. The MAG ACC004 seems different from

the other *Ca. Accumulibacter*-related MAG.

The clustering of the COG related to the inorganic ion transport and metabolism in the main MAG (Figure V.11), reveal that the MAG related to Saccharibacteria can be distinguished from the others by the presence of very few of these COG. This can be due to the apparently small size of their genomes or because their protein sequences are very different from the bacteria that were used to build the COG database. A group of COG related to the transport of metals is present mostly in the MAG related to *Ca. Accumulibacter* and to a lesser extent the MAG related to the Betaproteobacteriales. In the COG present particularly in the Actinobacteria, appear an ABC-type dipeptide/oligopeptide/nickel transport system and a superoxyde dismutase.

In this group of COG also, very few were retrieved in the Saccharibacteria related MAG. COG related to chemotaxis were mostly found in the MAG related to Proteobacteria and in particular to Betaproteobacteriales. The MAG SBR001\_d427 and SBR001\_d740 have exactly the same pattern of COG, and in general, the MAG carrying the same 16S rRNA gene sequences have similar patterns for the COG, giving evidences for the good quality of the assembly and grouping of these MAG.

Some of the MAG recovered in this study had estimated completeness higher than 80 % and have to date no complete genome available in the NCBI database. A brief and very incomplete survey of these MAG is presented in the next paragraphs. The genomes and annotations performed with prokka are available in electronic supplementary material A.4.1 (Folder A.1)

### **Sbr-gs28**

Two MAG related to the genus *Sbr-gs28*, from the class Actinobacteria, with an estimated completeness of 96 and 97 %, respectively, and an estimated percentage of contamination of 0 %, were obtained. 3216 coding sequences were detected in the SBR001\_d427, with 1881 having homologies with know protein sequences. Sequences homologous to genes involved in biofilm formation were detected in these genomes: a biofilm regulatory protein A (brpA, FLBBEHNN\_00399), a protein involved in the ESX-5 secretion system (eccC5, FLBBEHNN\_00033), a putative N-acetylglucosaminyltransferase involved in the formation of a biofilm adhesin polysaccharide (pgaC, FLBBEHNN\_01561), a protein for the synthesis initiation of the extracellular polymeric substance (EPS) xanthan (gumD, FLBBEHNN\_00568) and several other biomilm formation related proteins. SBR001 may therefore take part in the biofilm formation in AGS.

Several genes putatively involved in quorum-sensing were also found: an alkanal monooxygenase alpha chain (LuxA, FLBBEHNN\_00360), a (4S)-4-hydroxy-5-phosphonooxypentane-2,3-dione isomerase, taking part in the degradation of phospho-AI-2 (lsrG, FLBBEHNN\_00483), two homoserinelactone efflux protein (rhtB, FLBBEHNN\_01041 and FLBBEHNN\_03100). This suggests that the expression of specific genes may be regulated by quorum-sensing in SBR001.

Genes putatively coding for proteins involved in the toxin/anti-toxin system were found :

for example, the toxins gene *higB* (FLBBEHNN\_00361) and *mazF* (FLBBEHNN\_00637) and the anti-toxins such as *higA* (FLBBEHNN\_00362), *mazE3* (FLBBEHNN\_00636) *relB* and *relJ* (FLBBEHNN\_02598 and FLBBEHNN\_01018). A CRISPR-Cas loci was also detected with repeat regions and putative CRISPR-associated proteins (endoribonuclease, nuclease/helicase).

The presence of a putative superoxide dismutase (*sodA*, FLBBEHNN\_00949) and other putative proteins involved in the protection against oxidative stress : a peroxiredoxin (*bpc*, FLBBEHNN\_02125, not found in SBR001\_d740) and a thiol peroxidase (*tpx*, FLBBEHNN\_01426), suggest that SBR001 can survived in an aerobic environment.

Numerous sequences possibly coding for proteins involved in fermentation were detected in SBR001: an adenylate kinase, (*adk*,FLBBEHNN\_00938), a putative zinc-binding alcohol dehydrogenase (FLBBEHNN\_00278), an acetate kinase (*ackA*, FLBBEHNN\_00343), a L-lactate dehydrogenase (*ldhA*, FLBBEHNN\_00475), a succinyl-CoA:acetate CoA-transferase (FLBBEHNN\_00557), diverse components of a pyruvate dehydrogenase (*PdhC*, FLBBEHNN\_00648, FLBBEHNN\_02133), (*aceE*, FLBBEHNN\_01591) and (*aceF*,FLBBEHNN\_01058), an NADP-dependent alcohol dehydrogenase C1 (*adhC1*,FLBBEHNN\_01421), two classical alcohol dehydrogenases ( FLBBEHNN\_01942 and FLBBEHNN\_02107), a complex of methylmalonyl-CoA carboxyltransferase (FLBBEHNN\_00688, \_00689 and \_00691), and several other. This may indicate a that this organism can ferment several organic compounds.

Hydrolysis is supported by the presence of diverse proteases : a neopullulanase (*tvaII*, FLBBEHNN\_00482), an extracellular exo-alpha-L-arabinofuranosidase (*abfB*,FLBBEHNN\_00694) or a minor extracellular protease (*vpr*, FLBBEHNN\_00791).

Genes putatively coding for porteiins implicated in the formation of glycogen are present in these MAG : two 1,4-alpha-glucan branching enzyme (*glgB*, FLBBEHNN\_00163) and FLBBEHNN\_01831) a glucose-1-phosphate adenylyltransferase (*glgC*,FLBBEHNN\_00965), and other proteins of the *glg* operon (*glgE2*,*glgX*, *glgP*). A putative glycogen accumulation regulator sequence (*garA*, FLBBEHNN\_01896) was also detected.

Sequences coding for genes implicated in denitrification were also found, as reported in section V.3.5, as well as putative nitrate/nitrite sensors and transporters.

Homologues of genes implicated in the regulation of sporulation were detected : a sporulation transcription regulator (*WhiA*,FLBBEHNN\_01172), sporulation initiation inhibitor protein (*soj*,FLBBEHNN\_00075), with diverse sporulation proteins (*spoVK*,FLBBEHNN\_01637), (*paiB*,FLBBEHNN\_01744). These MAG contain also several sequences homologous to genes involved in betalactame and other antibiotic resistance (*MacA*, *MacB*, tetracycline repressor).

### Unclassified DB1-14

An MAG related to the order DB1-14 (DBO002), from the class Alphaproteobacteria, with an estimated completeness of 84 % and an estimated percentage of contamination of 0 % was obtained. 2008 coding sequences were detected in the DBO002\_d740 (DBO002), with 1138 having homologies with know protein sequences.

Several sequences homologous to genes involved in biofilm regulation (bigR, CPBMCMNO\_01044, kinB, CPBMCMNO\_01871) and formation through biofilm exopolysaccharide production (icaB, CPBMCMNO\_01899, epsE CPBMCMNO\_01938, epsJ CPBMCMNO\_01052), in particular in the biosynthesis of alginate (algA, CPBMCMNO\_01148 and CPBMCMNO\_01929 and algC, CPBMCMNO\_01904) were found in DBO002. Two sequences similar to the genes encoding autoinducer 2 sensor kinase/phosphatase (LuxQ, CPBMCMNO\_00216, CPBMCMNO\_01636) were also found. Few sequences homologous to genes components of the toxin (higB, CPBMCMNO\_00421)/antitoxin (higA, CPBMCMNO\_00422) system are present in this MAG. A putative CRISPR-Cas system was detected with repeat regions and potential CRISPR-associated endonuclease sequences.

Homologous sequences to genes encoding for proteins involved in oxidative stress protection were detected : two alkyl hydroperoxide reductase (ahpD, CPBMCMNO\_01139 and C, CPBMCMNO\_01140), a superoxide dismutase (sodB, CPBMCMNO\_00590), a putative peroxiredoxin (bcp, CPBMCMNO\_01314) and an organic hyperoxide resistance protein (ohrB, CPBMCMNO\_00069).

The few sequences potentially coding for genes potentially involved in fermentation pathways, such as a L-lactate dehydrogenase (ldh, CPBMCMNO\_01689), a formate acetyltransferase (pflB, CPBMCMNO\_00232) or a dihydrolipoyl dehydrogenase (lpd, CPBMCMNO\_00407) may not be sufficient to support the hypothesis of a fermentative metabolism in DBO002. Two sequences similar to genes coding for aerotaxis receptors are present in ODB002, suggesting a putative preference for aerobic environments.

A unique sequence putatively involved in glycogen formation, a glycogen synthase (CPBMCMNO\_01936), was found. PHA formation, on the other hand is supported by the presence of homologous sequences of the whole poly(3-hydroxyalkanoate) polymerase operon (phaABC, CPBMCMNO\_00897 to CPBMCMNO\_00899), an acetate kinase (ackA, CPBMCMNO\_00071) and two acetyl-CoA synthetase (acsA, CPBMCMNO\_01026 and CPBMCMNO\_01639).

Numerous sequences homologous to genes involved in chemotaxis and flagel fromation were fond in DBO002. Sequences putatively encoding antibiotic resistances in the form of penicillin-binding proteins, multidrug export proteins, a beta-lactamase and macrolide export proteins were also detected in DBO002.

### Unclassified Saccharibacteria

Several MAG related to the phylum Saccharibacteria were obtained with estimated completeness of 63 % and an estimated percentage of contamination of 0 %. The size of the genome and the number of coding sequences in these MAG being close to the complete genome of *Saccharimonas aalborgensis*, the completeness may be underestimated for the genomes of this phylum. 1098 coding sequences were detected in the SAC013\_d740 (SAC013), with 502 having homologies with known protein sequences.

Two sequences homologous to genes coding for the toxin (relE, CMHCJOKK\_00221 and relE3, CMHCJOKK\_01125) components of the toxin/antitoxin system are present in SAC013. No CRISPR system component was detected. A putative peroxiredoxin (bcp, CMHCJOKK\_00684) and a superoxide dismutase (sodA, CMHCJOKK\_00321) homologues seem to indicate a tolerance of SAC013 for aerobic environments. A sequence similar to an anaerobic regulatory protein (fnr, CMHCJOKK\_00594) may encode for a protein helping SAC013 to adapt its metabolism to an environment alternating between presence and absence of oxygen.

Sequences potentially coding for proteins involved in the catabolism of amino acids such as a periplasmic L-asparaginase (ansB, CMHCJOKK\_01108) or a L-threonine dehydratase (tdcB, CMHCJOKK\_00028) were detected. Homologues to glycogen synthases (CMHCJOKK\_00488, CMHCJOKK\_00619 and CMHCJOKK\_01072) a 1,4-alpha-glucan branching enzyme (glgB, CMHCJOKK\_00935) and an alpha-maltose-1-phosphate synthase potentially involved in glycogen pathway (glgM, CMHCJOKK\_01125) may indicate that SAC013 can form glycogen reserves. Sequences potentially encoding for proteins involved in sporulation (2) and antibiotic resistance (9) were detected.

### V.3.8 Other MAG

The functional annotation of the MAG with an estimated completeness higher than 80 % according to the BUSCO analysis, was briefly skimmed through. Sequences homologous to genes coding for toxin/antitoxin, resistance to oxidative stress and antibiotic resistance were found in all the considered MAG. A putative CRISPR-Cas system was detected in the MAG related to *Ca. Accumulibacter* (ACC003, ACC004, ACC007b), *Ca. Competibacter* (COM001), *Niabella* (NIA001b), *Propioniciclava* (PRO001) and *Thiothrix* (THI001).

Sequences homologous to genes coding for proteins involved in the glycogen synthesis pathway were found in the MAG related to *Ca. Accumulibacter*, *Ca. Competibacter*, *Defluviimonas* (DEF002), *Thiothrix*, *Niabella*, *Cytophagaceae* (CYT003), and *Propioniciclava*. The genomic potential to store PHA was detected in the MAG related to *Ca. Accumulibacter*, *Ca. Competibacter*, *Defluviimonas* and *Thiothrix*.

Homologues of genes putatively involved in fermentation pathways were found with a more or less high diversity in all the considered MAG. There are many different proteins involved in fermentation pathways and some of them can be found in fermenting non-fermenting

bacteria. A more thorough analysis of the fermentation-related genes would be required to gather more information about potential fermentative metabolism in the MAG.

Sequences potentially encoding extracellular hydrolases were found in the MAG related to Cytophagaceae, *Dokdonella* (DOK002), *Dechloromonas* (DEC001) and *Niabella*.

Sequences homologous to genes involved in biofilm formation were found in all the considered MAG. In most of them (ACC003, ACC004, CYT003, COM001, DEF002, DOK002, NIA001b), sequences putatively encoding proteins involved in alginate formation were detected. Putative quorum-sensing related genes were found in all the considered MAG except the one related to *Propioniciclava* (PRO001).

### V.3.9 Coverage of hydrolysis- and biofilm-related genes in the different samples

The heatmap of the coverage of selected EC numbers corresponding enzymes involved in hydrolysis and quorum sensing show differences between the three samples collected on day 71, 427 and 740, when the AGS was treating simple monomeric, complex monomeric and complex polymeric wastewater respectively (Figure V.13). Homologues of the genes coding for enzymes involved in the hydrolysis of organic compounds had, as expected, a higher coverage in the contigs corresponding to the biomass treating the complex polymeric wastewater.

The enzyme (4S)-4-hydroxy-5-phosphonooxypentane-2,3-dione isomerase is part of a degradation pathway of the bacterial autoinducer AI-2. The acyl-homoserine lactone acylase is also involved in quorum-quenching as it degrades N-Acyl homoserine lactones another quorum sensing signaling molecule. This enzyme was detected with a higher coverage in the contigs corresponding to sample d740 collected when the reactor was treating complex polymeric wastewater. The sequences coding for the quorum sensing signaling molecules were not detected with this analysis because they are not enzymes.

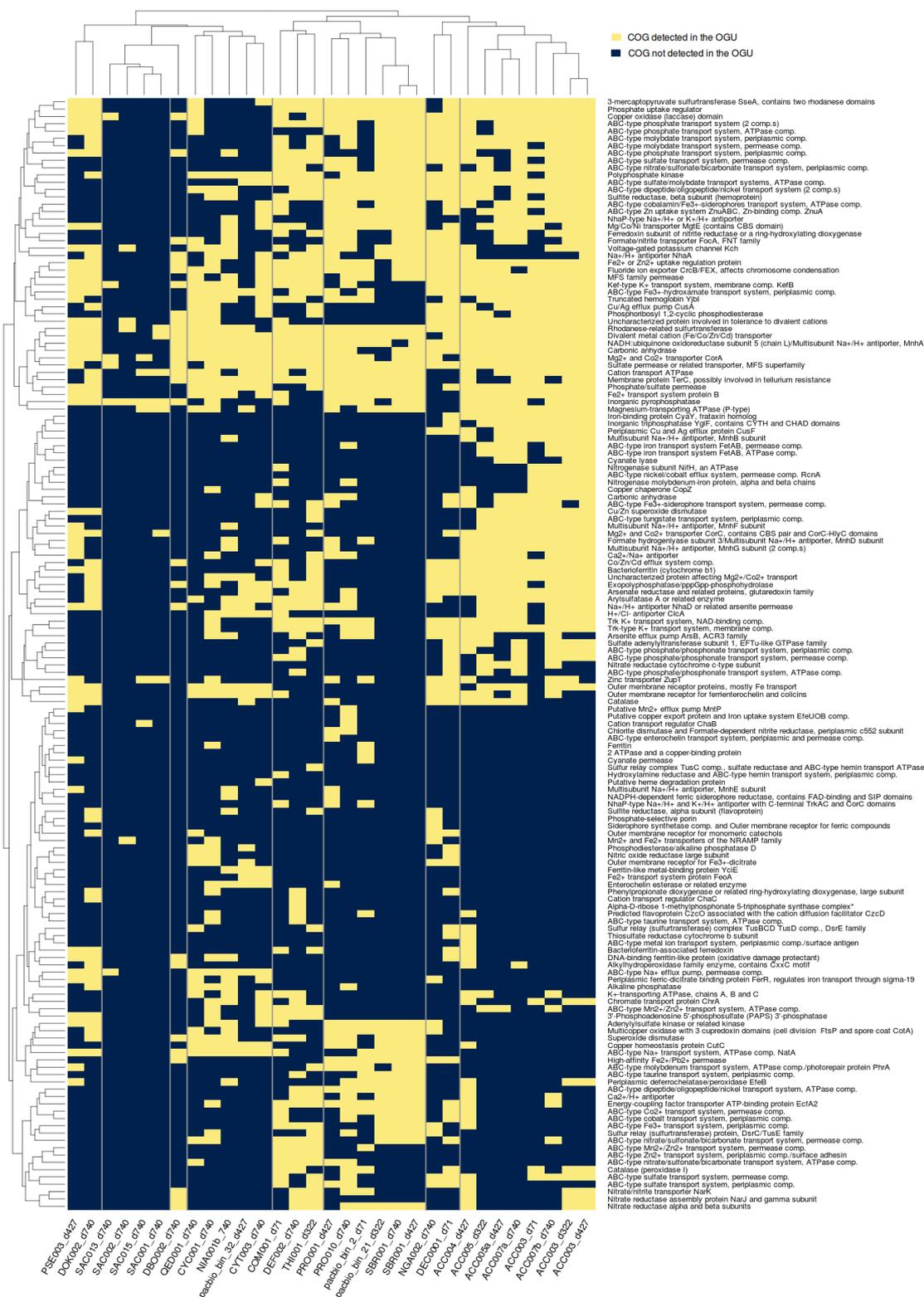


Figure V.11 – Presence/absence of the COG related to inorganic ion transport and metabolism in the main MAG.

## Chapter V. Metagenomic analysis of different AGS samples



Figure V.12 – Presence/absence of the COG related to signal transduction mechanisms in the main MAG.

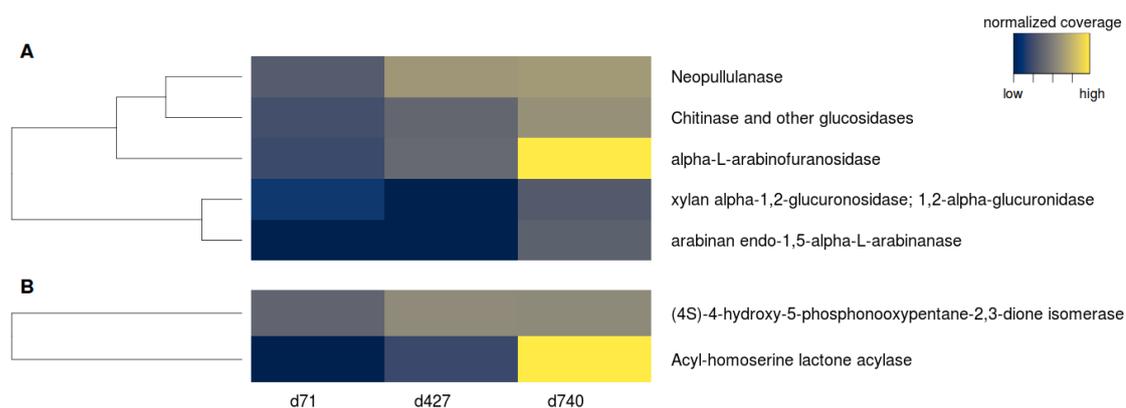


Figure V.13 – Average coverage of the selected EC number corresponding to enzymes related to hydrolysis (A) and quorum-sensing (B) detected in the DNA extracted from the biomass collected on day 71, when the sludge was treating simple wastewater, on 427, when the sludge was treating complex monomeric wastewater and on day 740 when the sludge was treating complex polymeric wastewater. The data from day 322 were not included in the analysis because the quality of the corresponding assembly was very different from the one of the three other samples.

### V.4 Discussion

#### V.4.1 Impact of the DNA extraction methods on the abundance of Actinobacteria

The DNA extracted from AGS samples collected when it was fed with different wastewater types and sequenced with PacBio technology was assembled into contigs which were grouped into operational genomic units (MAG). A total of 32 MAG with an estimated completeness higher than 50 % were recovered from the classes of Gammaproteobacteria, Actinobacteria, Sphingobacteriia, Cytophagia and the phylum Saccharibacteria. Unfortunately, PacBio contigs from the class of Actinobacteria were likely underrepresented compared to the amplicon sequencing data of the corresponding days (see Figures V.1 and IV.7). The extraction method chosen for PacBio sequencing was the one producing the longest DNA fragments, but likely had the disadvantage to keep intact part of the gram+ cells that are difficult to lyse (e.g., Actinobacteria) [Juretschko et al., 2002, Kong et al., 2007]. Albertsen et al. [2012] tested different extraction methods with different bead beating intensities and compared the abundance of the resulting sequences with qFISH. They showed that the class of Actinobacteria was not only underrepresented in the DNA sequences from extraction without bead beating, but also with the bead beating intensities recommended by the DNA extraction kits manufacturer instructions. Increased bead beating intensities increased the relative abundance of extracted Actinobacteria sequences but it also reduced the DNA integrity [Albertsen et al., 2012], which is not advisable for PacBio sequencing since the major limitation for the length of the output sequences is the length of the input DNA.

#### V.4.2 Evaluation of the assembly and binning

The assemblies of the Illumina reads produced unsatisfying results compared to the hybrid and the PacBio only assemblies. In particular, the resulting contigs were small, with a N50 around 10 kbp, as it is generally the case for this type of assemblies [Olson et al., 2017]. Moreover, an important proportion of 16S rRNA genes were incomplete in these assemblies, which is likely indicative of missassembly events. Megahit and metaspades used for these assemblies are considered as good assemblers, yet many ambiguities due to microdiversity, repeat regions, intra- and inter-genomic duplications are simply not possible to resolve with short-reads [Olson et al., 2017, Sloan et al., 2013]. The introduction of PacBio long-reads in the metaspades assemblies enhanced their quality both in terms of contig lengths and proportions of complete 16S rRNA gene sequences recovery. Such improvement of metagenomes sequences assembly quality was often reported [Olson et al., 2017, Slaby et al., 2017].

The PacBio assembly produced significantly fewer and longer contigs. After inspecting the mapping of the raw PacBio long-reads against the contigs obtained after different steps of polishing, it is likely that some of these contigs were constructed by merging several single nucleotide polymorphisms (SNP), small deletions or insertions of closely related strains. Therefore, the MAG formed with these contigs probably represent consensus genomes. The

quality of the contigs having a high coverage was good, according to the mapping of the raw long-reads on the final assembly. The potential high quality of PacBio only assemblies for metagenomes was previously noticed with synthetic data [Olson et al., 2017]. Yet, the contigs constructed from a small coverage of long-reads had a lower quality according to the mapping, therefore, the assembly of long-reads alone may not be adapted to recover the genomes of low-abundance bacteria.

The good quality of the main MAG resulting from the grouping of PacBio contigs is supported by several evidence. The number of coding sequences was approximately the same as in the closest reference genomes. Errors like insertions or deletions would interfere with the open reading frames (ORF) and introduce false stop-codon, thus lowering the numbers of detected CDS. The few MAG with a same 16S rRNA gene sequence recovered from the samples of different days contains a strongly similar information in the sense that the annotation of these genomes is almost equal. Sets of single-copy orthologous genes are generally used to estimate the completeness and contamination of metagenome assembled genomes [Albertsen et al., 2013, McIlroy et al., 2014, Sharon et al., 2013]. In this study, having the less BUSCO in common was used as a criterion to group the contigs together. It was therefore expected that the estimation of duplication in the MAG would be low ( $< 1\%$  for most of the MAG). A set of truly universal essential single copy genes for bacteria has likely not been found yet [Dupont et al., 2012], and therefore the estimation performed by CheckM is based on different sets of genes according to the phylogeny of the MAG as detected by the program. Yet, if the lineage is not assessed accurately, the set of marker genes chosen for the analysis may not be adapted. The estimation of completeness and contamination depends on the presence of closely related genomes in the database and their accuracy therefore varies accordingly. For example, the most complete MAG from Saccharibacteria were estimated complete around 65 % . Arlbetsen et al. [2013] recently assembled four genomes from Saccharibacteria. These genomes have approximately the same size as the Saccharibacteria-related MAG found in this study and a similar estimated completeness. 6.5 % of the single copy essential genes used to evaluate the completeness of the genomes were not found in any of these four ones supporting the hypothesis that Saccharibacteria members have smaller genomes due to parasitic/commensalistic/symbiotic lifestyles [Bor et al., 2019, McLean et al., 2018]. The completeness of Saccharibacteria genomes using single copy essential genes is likely biased and could be adapted in the future with the information recently collected on different genomes belonging to this phylum.

The method and the thresholds used to group the PacBio contigs into MAG were chosen so that the contamination was minimized, perhaps in some cases to the detriment of completeness. The MAG can probably be completed manually with the help of the assembly graphs. Yet, the statistics on the collected MAG are comparable to the ones of similar metagenomic studies using short-reads or a combination of short- and long-reads to reconstruct genomes [Hao et al., 2018, Lawson et al., 2017]. This suggest that the assembly of PacBio long-reads is promising for the recovery of genomes from metagenomes of medium complexity.

### V.4.3 The diversity of *Ca. Accumulibacter* complicates the assembly of the genomes from different clades

The assembly graphs show two different kinds of contigs: isolated contigs and contigs linked with other ones producing bubbles and multiple possible paths. Most of the contigs attributed to *Ca. Accumulibacter* genus belongs to the second category. In comparison, the MAG corresponding to *Dechloromonas* was assembled in a single contig despite a lower coverage than some *Ca. Accumulibacter*-related MAG. Difficulties to assemble *Ca. Accumulibacter* sequences despite a high coverage was previously reported [Slaby et al., 2017]. *Ca. Accumulibacter* is an extremely diversified genus [He et al., 2007, Leventhal et al., 2018, McMahon et al., 2007]. According to the results of this study, part of this diversity was present in the AGS of the four metagenomes sequenced, while only one 'clone' related to *Dechloromonas* was found among the main MAG.

Moreover, the quantity and phylotype of the *Ca. Accumulibacter*-related MAG varied through the different samples, possibly due to the changes in the wastewater composition and/or other factors (e.g., phage). This high diversity inside the *Ca. Accumulibacter* genus and the shifting of the most abundant clades were previously reported in the biomass of reactors performing biological phosphorus removal [He et al., 2007, Leventhal et al., 2018, McMahon et al., 2007]. In this study, the clade IA, IIA and other phylotypes probably from type I were present in the AGS treating simple and complex monomeric wastewater, ACC004 quite different from the other MAG, to which no type could be assigned, was present in the AGS treating complex monomeric wastewater. The clades IA and IIA were likely not abundant in the AGS treating complex polymeric wastewater, instead ACC007, related to type I was the main *Ca. Accumulibacter* phylotype. Further studies will be needed to determine if these changes were linked with the changes in the wastewater composition.

The presence of multiple closely related strains within the same sample tends to complexify the assembly, with highly similar parts and others that are markedly different [Nurk et al., 2017, Sloan et al., 2013]. In addition, *Ca. Accumulibacter* bacteria seem to harbour an important CRISPR-Cas system, which contains variable parts of bacteriophage and increase the diversity of the sequences. Part of the CRISPR-Cas sequences contribute to the 'ball of wool' in the assembly graph from sample d71. A thorough analysis of these sequences would be required in order to understand their structure and potentially link them to one of the MAG. It is also possible that *Ca. Accumulibacter* have active mechanisms of genome recombination that make the genomes unstable through the closely related strains and even in time. This could be the subject of further investigations.

### V.4.4 The different clades of *Ca. Accumulibacter* present in the AGS

Similarly to previous studies, the classification of *Ca. Accumulibacter*-related MAG based on the *ppk1* genes corroborated the classification based on 16S rRNA genes [He et al., 2007]. The classification based on the *ppk1* was more precise as it distinguished between the sequences

of MAG having the same 16S rRNA gene sequence (ACC005 and *Ca. Accumulibacter* sp. BA-93, ACC007a and ACC007b from sample d740). This observation was the motivation for choosing *ppk1* to distinguish different *Ca. Accumulibacter* genotypes [He et al., 2007]. Based on the different phylogenetic trees including *Ca. Accumulibacter* sequences from clades IA, IC, IIA, IIC and IIF, the MAG ACC001 is close to the clade IIA, ACC005 belongs to clade IA, ACC003 and ACC007 belong to the type I but the specific clade could not be determined due to the absence of representant of some clades in these analyses. The clade of ACC004 could not be determined, but this genotype is a little apart from the other *Ca. Accumulibacter*-related MAG and references included in this study. ACC004 distinguish itself from the other genomes of the genus also by a fewer number of phosphorus-cycling related genes detected (only one *Pit* gene and one *PhoU*). This suggest that ACC004 may not have the same capability to accumulate and release inorganic phosphate as the other MAG of the same genus. The presence of *Ca. Accumulibacter* clades which do not constitutively accumulate large amount of polyphosphate was reported [Kong et al., 2004, Zilles et al., 2002], but the hypothesis that ACC004 accumulates lower polyP than the other *Ca. Accumulibacter* needs to be supported by experimental evidences.

Based on key gene sequences, the MAG ACC004\_d427 and ACC007b have the genomic potential to denitrify from nitrate through a periplasmic nitrate reductase, the MAG ACC001a and ACC005\_d71 to denitrify from nitrite with a possibility that a periplasmic nitrate reductase was not found in the two latest MAG due to their incompleteness, since it is present in the reference genomes of the corresponding clades. These results are in line with the conclusions of previous metagenomic studies on type IA and IIA. The complete denitrification pathway was not found in the relatively complete MAG ACC003, but the genomic potential to respire nitrate (through a respiratory nitrate reductase) and nitrite was.

#### V.4.5 Marker genes for PAO metabolism

Phosphate is an essential micronutrient for bacteria, therefore polyphosphate kinase, exopolyphosphatases and phosphate transporters are commonly found in bacteria that do not accumulate large amounts of polyphosphate [Albi and Serrano, 2016, Kornberg et al., 1999]. The low-affinity inorganic phosphate transporter (*Pit*) was a good candidate tho discriminate PAO from non-PAO since in was not found in *Ca. Competibacter* [McIlroy et al., 2014], yet it was detected in the genome of GAO bacteria [Wang et al., 2014a]. The uncharacterized protein encoded by the gene *PhoU* is thought to be related to the regulation of phosphate uptake. *Pseudomonas* and *Escherichia coli* mutants having a *phoU* gene knocked out, were reported to accumulate unusually high amount of polyphosphate and alarmone guanosine tetraphosphate (ppGpp) intracellularly and be less resistant to stress [de Almeida et al., 2015, Morohoshi et al., 2002]. Multiple homologues of these genes were found in the MAG and the reference genomes of *Ca. Accumulibacter*, but only one in the reference genomes of *Tetrasphaera*. Moreover homologous sequences of *PhoU* were found in non-PAO MAG such as *Thiothrix* or *Zoolgoea*. A marker gene for PAO metabolism, if it exists, is still to be found.

### V.4.6 Biofilm formation

The genes encoding proteins involved in biofilm formation were detected in all the main MAG, in particular, proteins related to alginate production were present in most of the main MAG. Yet, this was also observed in metagenomes from activated sludge systems [Albertsen et al., 2012]. Quorum-sensing was shown to play an important role in granule formation. For example, the presence of autoinducer acyl-homoserine lactone, a quorum-sensing protein, can induce the granulation of floccular activated sludge [Barr et al., 2016, Schaefer et al., 1996, Tan et al., 2014]. Several homologues of quorum-sensing related genes, including acyl-homoserine lactone export, were detected in almost all the main MAG of this study. It is possible that the potential to form biofilm is present in most of the bacteria constituting the AGS in this study and that it is actually triggered by external factors influencing quorum sensing signaling. For example, the amount of COD incorporated by the bacteria could influence positively expression of genes involved in biofilm formation [Burne, 1998, Jefferson et al., 2004].

### V.4.7 Repartition of COG in the MAG

The PCA performed on the presence/absence of COG in the main MAG revealed a link between the functional orthology and the phylogeny. Such a relation was already noticed with a metagenome from sponge microbiomes [Slaby et al., 2017]. Yet, in this study, all the analysis based on the gene prediction are biased. Most of the proteins have not been characterized yet or their function is only inferred from homology, which does not guarantee a homologous function [Martin et al., 2006]. The putative functions detected in the present metagenomes can be used to make hypothesis and serve as a template for metatranscriptomic, metaproteomic, or other molecular analyses such as (MAR-FISH) [Oyserman et al., 2016].

## V.5 Conclusions

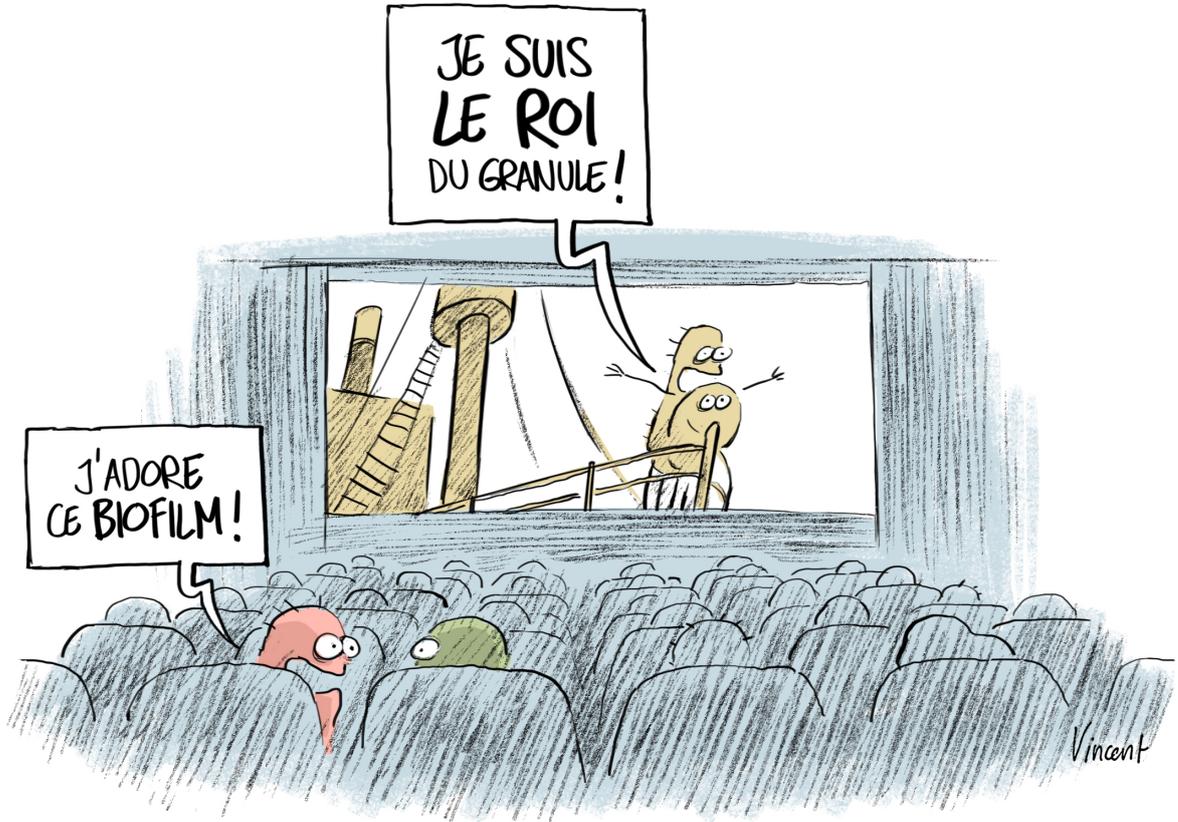
The metagenomes of four AGS samples collected when the reactor was treating simple, complex monomeric and complex polymeric synthetic wastewater were sequenced in order to get indications about the roles of the different bacterial populations in the nutrient removal and granulation processes and to provide a template for future metatranscriptomic experiments.

PacBio and Illumina sequencing were employed and different assemblies with the sequences obtained from one of these technologies or a combination of both were tested. Based on the contig lengths and the proportion of complete 16S rRNA gene sequences, the PacBio assembly proved to be the best of the tested assemblies and was therefore chosen for the rest of the analysis. Operational genomic units (MAG) were built from this assembly by using the Illumina read coverages, with universal single copy genes and tetranucleotide frequency to bin the PacBio contigs. 31 of these MAG had were estimated to be at least 50 % complete. Several MAG affiliated with unclassified genera (DB1-14, Saccharibacteria) or from which no complete genome was found in the public databases (*Sbr-gs28*, *Propioniciclava*) had an estimated

completeness higher than 80 % and delivered few clues about their potential metabolism. The few MAG recovered from different samples but carrying an identical 16S rRNA gene sequence had almost identical gene annotations, testifying of the reproducibility of the analysis. This supports the fact that despite a high error rate, PacBio technology can advantageously be used to sequence low-to-medium complexity genomes. The data collected in this study will be further exploited to investigate more thoroughly the genomic information it contains and will hopefully serve as a basis for future metatranscriptomic analysis.

Multiple strains of *Ca. Accumulibacter* were found in each samples likely making the assembly of their genomes very complex. This diversity of *Ca. Accumulibacter* evolved through the changes in the wastewater composition feeding the AGS samples and were not the same with the sample treating simple, complex monomeric or complex polymeric wastewater. Whether these changes are linked with the wastewater composition or with other factors will need further investigations, but different metagenomic potential of the different MAG were already noticed, in line with previous studies. In particular, the potential to respire nitrate is not provided by the same type of enzyme in the different MAG and the full denitrification pathway is potentially present in some of the MAG but not in others. Moreover, one of the MAG seems to carry significantly less phosphate-cycling related genes than the others, suggesting that it may have a lower phosphorus removal capability.





Vincent



## VI General discussion

### VI.1 Bacterial communities in AGS fed with different wastewater types

Two experiments were designed to study the influence of the wastewater composition on the bacterial communities of aerobic granular sludge (AGS), as well as their settling properties and nutrient removal performance with two different approaches. The experiment carried out at Eawag (hereafter referred to as Eawag experiment) studied the influence of wastewater composition, in terms of diffusible carbon, on the transformation of activated sludge into AGS (chapter III). The experiment carried out at EPFL (hereafter referred to as EPFL experiment) studied the effect of progressive complexification of the synthetic wastewater by introducing fermentable compounds in a first step and polymeric compounds in a second step on already formed AGS acclimated to simple wastewater containing acetate and propionate as carbon source (chapter IV).

The two principal coordinates of the bacterial communities during stable states from both experiments are shown in Figure VI.1. They are clearly separated by experiment on the first axis. The comparison of bacterial communities from different wastewater treatment plants usually produces similar results, with an important core community and yet a clear clustering of the samples according to their sampling site [Saunders et al., 2016].

The wastewater composition corresponding to stable state 7 and R1 (simple wastewater) and those corresponding to stable state 6 and R2 (synthetic complex polymeric wastewater) were similar in terms of carbon substrates. Yet the phosphate concentrations were very different between the two experiments:  $6 \text{ mg l}^{-1}$  of  $\text{P-PO}_4$  at Eawag, in order to be close to the values typically measured in Swiss municipal wastewater,  $22 \text{ mg l}^{-1}$  at EPFL to promote the growth of phosphate-accumulating organisms (PAO). Moreover, the sizes and operation of the reactors were not exactly the same for the two experiments. As mentioned above, the seed sludge for the Eawag experiment was activated sludge from an EBPR wastewater treatment plant, while it was AGS acclimated to simple wastewater for the EPFL experiment. The bacteria brought by

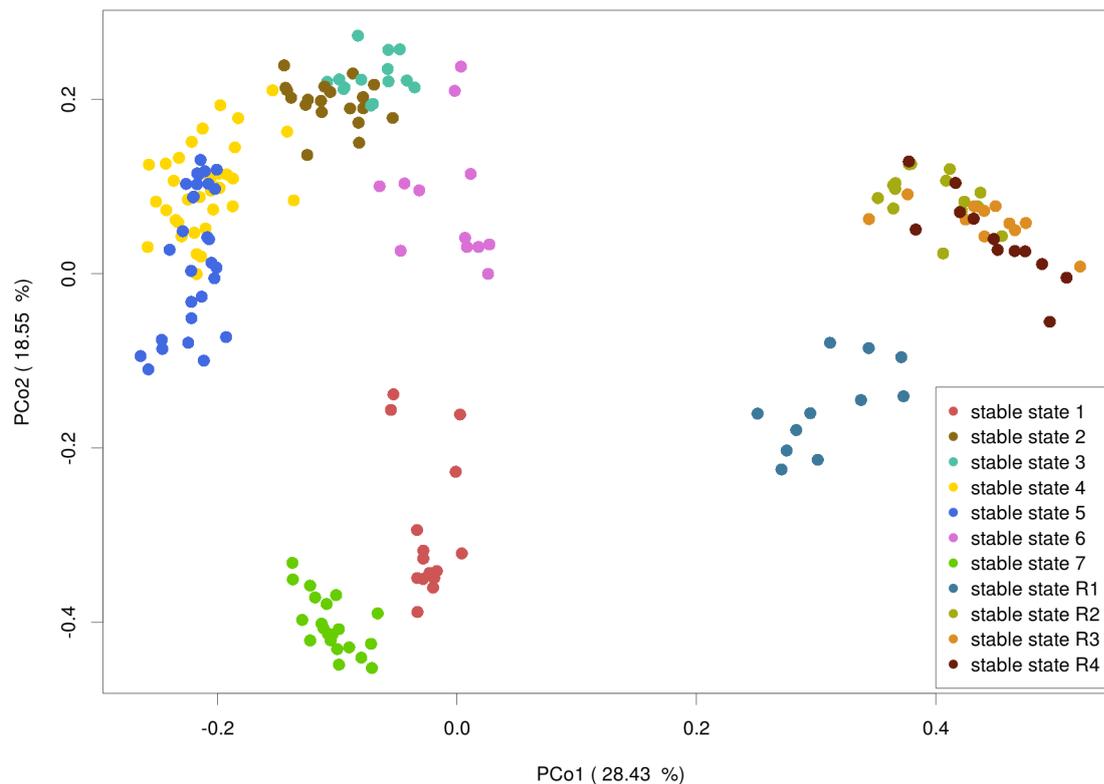


Figure VI.1 – Principal coordinate analysis of the Bray-Curtis distance of the bacterial communities at stable states in the EPFL (stable states 1-7) and Eawag (R1-R4) experiment. Stable states 1, 7 and R1 correspond to simple synthetic wastewater, stable states 2, 3, 4 and 5 correspond to complex monomeric synthetic wastewater, stable states 6 and R2 correspond to complex polymeric synthetic wastewater and stable states R3 and R4 correspond to municipal wastewater. 46.98 % of the total variability of the bacterial communities are shown in this projection. The graphs showing the projection on the third axis are presented in supplementary material (Figure A.18)

the municipal wastewater can influence the sludge community [Saunders et al., 2016]. This is not part of the reasons explaining the differences in the bacterial communities shown in the first axis of the PCoA because the bacterial communities from the reactors treating synthetic and municipal wastewater are grouped together in this projection.

A total of 425 taxa (genus level) were detected in the bacterial communities of the Eawag experiment, 293 in the ones of the EPFL experiment. From those, 223 were common to both experiments and 28 were abundant in both experiment. Among the abundant genera (>1 % at stable state) detected only in the Eawag experiment were *Trichococcus*, several *Competibacteraceae* (*Ca. Contendobacter*, *CPB\_CS1*), *Ca. Mictothrix*, and *Ca. Epiflobacter*. The *Competibacteraceae CPB\_S18* was the abundant genus detected only in the EPFL experiment.

In the two experiments, the bacterial communities treating simple wastewater were significantly different from the ones treating complex wastewater (Figure VI.1). AGS treating simple

## VI.1. Bacterial communities in AGS fed with different wastewater types

wastewater were characterized by very low proportions of Actinobacteria and Saccharibacteria and high proportions of Gammaproteobacteria. AGS treating complex wastewater were characterized by a more diverse bacterial community, in particular with the presence of Saccharibacteria and Actinobacteria. The main genera from the two experiments are presented in table VI.1 depending if they were enriched in simple wastewater, in complex polymeric wastewater or if the difference of mean proportion between the two wastewater types was not significant. The majority of the genera were found in both wastewater type, yet numerous taxa, including several Actinobacteria, were underrepresented in the simple wastewater compared to complex polymeric wastewater. Holling et al. [1973] compared the microbial communities of AGS treating a simple synthetic (acetate) and a municipal wastewater and found 80 % of Proteobacteria and almost no Actinobacteria in the first AGS and less than 50 % of Proteobacteria and around 2 % of Actinobacteria in the second AGS. Percentages of < 1 % to 33 % of Actinobacteria were reported in AGS treating complex polymeric wastewater [Cetin et al., 2018, Kang et al., 2018, Liu et al., 2017, Swiatczak and Cydzik-Kwiatkowska, 2018]. Yet it is possible that the abundance of Actinobacteria was underestimated by amplicon-sequencing based evaluations [Albertsen et al., 2015]. For example Cetin et al. [2018] reported good P-removal performance, but no PAO in the most abundant genera detected in the AGS.

Table VI.1 – Main taxa common to Eawag and EPFL experiments enriched in the simple wastewaters, the complex polymeric wastewaters or in both wastewater types.

simple wastewater	complex polymeric wastewater	both types
<i>Thiothrix</i>	<i>Micropruina</i>	<i>Tetrasphaera</i>
<i>Ca. Accumulibacter</i>	Saccharibacteria (p)	<i>Zoogloea</i>
<i>Ca. Competibacter</i>	<i>Kouleothrix</i>	<i>Nitospira</i>
<i>Flavobacterium</i>	<i>Paracoccus</i>	<i>Rhodobacter</i>
DB1-14 (o)	Saprospiraceae (f)	<i>Dechloromonas</i>
<i>Ca. Nitrotoga</i>	<i>Thauera</i>	<i>Propioniciclava</i>
<i>Flavobacterium</i>	<i>Dokdonella</i>	Anaerolineaceae (f)
Hyphomonadaceae (f)	Xanthomonadaceae (f)	CYCU-0281
<i>Nitrosomonas</i>	<i>Propionicimonas</i>	CPB_C22&F32
Armatimonadetes (p)	<i>Nocardioides</i>	<i>Meganema</i>
	<i>Diaphorobacter</i>	Comamonadaceae (f)
	<i>Shinella</i>	<i>Pseudoxanthomonas</i>
	<i>Iamia</i>	<i>Propionivibrio</i>
		Cytophagaceae (f)

Similar differences also exist between the bacterial communities of EBPR activated sludge fed with simple synthetic wastewater and real wastewater [Swiatczak and Cydzik-Kwiatkowska, 2018, Weissbrodt et al., 2014b]. The bacteria enriched in the AGS treating complex wastewater compared to AGS treating simple wastewater, and in particular members of Actinobacteria and Saccharibacteria, may be part of the AGS bacterial core communities and are missed when working with simple wastewater.

### VI.2 Functional groups in AGS

To study the evolution of the different functional groups in the microbial communities during the two experiments, the proportions of different functional groups were assessed by assigning the functional information provided by the MiDAS microbe browser [McIlroy et al., 2015a] to the corresponding genera of the bacterial communities.

#### VI.2.1 PAO and GAO

At Eawag, the proportions of PAO diminished from 10 % to below 5 % during the start-up from EBPR activated sludge to AGS with the synthetic and municipal wastewater (Figure VI.2). In the reactors treating complex wastewater, the relative abundance of PAO, mainly *Tetrasphaera*, remained around 10 % during the transition, and then was lower at stable state (1-3 %). This is lower than the proportions of PAO generally reported in EBPR treatment plants (12 - 36 %) [Mielczarek et al., 2013, Muszynski and Zaleska-Radziwill, 2015, Nielsen et al., 2010]. The reasons for the low abundances of PAO in the stable states are unclear. Certainly the low phosphate concentrations in the influent wastewater did favor GAO over PAO [de Kreuk et al., 2005, de Kreuk and van Loosdrecht, 2004, Weissbrodt et al., 2014a, Zeng et al., 2003b]. It is also possible that these AGS contain PAO that are not yet identified.

The proportion of PAO at stable state was lowest in the AGS treating complex synthetic wastewater (R3). Yet, 1.5 % of *Microlunatus*, which was not in this version of MiDAS browser, were detected in this AGS. A PAO metabolism with the uptake of glucose and glutamate was detected in *Microlunatus* phosphovorus [Santos et al., 1999]. It is therefore possible that part of the *Microlunatus* in R2 had a PAO metabolism, but it is difficult to confirm this hypothesis afterward. If the *Microlunatus* was counted as a PAO, the total of PAO in R2 would be similar to the ones detected in the AGS treating municipal wastewater (R3 and R4).

The proportions of GAO was lower than 4 % in the EBPR activated sludge inocula. In the sludge treating simple wastewater, it increased and fluctuated around 8 % during the transition to AGS and stabilized around 18 % at stable state. In the AGS treating complex wastewater, the proportions of GAO, mainly *Micropruina*, increased during the transition to AGS, and then were around 13 % at stable state. These proportions were lower in the AGS treating municipal wastewater (< 5 %) at stable state.

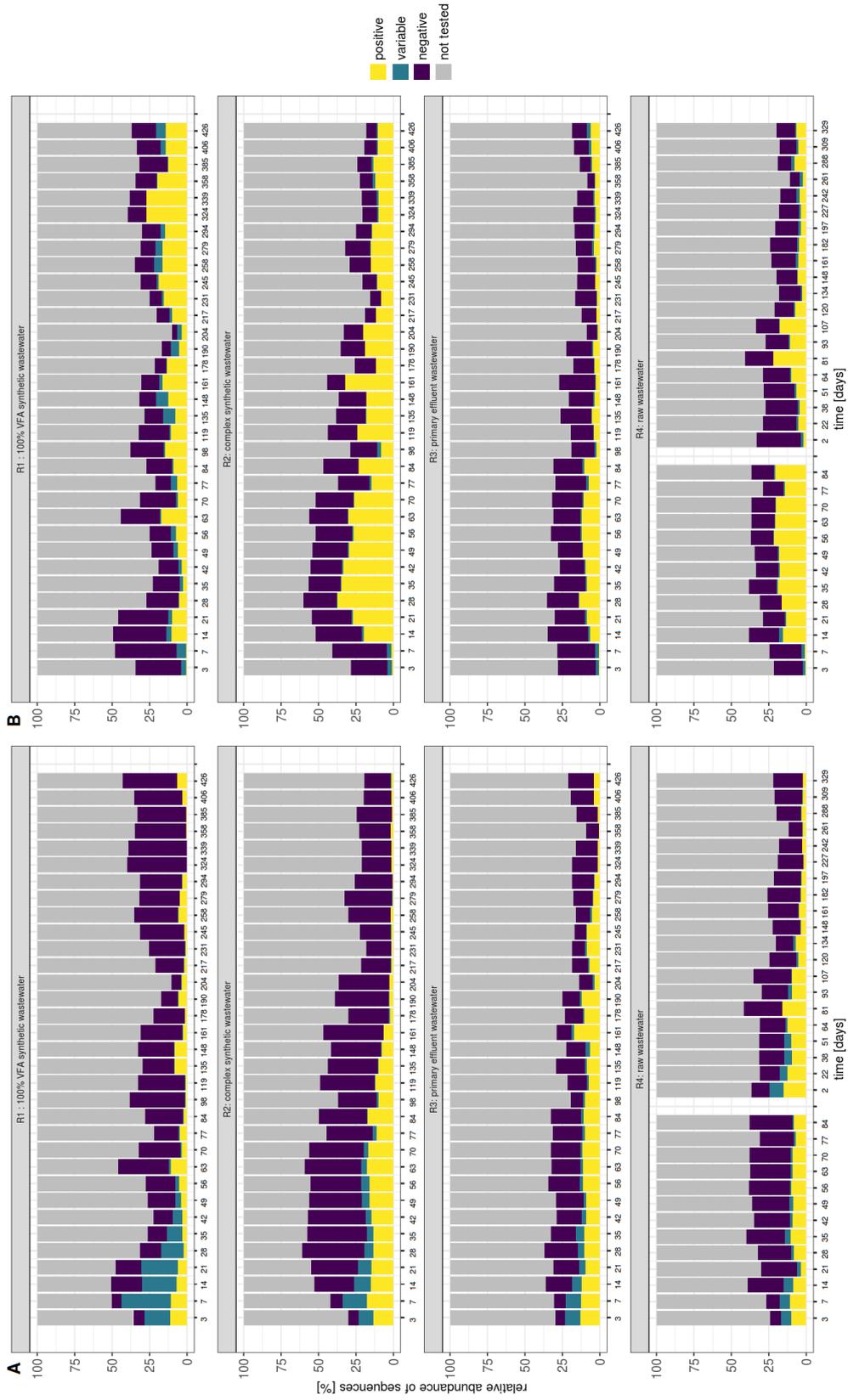


Figure VI.2 – Relative abundance of PAO (A) and GAO (B) during the start-up from EBPR activated sludge to AGS in reactors treating different wastewater types in Eawag experiment. The functions were attributed to taxa defined at genus level based on the functional information provided on MiDAS browser.

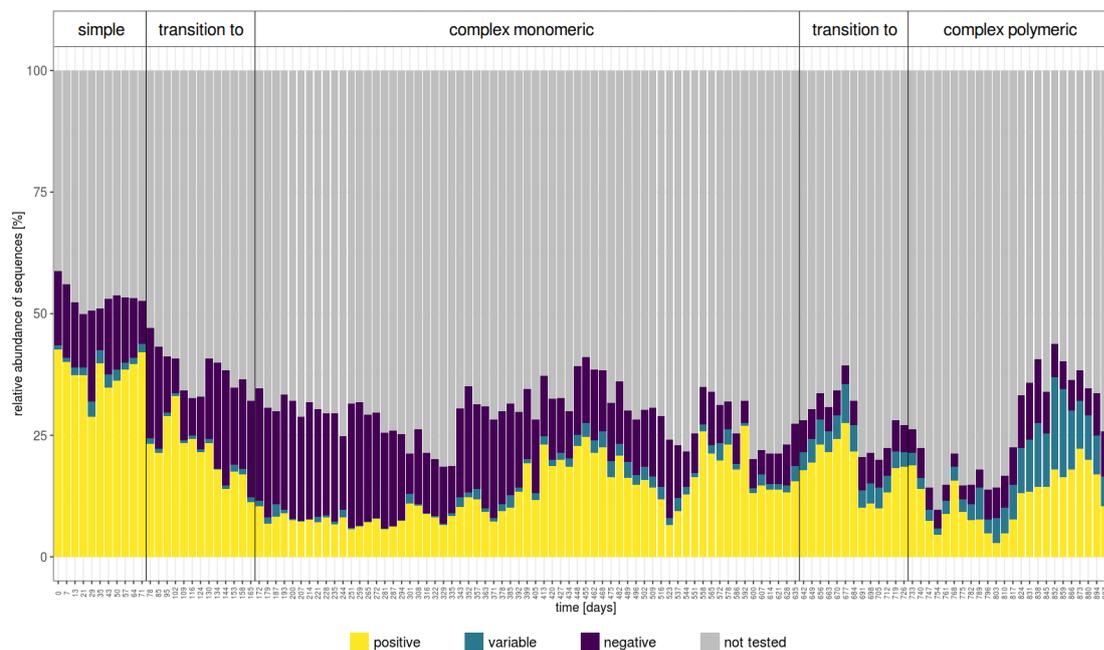


Figure VI.3 – Relative abundance of PAO during the transition from simple monomeric, to complex monomeric and to complex polymeric influent wastewater in EPFL experiment.

At EPFL, the proportions of PAO were around 40 % at the beginning of the experiment, in the AGS treating simple wastewater (Figure VI.3). This is higher than the proportions measured in activated sludge from EBPR treating municipal wastewater and was probably a consequence of the high phosphate concentration in the influent wastewater [Weissbrodt et al., 2014a]. The proportion of PAO fluctuated a lot during this experiment. It is likely that the proportions of *Tetrasphaera* were underestimated leading to an important underestimation of the proportion of total PAO when *Tetrasphaera* was the dominant PAO [Albertsen et al., 2015].

The proportions of GAO also fluctuated during the EPFL experiment (Figure VI.4). The proportions of GAO increased during the transition from simple to complex monomeric wastewater which is also the period corresponding to very low proportions of *Ca. Accumulibacter*.

The compositions of the guild of PAO (including putative PAO) according to the different wastewater types in the EPFL and Eawag experiments during stable states are presented in Figure VI.5. In the reactors fed with simple wastewater, the PAO were mainly Betaproteobacteriales, whereas with the synthetic complex wastewater, the guild of PAO was more diverse with *Tetrasphaera* representing a higher proportion of the guild. The guild of PAO in the reactors treating municipal wastewater had a higher richness with the presence of the PAO *Ca. Obscurimonas* and *Ca. Accumulimonas* which were not detected in the AGS treating synthetic wastewater.

The compositions of the guild of GAO according to the different wastewater types in the EPFL and Eawag experiments during stable states are presented in Figure VI.6. The trends

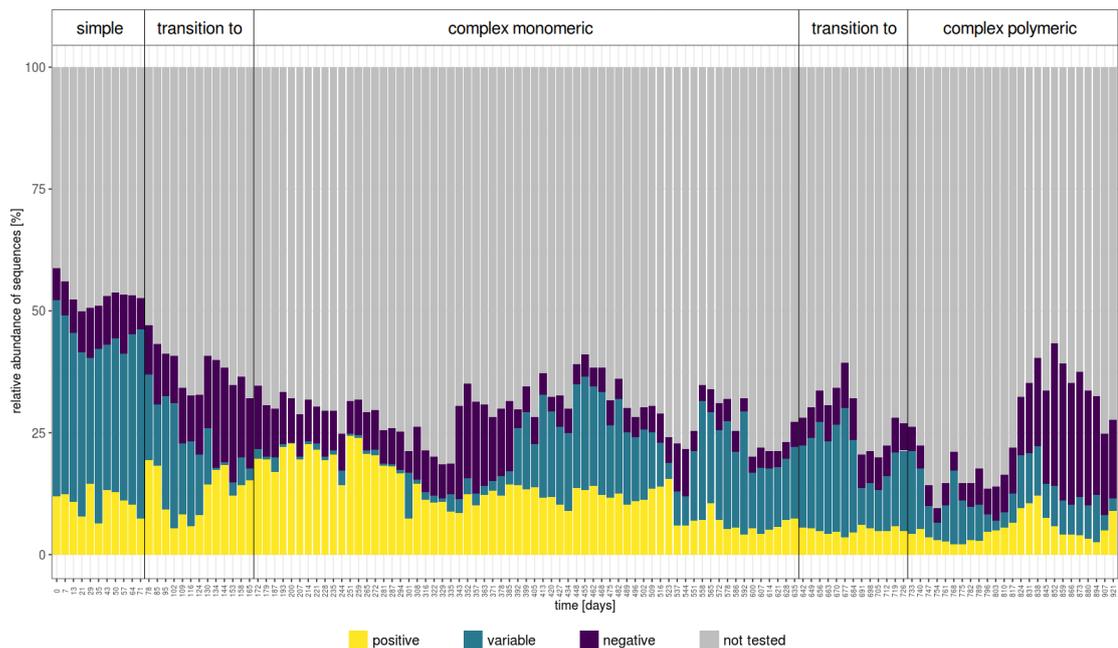


Figure VI.4 – Relative abundance of GAO during the transition from simple monomeric, to complex monomeric and to complex polymeric influent wastewater in EPFL experiment

observed with the guild of PAO are similar for the one of GAO. The compositions of the guilds treating the same wastewater type were similar in both experiments. In the AGS treating simple wastewater, the GAO belong mostly to Gammaproteobacteria and more particularly to the Competibacteraceae family. In the AGS treating complex polymeric wastewater (synthetic and municipal), the GAO are mostly Actinobacteria, with a high proportion of *Micropruina*. This corresponds to the observations made on the sludge collected in EBPR wastewater treatment plant, where *Micropruina* was often representing a high proportion of the guild of GAO [Stokholm-Bjerregaard et al., 2017]. The guild of GAO had also a higher richness in the AGS treating municipal wastewater.

### VI.2.2 Fermenting bacteria

The proportions of fermenting bacteria during the start up of AGS from EBPR activated sludge are shown in Figure VI.7. These proportions were around 5 % in the inocula. In the reactor treating simple wastewater it diminished below 1 % at stable state. In the reactors treating complex wastewater, the proportions of fermenting bacteria increased during the transition phase and then returned close to their initial values in the reactors treating complex wastewater at stable state.

## Chapter VI. General discussion

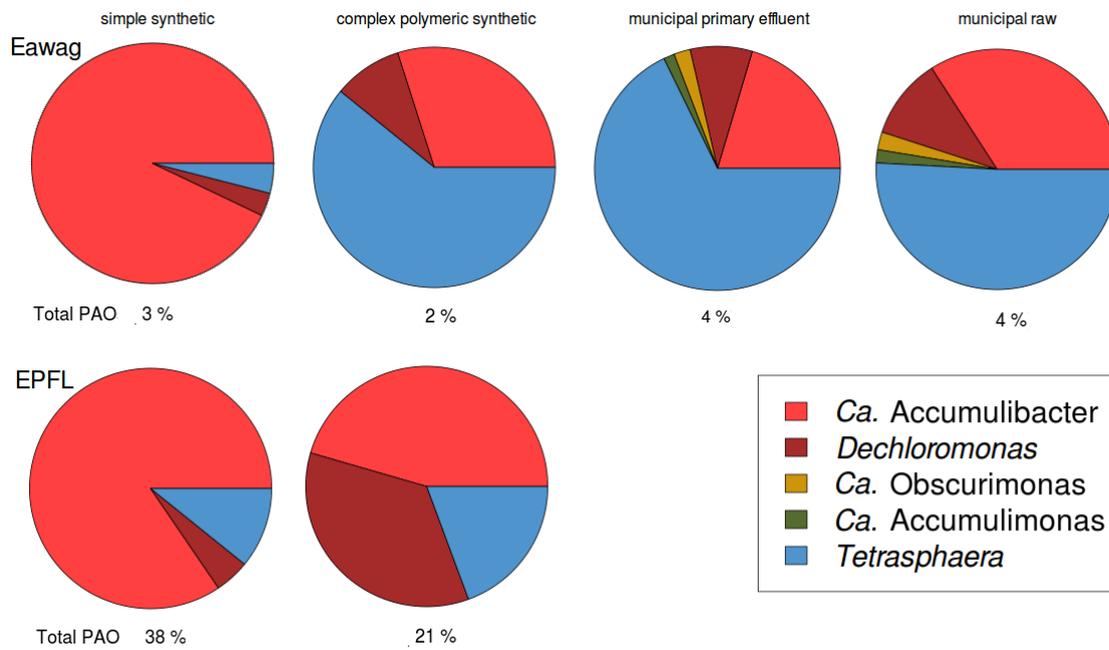


Figure VI.5 – Composition of the guild of PAO in the Eawag and EPFL experiments, according to the different wastewater types. Functions were provided by MiDAS browser

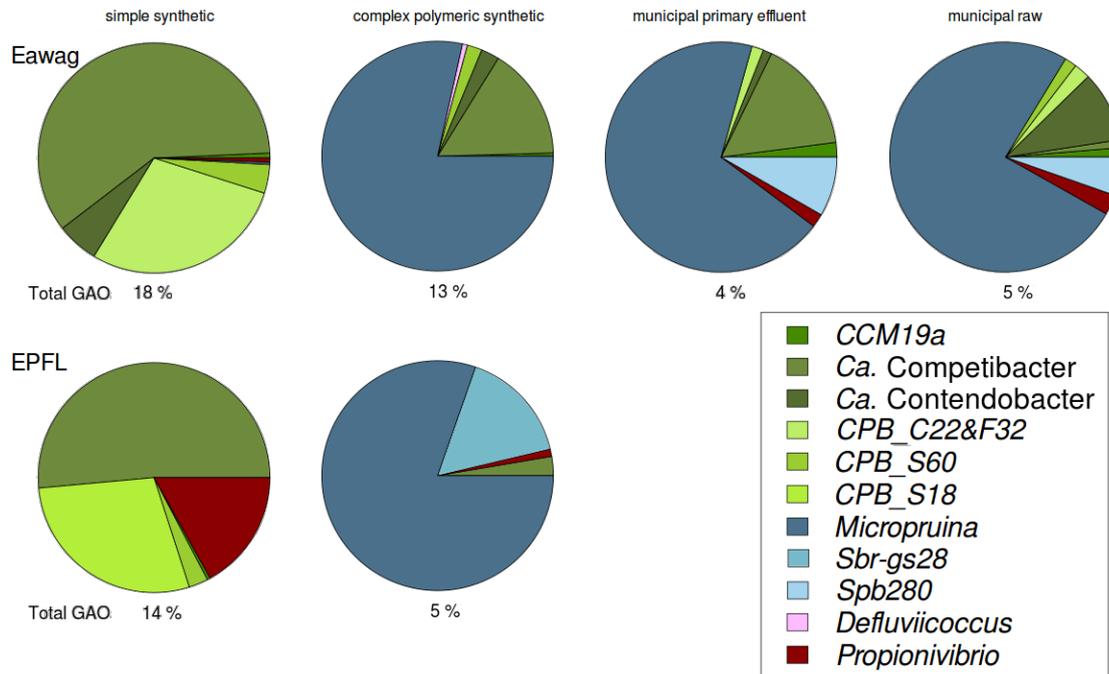


Figure VI.6 – Composition of the guild of GAO in the Eawag and EPFL experiments, according to the different wastewater types. Functions were provided by MiDAS browser.

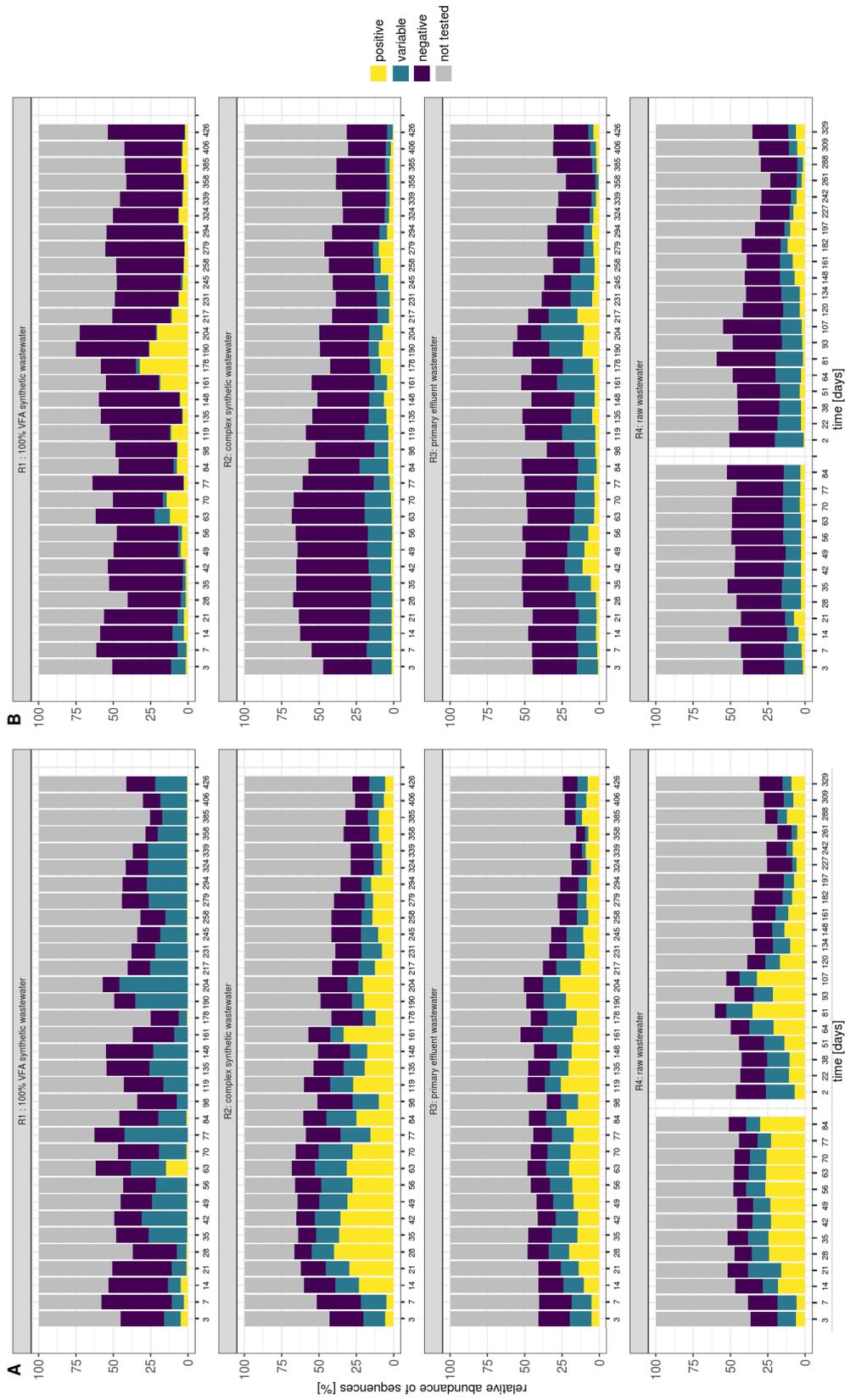


Figure VI.7 – Relative abundance of fermenting (A) and filamentous (B) bacteria during the start up from EBPR activated sludge to AGS in reactors treating different wastewater types in Eawag experiment. The functions were attributed to taxa defined at genus level based on the functional information provided on MiDAS browser.

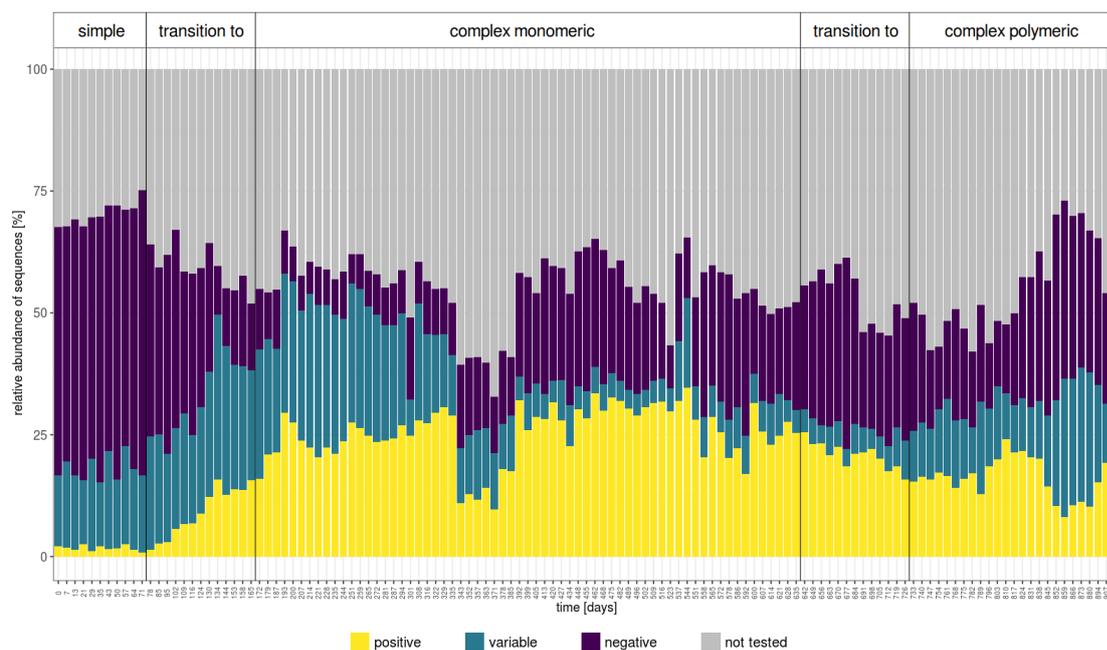


Figure VI.8 – Relative abundance of fermenting bacteria during the transition from simple monomeric, to complex monomeric and to complex polymeric influent wastewater in EPFL experiment. The functions were attributed to taxa defined at genus level based on the functional information provided on MiDAS browser.

The proportions of fermenting bacteria in the AGS treating simple wastewater and after the transitions to complex monomeric and to complex polymeric wastewater are presented in Figure VI.8. The low proportions of fermenting bacteria detected in the AGS treating simple wastewater increased up to 25 % during the transition to complex wastewater. This proportion was lower during few weeks following the duplication of the reactors, coming with an increased food to microorganism ratio. In this graph, the potentially fermenting bacteria from uncharacterized genera are not included in the fermenting or putative fermenting bacteria. For example, the unclassified members of Saccharibacteria that may ferment glucose if they have a similar metabolism as *Ca. Saccharimonas* [Albertsen et al., 2013] are colored in grey. Therefore the fluctuations in the proportions of fermenting bacteria after the transition to complex polymeric wastewater may be an artifact of the analysis.

In the two experiments, almost no fermenting bacteria were detected in the AGS treating simple wastewater, whereas 12 % to 30 % were detected in the AGS treating complex wastewater at stable states. Similar percentages of fermenting bacteria with a majority of Actinobacteria were also reported in biological nutrient removal wastewater treatment plants [Kong et al., 2008, Nielsen et al., 2012a].

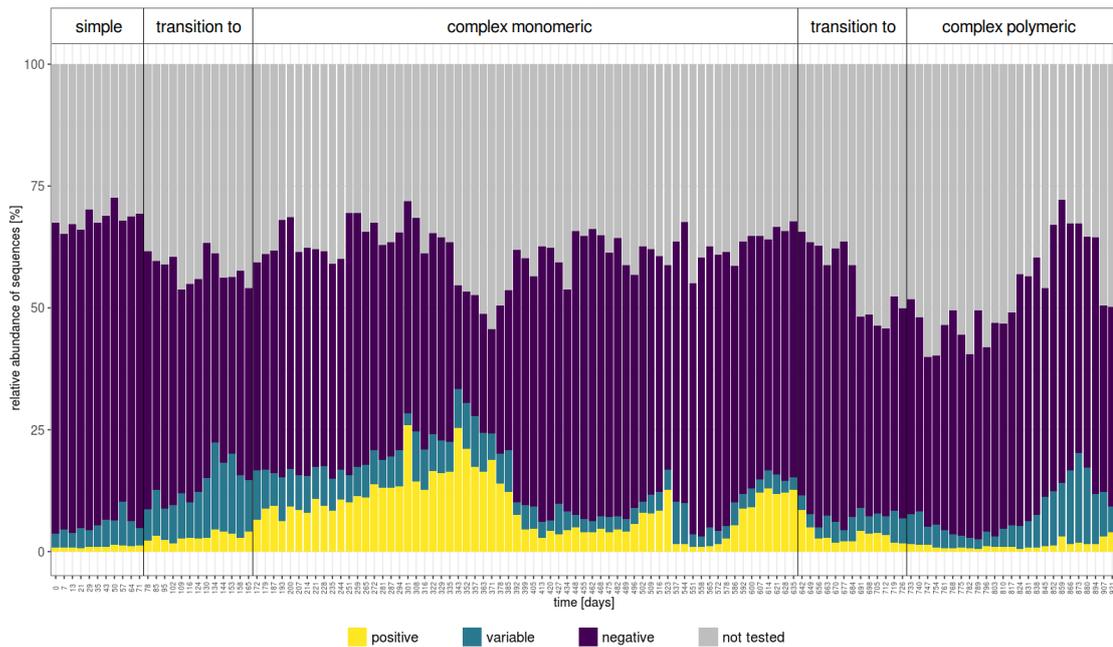


Figure VI.9 – Relative abundance of filamentous bacteria during the transition from simple monomeric, to complex monomeric and to complex polymeric influent wastewater in EPFL experiment. The functions were attributed to taxa defined at genus level based on the functional information provided on MiDAS browser.

#### VI.2.3 Filamentous bacteria

Filamentous bacteria are often held responsible for deterioration of the settling properties of activated sludge or AGS [Liu and Liu, 2006, Martins et al., 2011, Nielsen et al., 2000]. Indeed overgrowth of filamentous bacteria were often linked with bulking episodes in activated sludge. Yet, filamentous bacteria are generally present in reasonable proportions in AGS and they may favor granule formation and stability by providing a structure to which the other bacteria can attach [Gonzalez-Gil and Holliger, 2014, Liu et al., 2005]. This is confirmed by the results collected in this study. In both experiments, the increased proportions of filamentous bacteria were not related to impaired settling properties.

### VI.3 Granulation and settling properties

In both experiment, the AGS treating polymeric wastewater had significantly lower settling properties and higher proportions of flocs compared to the AGS treating monomeric wastewater. In the Eawag experiment, the evolution of the size fraction of the sludge was very similar during the start-up of AGS treating synthetic and municipal complex polymeric wastewaters. During this experiment, the biomass was mainly composed of small granules with a size between 0.25 mm to 0.63 mm. The proportion of granules remained above 50 % in the reactor treating raw wastewater, while it was more fluctuating in the two others. Yet, the settling

properties, characterized by the sludge volume index (SVI) were positively correlated to the amount of diffusible carbon in the influent wastewater.

At EPFL, the low SVI measured after the first transition from AGS fed with simple to complex monomeric wastewater showed that excellent settling properties of AGS can be maintained with a high proportion of fermentable compounds (glucose and amino acids) in the influent wastewater. On the other hand, the introduction of polymers (up to 33 % of the COD) totally impaired the granulation of the sludge, which stopped producing new granules for more than two months. After this adaptation period, the AGS resembled the one observed at Eawag with a proportion of flocs estimated around 30 % and a  $SVI_{10}$  around  $80 \text{ ml g}^{-1}$ .

The negative influence of polymeric compounds in the wastewater on the AGS settling properties have already been highlighted. It was often attributed to the leakage of COD in the aerobic phase resulting from non-hydrolyzed polymers [de Kreuk et al., 2010, deBeer and Stoodley, 1995]. The COD available with aerobic conditions can benefit the growth of filamentous bacteria. In small proportions, they are thought to provide a backbone for granules, thus being part of the granulation process [Bin et al., 2015, Gonzalez-Gil and Holliger, 2014]. Yet their overgrowth is hold responsible of filamentous outgrowth at the surface of the granules, floc formation and the resulting deterioration of settling properties [Morgenroth et al., 1997, Pronk et al., 2015a]. However, the presence of filamentous bacteria is certainly not the only cause for the lower settling properties observed with the AGS treating polymeric wastewater compared to the ones treating monomeric wastewater. Indeed, the bacterial communities of flocs were not significantly enriched with filamentous bacteria. The larger differences between the proportions of filamentous bacteria in flocs and granules were observed when the proportion of flocs was the lowest (stable state 2, 4 and 5, R1). Moreover, the proportion of known filamentous bacteria were detected in low amount ( $< 10\%$ ) in the sludge treating polymeric wastewater (Figure VI.9 and VI.7), while it was sometimes higher than 20 % in the AGS treating simple wastewater showing excellent settling properties.

Like other biofilms, granules are mainly composed of microorganisms and the extracellular polymeric substances (EPS) they produce [Liu et al., 2004]. Some bacteria may have a higher capability than others to produce EPS and therefore contribution to granulation may depend on the bacterial populations [Li et al., 2014]. A brief overview of the most complete reconstructed genomes from the AGS collected during the EPFL experiment tends to confirmed that the main bacteria present in the sludge treating monomeric or polymeric wastewater, carry genes potentially coding for proteins involved in biofilm formation.

Therefore the different propensities of the AGS to constitute dense biofilm via the production of EPS is likely regulated by external factors. The abundance of readily available COD was shown to promote EPS formation in numerous cultivated bacteria [Burne, 1998, Jefferson et al., 2004]. Conversely, COD limitations promote planktonic mode, allowing the bacteria to move and find other food sources [Jefferson, 2004]. High amounts of COD in the influent wastewater favor granulation and excellent settling properties [Cetin et al., 2018, Nancharaiah

and Reddy, 2018, Wang et al., 2009]. Increasing the volume exchange ratio and therefore the rate of incoming COD is recognized as a mean to promote granulation with low strength polymeric wastewater [Wang et al., 2009]. In the experiments presented in this thesis, AGS with very similar communities (stable states R3 and R4, 4, 5 and 6) had distinct settling properties depending on the quantity of diffusible COD provided by the influent wastewater, showing that the composition of the microbial community is not the only factor influencing the settling properties of the sludge.

Hence, granulation is proposed to be partly linked to the quantity of COD taken up by the EPS producing bacteria. Quorum-sensing could be involved in this regulation, as it was shown to play a role in biofilm formations [Merritt et al., 2003, Sakuragi and Kolter, 2007, Wang et al., 2012, Xia et al., 2012]. Putative quorum-sensing related genes were found in the genomes of bacteria present in the EPFL experiment.

### VI.4 Evolution of the PAO microdiversity

Based on the v1-v2 regions of the 16s rRNA gene sequences, 59 OTU (97 %) related to *Ca. Accumulibacter*, 17 to *Tetrasphaera* and 12 to *Dechloromonas* were found in the AGS collected during the EPFL experiment. Several of the full-length distinct sequences from the whole metagenome sequencing or reference genomes had identical v1-v2 regions, therefore the richness based on the partial 16s rRNA gene sequences may be underestimated. The analysis of P metabolism-related genes confirms that other genes, such as *ppk1*, can discriminate two *Ca. Accumulibacter* populations carrying the same 16s rRNA gene sequences [de Kreuk et al., 2007, He et al., 2007].

The high diversity of *Ca. Accumulibacter* likely complexified the reconstruction of their genomes even with the PacBio data. Even with a lower coverage, the genome of *Dechloromonas* was easily assembled in one contig, while those related to *Ca. Accumulibacter* were in multiple contigs and several contigs and the binning did not succeed to produce complete and not contaminated genomes for all the *Ca. Accumulibacter*-related MAG. The individual inspection of the contigs may help to improve the binning and get more complete MAG. Yet, the only complete genome available for *Ca. Accumulibacter* was reconstructed from a highly enriched sludge [Martin et al., 2014]. Other studies have assembled *Ca. Accumulibacter* genomes from metagenomes, but these genomes are fragmented, incomplete and generally contains contaminations [Mao et al., 2014, Skennerton, 2012]. The diversity of *Ca. Accumulibacter* may not be the only factor complexifying its assembly. Their genome likely contains long repeats that are difficult to resolve [Schmid et al., 2018]

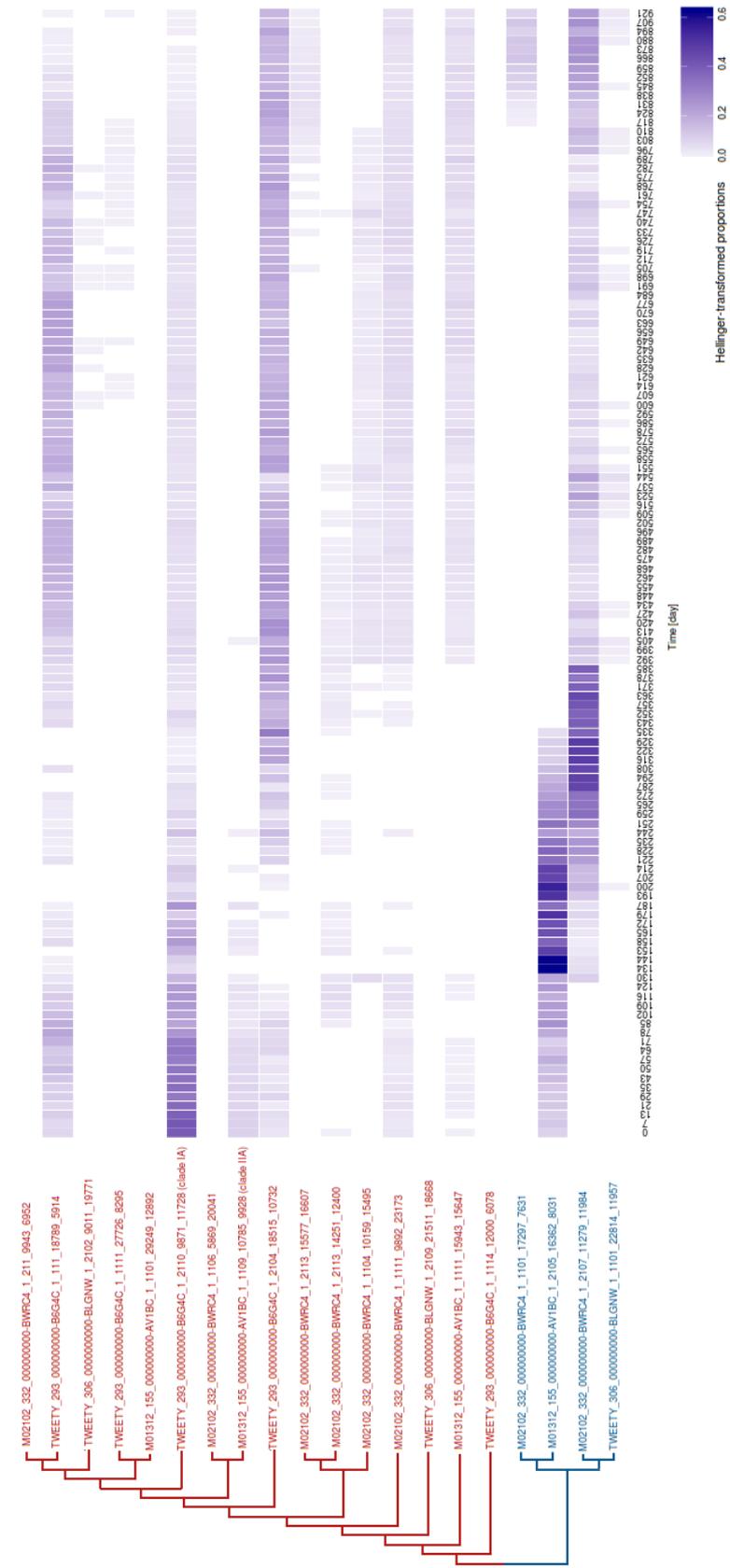


Figure VI.10 – Hellinger transformed proportions of *Ca. Accumulibacter* (in red) and *Tetrasphaera* (in blue) related OTU during the transition from simple to complex monomeric (day 48 to 167) and to complex polymeric wastewater (day 634 to 729). The phylogenetically close OTU were merged for the readability of the graph.

The repartition of the PAO into different clusters (groups of phylogenetically related OTU) across the EPFL experiment is presented in Figure VI.10. For the two genera, *Ca. Accumulibacter* and *Tetrasphaera*, this repartition evolved during the experiment. At the beginning of the experiment, when the AGS was treating simple wastewater, a cluster related to *Ca. Accumulibacter* class IA was relatively more abundant than the others, but its proportion decreased with the transition to complex monomeric wastewater. The cluster related to clade IIA was present mainly at the beginning of the experiment. These two clades along with clade IIC were detected in the AGS collected few months before the EPFL experiment [Leventhal et al., 2018]. It is possible that one of the clusters presented here is related to clade IIC, but based on the phylogeny of *ppk1* and *ppk2* homologues detected in the whole metagenome sequences, no evidence of the presence of clade IIC was found. Most of the clusters detected at the beginning of the experiment were also present in the AGS at the end of the experiment, with fluctuations in between. Yet, the dominant cluster was not the same before and after the changes of wastewater. Such fluctuations of different clades have been reported [Flowers et al., 2013]. It is however difficult to identify the causes of these fluctuations. The composition of the carbon source may have played a role, either directly [Qiu et al., 2019] or indirectly by selecting different flanking populations. A bacteriophage attack can have played a role by selectively infecting the strains having non-adapted defense mechanism [Skennerton et al., 2011]. The aeration strategy during the aerobic phase was changed from constant aeration to alternated aeration and idle periods of 10 minutes from day 48 on. If the different *Ca. Accumulibacter* clades present in the AGS had different denitrification capabilities, this change may also have played a role in the new repartition *Ca. Accumulibacter* clusters. Kim et al. [2013] noticed a shift from a *Ca. Accumulibacter* population with a majority of clades IIA, IA and IF to a majority of clades IIC, IA and IIF after the introduction of an anoxic phase between the anaerobic and oxic phases. It is also possible that flocs and granules do not have the same repartition of *Ca. Accumulibacter* clades [Barr et al., 2016].

Based on previous studies and on the results presented in chapter V, it is likely that the bacteria in the different clusters have different metabolism [Flowers et al., 2013, Peterson et al., 2008], in particular regarding to P-accumulation, carbon uptake [Gonzalez-Gil and Holliger, 2014, Qiu et al., 2019] and denitrification [Flowers et al., 2009, Kim et al., 2013, Oehmen et al., 2010]. In this study, correlation between the main taxa and the reactor performance were computed, giving few significant results. With genera such as *Ca. Accumulibacter* it could be worth to compute these correlations at the OTU level. Yet, the software used to cluster the sequences in OTU (cd-hit) does not actualize the cluster center during the algorithm and may produce clusters for which the composition can vary a lot with similar inputs and may not reflect the biology behind. New clustering algorithm which recompute the center of the clusters during the process could be tested to improve the stability of the clusters [James et al., 2018].

In comparison with *Ca. Accumulibacter*, *Tetrasphaera* was less diverse but still composed of 4 cluster of 4-5 OTU each. Up to 156 OTU (97 %) were related to *Tetrasphaera* were retrieved in EBPR sludge, but generally one of them is significantly more abundant than the others [Mielczarek et al., 2013, Nguyen et al., 2011]. The cluster present at the beginning of the

experiment became very abundant following the abrupt decrease of *Ca. Accumulibacter* around day 134, and then was progressively replaced by another cluster. Two other clusters appeared during the experiment, one did three months after the transition to polymeric wastewater. Bacteria from the genus *Tetrasphaera* have shown hydrolyzing capabilities [Xia et al., 2008a]. It is not known yet if this ability is common to all the *Tetrasphaera* strains. Some *Tetrasphaera* species seem to harbour slightly different metabolisms, for example regarding PHA synthesis [Kristiansen et al., 2013].

### VI.5 Nitrogen-removal according to the wastewater type

Ammonium-removal was, in average, above 90 % in all the reactors of the two experiments whatever the wastewater type. It was not the case of total nitrogen (TN) removal due to the presence of variable concentrations of nitrate in the effluent. In the EPFL experiment, few simultaneous nitrification-denitrification (SND) was noticed during the aerobic phase and the TN-removal was around 50 %. The full aeration was therefore replaced by intermittent aeration. It extended a little the length of the cycles but improved substantially the TN-removal (up to 80 %). During the transition from simple to complex monomeric wastewater, TN-removal performance was impaired. This coincided with the abrupt decrease of *Ca. Accumulibacter* in the AGS and the drop of P-removal performance. As mentioned above, it is likely that most *Ca. Accumulibacter* present in BNR sludge are implicated in denitrification. Therefore, the sudden decrease of *Ca. Accumulibacter* could be a cause of the impaired denitrification. The total COD which decreased slightly during this transition can be another valid explanation for the lower TN-removal. After the transition, the COD of the wastewater was increased to 600 mgO<sub>2</sub> l<sup>-1</sup> and both good P- and TN-removal were recovered.

In both experiments, the TN-removal was slightly lower with the polymeric wastewater in comparison with the monomeric wastewater. The potential denitrification pathways detected in the metagenome assembled genomes (MAG) of the sludge treating monomeric and polymeric wastewater, along with other studies [Wang et al., 2014b] show that denitrification capabilities are widespread across the bacterial kingdom. Numerous putative denitrifying bacteria were found in the AGS treating monomeric and polymeric wastewater. Moreover, the two reactors treating municipal wastewater, had very similar bacterial communities but the AGS treating clarified wastewater had a lower TN-removal (45 %) than the AGS treating raw wastewater (63 % for run #2). This shows that the combination of operational conditions and the bacterial community composition acts on the TN-removal performance of the AGS, a sufficient COD content being an important factor. With polymeric wastewater, part of the COD may not be available for the bacteria that participate to the denitrification and are located below the surface of the granules. The denitrification, which requires COD, is therefore lower than with a wastewater containing only diffusible COD. If a sufficient amount of denitrifying PAO are present in the AGS, P- and TN- removal are combined allowing to save COD [Kim et al., 2013]. If the denitrifying capabilities of the classical PAO *Ca. Accumulibacter* were investigated in the reconstructed genomes, it was not the case for *Tetrasphaera* since the chosen extraction

method chosen for PacBio sequencing was probably too gentle to extract efficiently the DNA of gram+ bacteria such as *Tetrasphaera*. Several contigs related to *Tetrasphaera* in the hybrid assembly and further investigation may provide additional information about this fermentative PAO in the AGS collected at EPFL.

### VI.6 Functional groups in flocs and granules

The relative abundances of different functional groups in flocs and granules bacterial communities of AGS at stable states in the EPFL and the Eawag experiments are shown in Figure VI.11. Filamentous bacteria were slightly enriched in the flocs only in some stable states corresponding to monomeric wastewater (Stable state 2,4 and R1). In the stable states corresponding to polymeric wastewater they were a little more abundant in granules. It is possible that some filamentous bacteria were not taken into account in this analysis because they were not characterized at the genus level or because their tendency to form filaments have not been elucidated yet. Even so, these results suggests that filamentous bacteria were not the main cause for the important proportions of floc found in the systems treating polymeric wastewater.

The GAO were a little more abundant in the granules in all the stable states except stable state 6. The repartition of the granules and flocs during this stable state was a particular case because no new granules were forming. The PAO were a little enriched in the granules collected during the stable states of the EPFL experiment with the exception of stable state 2. The proportion of *Ca. Accumulibacter* in the AGS collected during this stable state were particularly low, possibly due to a bacteriophage attack. This may have disrupt the normal distribution of PAO among flocs and granules. At Eawag, there were no clear tendencies for the repartition of PAO in flocs and granules.

Fermenting bacteria were a little more abundant in granules in the complex polymeric wastewater at Eawag. No clear differences between the proportions of fermenting bacteria in flocs or granules were found in the other stable states.

## Chapter VI. General discussion

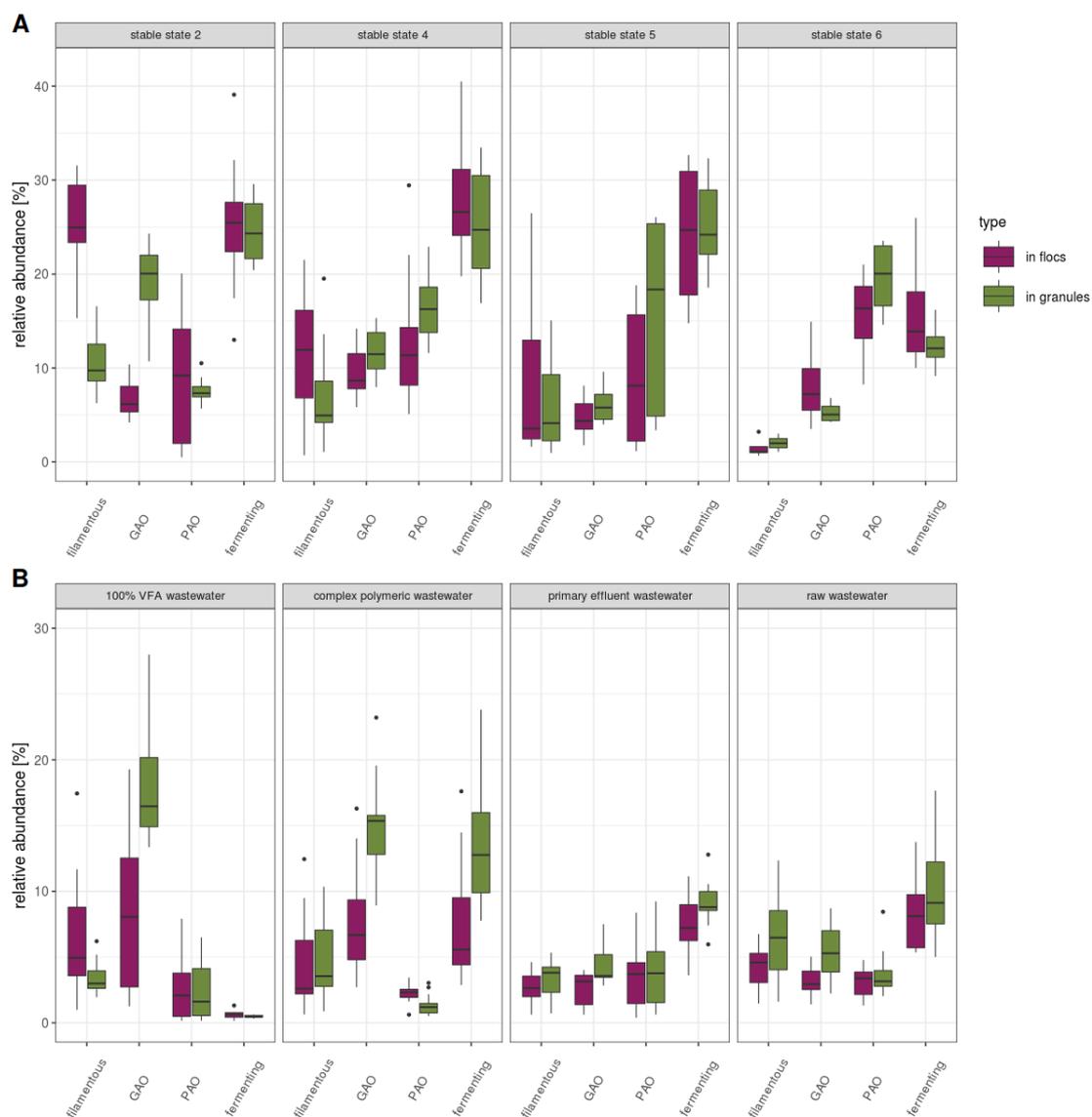


Figure VI.11 – Relative abundance of functional groups in flocs and granules in the different bacterial stable states during the LBE (A) and Eawag (B) experiments. The functions were attributed to taxa defined at genus level based on the functional information provided on MiDAS browser.

## VII Concluding remarks and perspectives

The two different approaches used to study the influence of wastewater composition on the AGS bacterial communities showed that they are very different in AGS treating simple and complex wastewater. Fermenting bacteria of the phylum Saccharibacteria and the class Actinobacteria were clearly underrepresented in AGS treating simple synthetic wastewater compared to complex wastewater (synthetic or municipal). Less important differences were observed in the abundance of other functional groups between simple and complex wastewater. Yet big differences were observed regarding their composition. In the simple wastewater, the PAO guild was mostly composed of *Ca. Accumulibacter*, whereas it was more diverse with the complex wastewater with higher proportions of *Tetrasphaera*. At the beginning of the EPFL experiment, a small proportion of *Tetrasphaera* in the AGS treating simple wastewater likely allowed a fast recovery of the P-removal despite the sharp decrease of *Ca. Accumulibacter* during the transition to complex monomeric wastewater. In AGS systems treating complex wastewater, the presence of these two PAO occupying slightly different ecological niches increased the functional redundancy and likely stabilizes the stability of P-removal.

The GAO were found in slightly higher proportions in the simple wastewater than in the complex wastewater. The guild of GAO was almost exclusively composed of Gammaproteobacteria mostly from the Competibacteraceae family in the AGS treating simple wastewater, whereas the Gammaproteobacteria were in minority in the guild of GAO in the AGS treating complex wastewater. In those AGS, Actinobacteria, in particular the fermenting GAO *Micropruina*, were dominant.

The AGS treating monomeric wastewater were mostly composed of large granules whereas AGS treating polymeric wastewater contained 20 to 40 % of flocs and mostly small granules as a consequence of non diffusible COD being hydrolyzed and consumed at the surface of the biofilm (granules or flocs). Few individual trends were noticed when comparing the bacterial populations in flocs and granules. The GAO, in particular *Ca. Competibacter* were slightly enriched in the granules. Similarly, the nitrite oxidizing bacteria *Nitrospira* and *Ca. Nitrotoga* were a little more abundant in the granules. On the other hand, some filamentous bacteria (eg. *Zoogloea*, *Thiothrix*) were slightly enriched in flocs likely due to their location at

the surface of the granules making them more exposed to erosion. Yet the overall bacterial composition in flocs and granules was globally very similar for a specific AGS. Flocs seem to be constitutive of AGS systems treating low-to-medium strength polymeric wastewater. It is likely that they are maintained in these systems due to a competitive advantage regarding polymeric substrate. The few differences in the bacterial communities of flocs and granules suggest that this advantage is mostly due to their different physical structure, which for instance allow them to capture the particules to hydrolyze more efficiently than granules. A moderate amount of flocs in AGS systems treating polymeric wastewater should therefore not be seen as a flaw as they probably have a specific role to play in the nutrient removal process (e.g., C, N, P). It would be worth to investigate this role further, e.g. by measuring specific metabolic activities of flocs and granules.

Denitrification was more efficient with monomeric wastewater compared to polymeric wastewater even though putative denitrifying bacteria were identified in the AGS corresponding to the different wastewater types. The smaller sizes of the granules and the limited diffusibility of polymeric COD were identified as the main causes for the lower denitrification observed with polymeric wastewater. This study illustrates that AGS with slightly different microbial communities can have similar settling properties, size distributions and nutrient removal performances, as it was the case for the AGS treating complex polymeric synthetic and raw municipal wastewater. On the opposite, the different behavior of the AGS cultivated with clarified and raw wastewater in terms of settling properties and nutrient removal performance, when their bacterial communities was very similar, confirms that the bacteria required a certain amount of organic COD to function optimally. Indeed, both granulation and denitrification were enhanced by a higher total COD content.

The lower settling properties observed for AGS cultivated with polymeric wastewater were not related to high proportions of filamentous bacteria, which were present in relatively low proportions (< 10 %) in the AGS independently of the wastewater type. Moreover granulation was observed with low proportions of the classical GAO (e.g., *Ca. Competibacter*) and PAO (e.g., *Ca. Accumulibacter*) suggesting that the capability to form granules is not limited to those bacteria. The fermentative GAO *Micropruina* was very abundant during the start-up of AGS from EBPR activated sludge and likely participated to the granule formation. If some bacteria contribute more than other to biofilm formation, for example by producing high amounts of extrapolymeric substances, the shift from planktonic to biofilm mode, is likely triggered by external factors via quorum-sensing or other regulation systems. The amount of available COD can set the biofilm mode on or off and could therefore be a major factor for granulation.

AGS treating municipal and complex polymeric synthetic wastewater were very similar regarding the bacterial communities, the biomass size distribution and settling properties. The latter wastewater type constitute therefore a more realistic surrogate for real wastewater than simple synthetic wastewater in laboratory studies.

---

The analysis of the whole metagenomes from AGS treating different wastewater types confirms that *Ca. Accumulibacter* can be very diverse in a single AGS and that the distribution of the different clades may be related to the wastewater type. The high microdiversity of *Ca. Accumulibacter*, the presence of long repeat regions (e.g., CRISPR-Cas system) and potentially important exchanges of genetic material between the members of different clades likely complexifies the binning of the contigs for this genus and would be worth to be studied further.

Several microbial populations found in AGS have not yet been characterized at the genus level and thus their role in granulation and nutrient-removal is unknown. The metagenomic study of the four AGS samples already provided some indications about the putative functions of the main populations present in these samples and can be carried on further, in particular by inspecting the contigs obtained with the hybrid assembly of PacBio and Illumina sequences which seem to contain more information on different taxonomic groups than the PacBio assembly alone. A special effort can be applied on the resolution of long nearly identical repeated regions in the diverse recovered genomes. In particular the CRISPR regions can be assembled by using specific tools and their inspection can potentially provide information about the past of the different clones. Yet the obviously better quality assembly obtained from the PacBio long-reads and the nearly complete genomes extracted from this assembly showed that this technology is well suited for the study of low-to-medium complexity metagenomes.

The metagenomic information collected provided information about the putative functions of the different bacterial populations in the studied AGS. It also provides a template for future meta-transcriptomic analysis. Altogether they can be integrated in an holistic approach of the AGS system based on the discovery of metabolic and cell signaling networks, an approach nowadays called systems biology. This will hopefully lead to a better understanding of the overall bacterial community interactions and functions allowing to predict more finely how the AGS system will evolve depending on the wastewater composition and the operational conditions. Such knowledge is essential to optimize wastewater treatment with AGS by adapting the infrastructures and process operations.

The development and increased accessibility of molecular methods have already enabled examining the composition of AGS bacterial communities and gain a better understanding of their functioning. Yet, many bacteria are still poorly characterized and little is known about the interactions of the different taxa inside the granules. Moreover, the majority of research were performed with AGS treating simple wastewater, which tends to harbor different bacterial communities, and different performance. There is therefore still a lot to discover on the individual and collective roles of bacteria in AGS systems treating real wastewater in order to optimize further this process.



# A Supplementary material

## A.1 Electronic supplementary material chapter II

Table A.1 – Abundance of the phages- and bacteria-related sequences in the metagenome O1  
'O1\_bact\_phage\_genus\_abundance\_cut.csv'

Table A.2 – Correlations between phages- and bacteria-related sequences in the metagenome O1  
'O1\_corr\_bacteria\_phages.csv'

Table A.3 – P-values of the correlations between phages- and bacteria-related sequences in metagenome O1  
'O1\_corr\_bacteria\_phages\_pvalue.csv'

Table A.4 – Abundance of the phages- and bacteria-related sequences in the metagenome  
'O2\_bact\_phages\_genus\_abundance\_cut.csv'

Table A.5 – Correlations between phages- and bacteria-related sequences in the metagenome O2  
'O2\_corr\_bacteria\_phages.csv'

Table A.6 – P-values of the correlations between phages- and bacteria-related sequences in metagenome O2  
'O2\_corr\_bacteria\_phages\_pvalue.csv'

### A.2 Supplementary information chapter III

Figure A.1: Evolution of sludge morphology in R1 fed by 100%-VFA synthetic WW (A, B, C, D) after 12, 57, 93, and 190 days of operation, in R2 fed with complex synthetic WW (E, F, G, H) after 12, 34, 57, and 100 days, in R3 fed with primary effluent WW (I, J, K, L) after 7, 147, 279, and 400 days, and in R4 fed with raw wastewater (M, N, O, P, for run #1, Q, R, S, T for run #2) after 7, 22, 57, and 84 days for run #1, and after 4, 100, 163, and 212 days for run #2. Orange size bars = 200  $\mu\text{m}$ , white size bars = 500  $\mu\text{m}$ , green size bars = 1.0 mm and blue size bars = 2.0 mm. Morphology of AGS fed by 100%-VFA influent wastewater (R1) significantly differed from the morphology of the AGS fed by complex synthetic wastewater (R2) and municipal wastewater (R3, R4). AGS of R1 was composed of large, round-shaped, dense and overall homogeneous shaped aggregates that dominated overall sludge morphology (A-D). Almost no flocs were observed. Operational issues of carbon leakage (Figure A.4) from anaerobic to aerobic phase were the cause of filamentous outgrowth at the granules surface, visible on (D), after 190 days of operation. Visual observations indicated that AGS fed with complex influent wastewater (R2, R3, R4) were composed of both flocs and small and dense aggregates. Formation of “finger-type” granules was not observed in R2, despite the feed of particulate substances (starch). The highest complexity wastewater fed systems R3 and R4 also resulted in the most complex sludge morphology. In these reactors, aggregates with dense cores formed, but with very heterogeneous sizes and shapes. The formation of “finger-type” granules was observed (P and T) but never over prolonged time. Large amounts of fibers or debris from the influent also accumulated in those systems (N and Q). First appearance of granules took much longer in complex wastewater fed reactors R2, R3 and R4 compared to R1. In R1 first granules were observed already after 12 days of operation (A). The formation of granules in the complex wastewater fed reactors R2, R3 and R4 was much slower in comparison. In R2 it took 30 days until first round-shaped and dense granules were observed (F). Granules only appeared after 100 days of operation in R4 run #2 and 279 days in R3. No large granules were observed in R4 during run #1, which was shorter than 100 days.

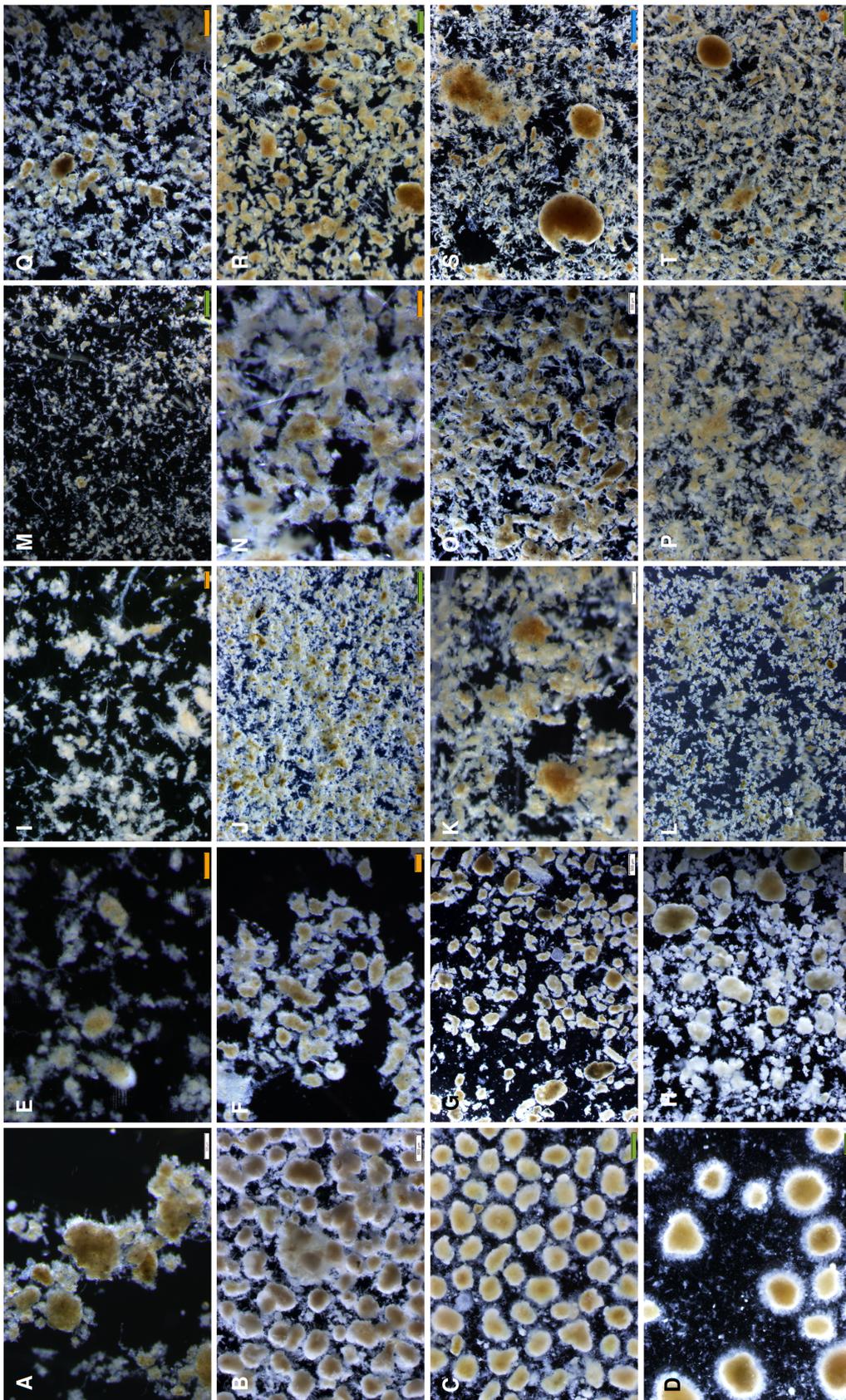


Figure A.1

## Appendix A. Supplementary material

---

Table A.7 – Recipe of 100%-VFA synthetic wastewater and complex synthetic wastewater for reactors R1 and R2, the concentrations are in  $\text{g l}^{-1}$ . The recipes only provide C and N species to the wastewater preparation. The recipes were prepared in 20-fold concentration, to provide total COD and TN concentrations of  $600 \text{ mg COD l}^{-1}$  and  $52 \text{ mg TN l}^{-1}$ , respectively

Component	100%-VFA synthetic WW	complex synthetic WW
Na acetate· 3H <sub>2</sub> O	12.8	4.3
Na propionate	4.8	1.6
(NH <sub>4</sub> )Cl	3.2	1.1
CaCl <sub>2</sub> · H <sub>2</sub> O	0.35	0.35
MgSO <sub>4</sub>	0.33	0.33
KCl	0.66	0.66
Glucose/Dextrose	-	1.9
Starch	-	1.4
Peptone	-	1.6
Alanine	-	0.27
Arginine	-	0.26
Aspartic acid	-	0.40
Glutamic acid	-	0.29
Glycine	-	0.45
Leucine	-	0.16
Proline	-	0.19

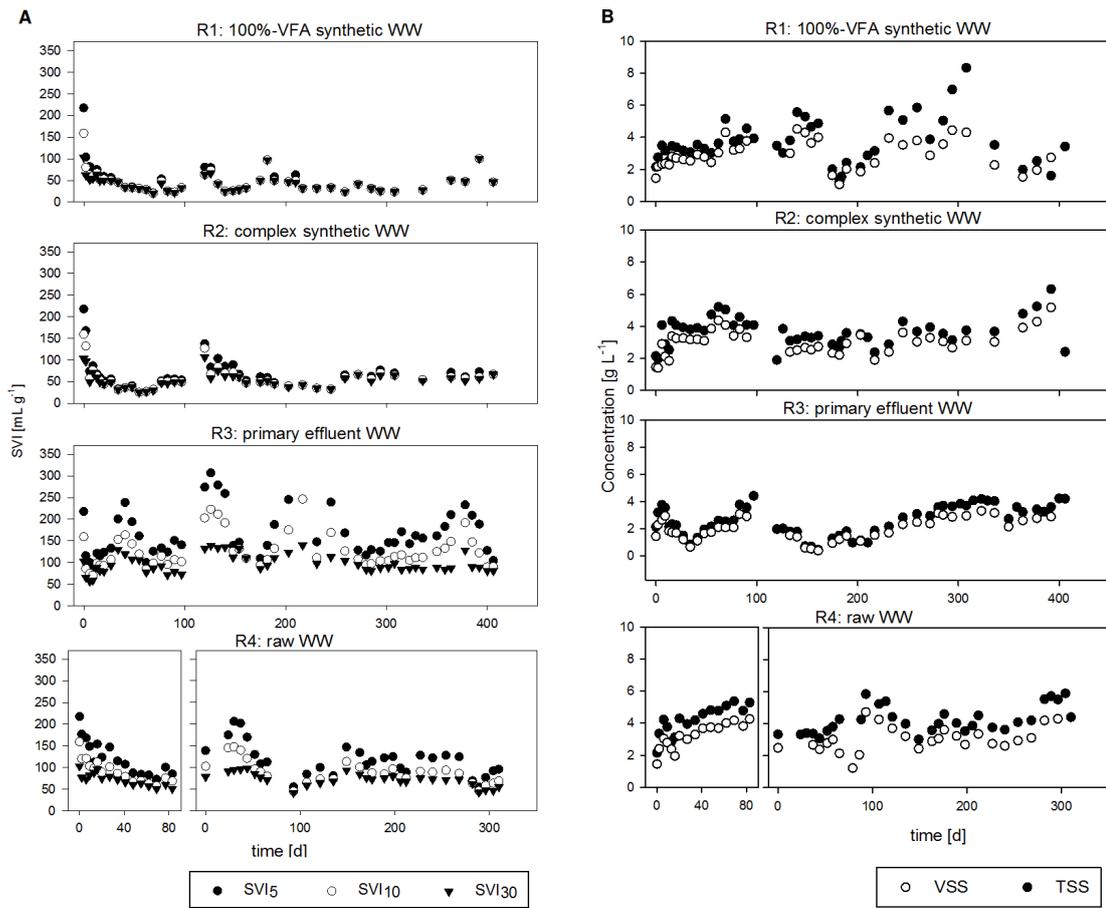


Figure A.2 – Evolution of SVI<sub>5</sub>, SVI<sub>10</sub> and SVI<sub>30</sub> (A) and evolution of TSS and VSS (B) in R1, R2, R3, R4 run#1 and run#2.

## Appendix A. Supplementary material

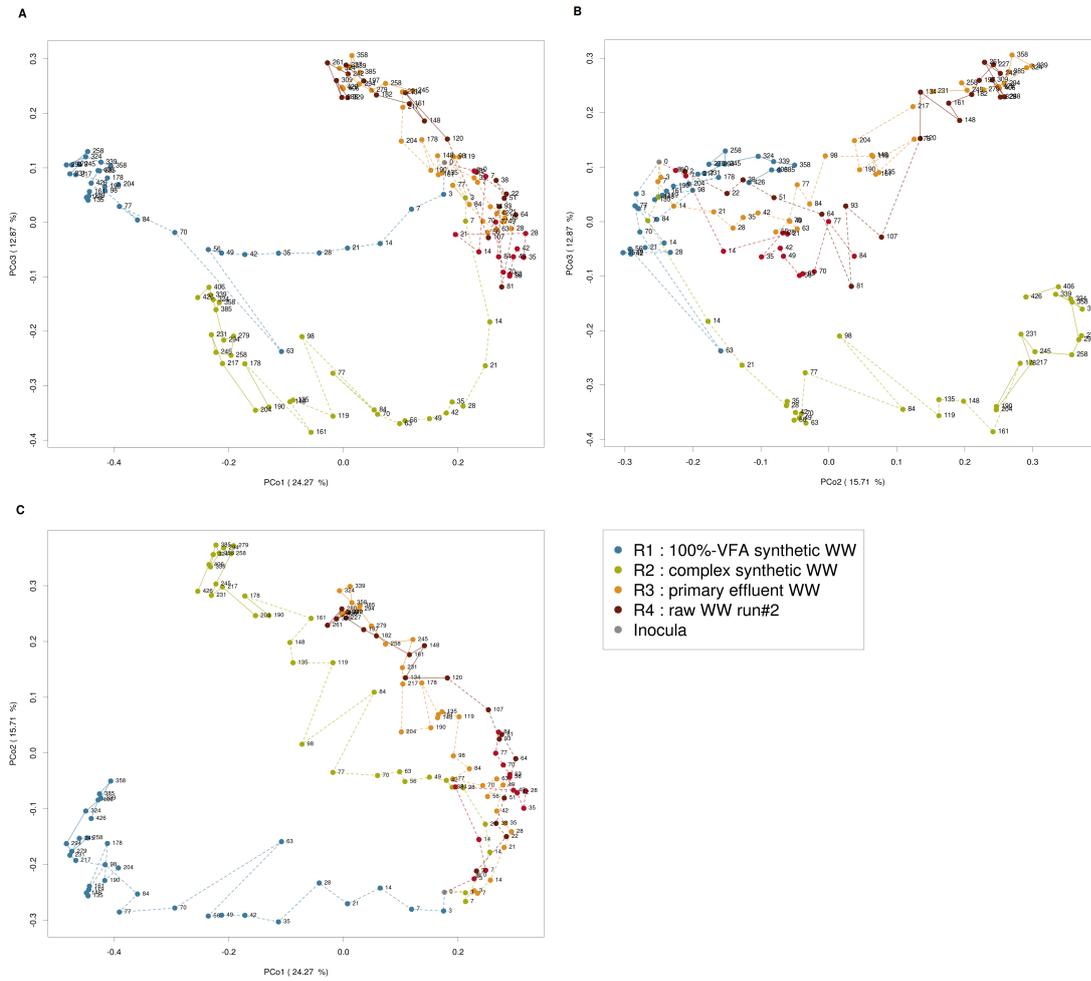


Figure A.3 – Principal component analysis of microbial communities structure evolution during the Eawag experiment on the 1<sup>st</sup> and 3<sup>rd</sup> axis (A) on the 2<sup>nd</sup> and 3<sup>rd</sup> axis (B) and on the 1<sup>st</sup> and 2<sup>nd</sup> axis (C).

## A.2. Supplementary information chapter III

Table A.8 – P-values of the t-test comparing the relative abundance of the main genera in the floccular and the granular fraction of the sludge.

Genus	R1	R2	R3	R4
480_2 (f)	$1.8 \times 10^{-1}$	$9.7 \times 10^{-2}$	$4.8 \times 10^{-1}$	$5.7 \times 10^{-1}$
<i>Ca. Accumulibacter</i>	$9.8 \times 10^{-1}$	$4.8 \times 10^{-1}$	$8.8 \times 10^{-1}$	$8.6 \times 10^{-1}$
<i>Azoarcus</i>	$5.6 \times 10^{-1}$	$8.5 \times 10^{-1}$	$1.7 \times 10^{-1}$	$4.9 \times 10^{-2}$
<i>Ca. Competibacter</i>	$4.6 \times 10^{-2}$	$9.0 \times 10^{-3}$	$4.5 \times 10^{-1}$	$2.8 \times 10^{-2}$
<i>CPB_CS1</i>	$2.0 \times 10^{-1}$	$4.0 \times 10^{-1}$	$6.5 \times 10^{-1}$	$2.1 \times 10^{-1}$
<i>CYCU-0281</i>	$6.5 \times 10^{-1}$	$4.4 \times 10^{-1}$	$6.8 \times 10^{-1}$	$9.7 \times 10^{-1}$
<i>Flavobacterium</i>	$9.1 \times 10^{-1}$	$1.2 \times 10^{-1}$	$5.7 \times 10^{-2}$	$4.4 \times 10^{-1}$
<i>Kouleothrix</i>	$4.8 \times 10^{-1}$	$6.3 \times 10^{-1}$	$7.2 \times 10^{-2}$	$4.3 \times 10^{-2}$
<i>Micropruina</i>	$6.3 \times 10^{-2}$	$3.0 \times 10^{-4}$	$3.9 \times 10^{-3}$	$6.1 \times 10^{-3}$
<i>Nitrospira</i>	$2.2 \times 10^{-6}$	$5.6 \times 10^{-2}$	$5.3 \times 10^{-1}$	$5.0 \times 10^{-8}$
<i>P58</i>	$8.9 \times 10^{-1}$	$9.4 \times 10^{-1}$	$1.9 \times 10^{-1}$	$8.3 \times 10^{-1}$
<i>Rhodobacter</i>	$3.5 \times 10^{-4}$	$3.0 \times 10^{-1}$	$9.8 \times 10^{-1}$	$4.9 \times 10^{-1}$
<i>Runella</i>	$2.8 \times 10^{-2}$	$3.3 \times 10^{-1}$	$9.8 \times 10^{-1}$	$4.9 \times 10^{-1}$
Saccharibacteria (p)	$5.7 \times 10^{-2}$	$2.8 \times 10^{-1}$	$6.1 \times 10^{-1}$	$2.6 \times 10^{-1}$
<i>Tetrasphaera</i>	$7.1 \times 10^{-1}$	$6.7 \times 10^{-1}$	$6.3 \times 10^{-1}$	$5.0 \times 10^{-1}$
<i>Trichococcus</i>	$2.9 \times 10^{-3}$	$5.9 \times 10^{-1}$	$1.8 \times 10^{-1}$	$3.9 \times 10^{-3}$
Xanthomonadaceae (f)	$6.8 \times 10^{-1}$	$2.2 \times 10^{-3}$	$1.4 \times 10^{-1}$	$7.8 \times 10^{-2}$
<i>Zoogloea</i>	$1.1 \times 10^{-2}$	$2.0 \times 10^{-4}$	$3.8 \times 10^{-1}$	$1.6 \times 10^{-1}$
<i>Terrimonas</i>	$2.7 \times 10^{-4}$	$5.8 \times 10^{-1}$	$4.8 \times 10^{-1}$	$3.1 \times 10^{-1}$
<i>Thauera</i>	$3.1 \times 10^{-1}$	$9.8 \times 10^{-1}$	$6.5 \times 10^{-1}$	$8.8 \times 10^{-6}$

Table A.9 – Mantel test comparing the Bray-Curtis distance matrices of the microbial communities, the settling and size characteristics of the sludge and its nutrient-removal performance, at bacterial stable state

Data-set	Mantel statistic	significance
Microbial communities vs settling	0.6454	0.001
Microbial communities vs nutrient removal	0.0470	0.195
Settling vs nutrient-removal	0.2138	0.009

Appendix A. Supplementary material

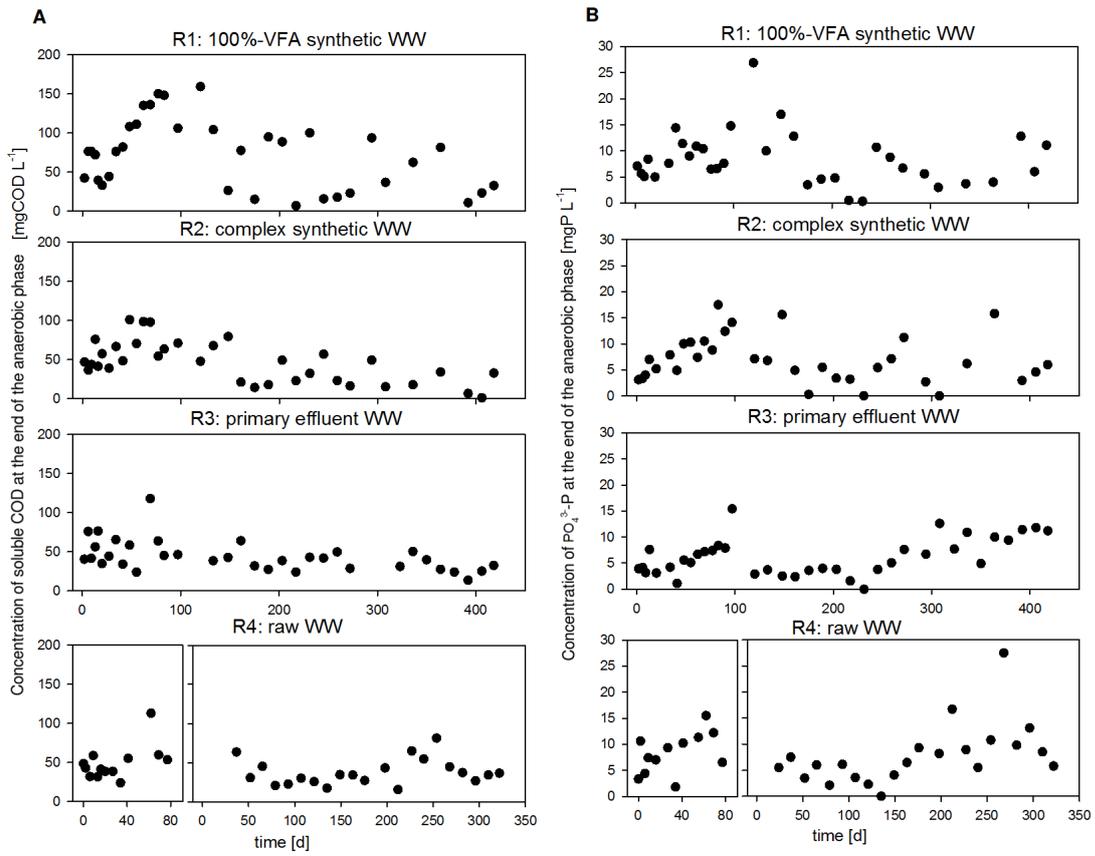


Figure A.4 – Soluble COD concentrations (A) and PO<sub>4</sub><sup>3-</sup>-P concentrations (B) after the anaerobic phase in R1, R2, R3, R4 run#1 (row 4, left) and R4 run#1 (row 4, right).

### A.2.1 Electronic supplementary material chapter III

Table A.10 – P-values of the t-test performed on the mean relative abundance of the taxa between different stable states or groups of stable states. The divergent, abundant and discriminant taxa are indicated. *'discriminant\_and\_p\_value\_t\_tests.ods'*

Table A.11 – Bray-Curtis distance matrices of the bacterial community compositions, the settling properties and size distribution of the sludge and its nutrient-removal performances at stable state. *'bray\_distance\_matrices\_stable\_state.xlsx'*

Table A.12 – Average relative abundance of the bacterial taxa (at genus level) in the inoculum and the four reactors during stable state. *'genera\_mean\_abundance\_at\_stable\_state.csv'*

### A.3 Supplementary information chapter IV

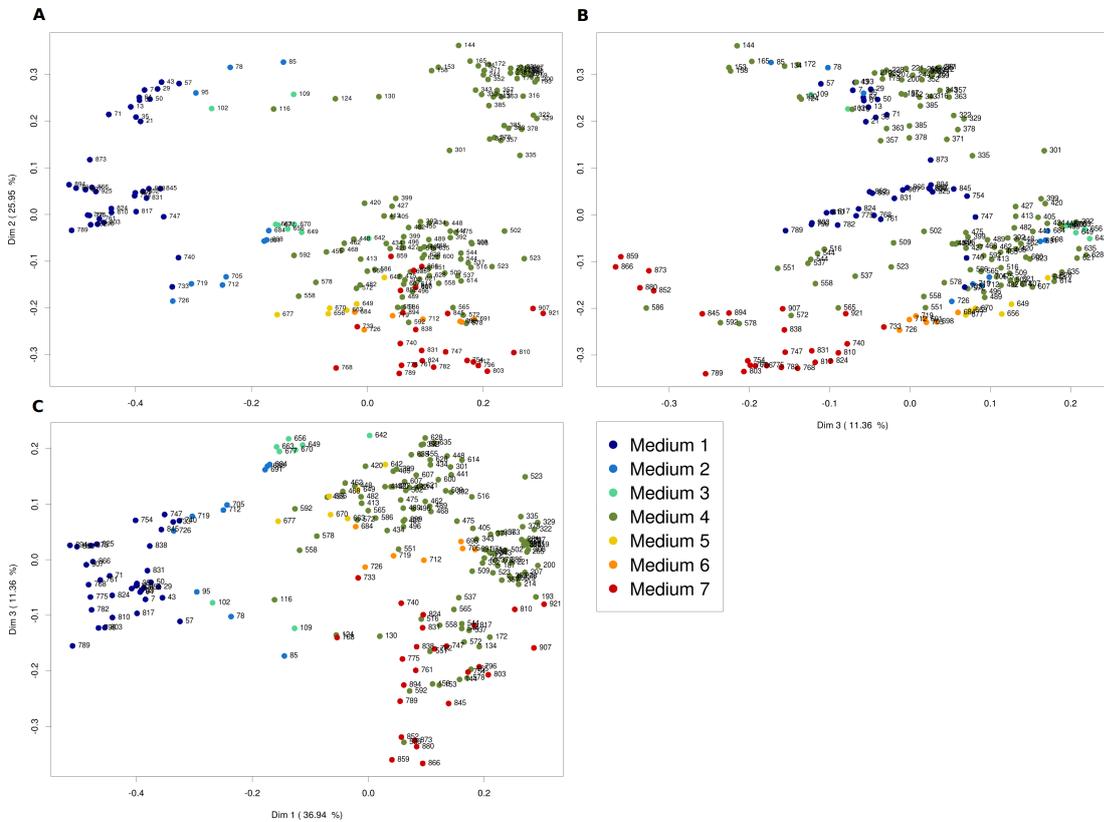


Figure A.5 – Principal coordinate (PCoA) plot based on the Bray-Curtis distance matrix of the bacterial communities (at genus level) in the EPFL experiment. Only the abundant taxa were considered for this analysis. The three first coordinates Dim 1, Dim 2 and Dim 3 explain 36.94%, 25.95% and 11.36% of the variance, respectively.

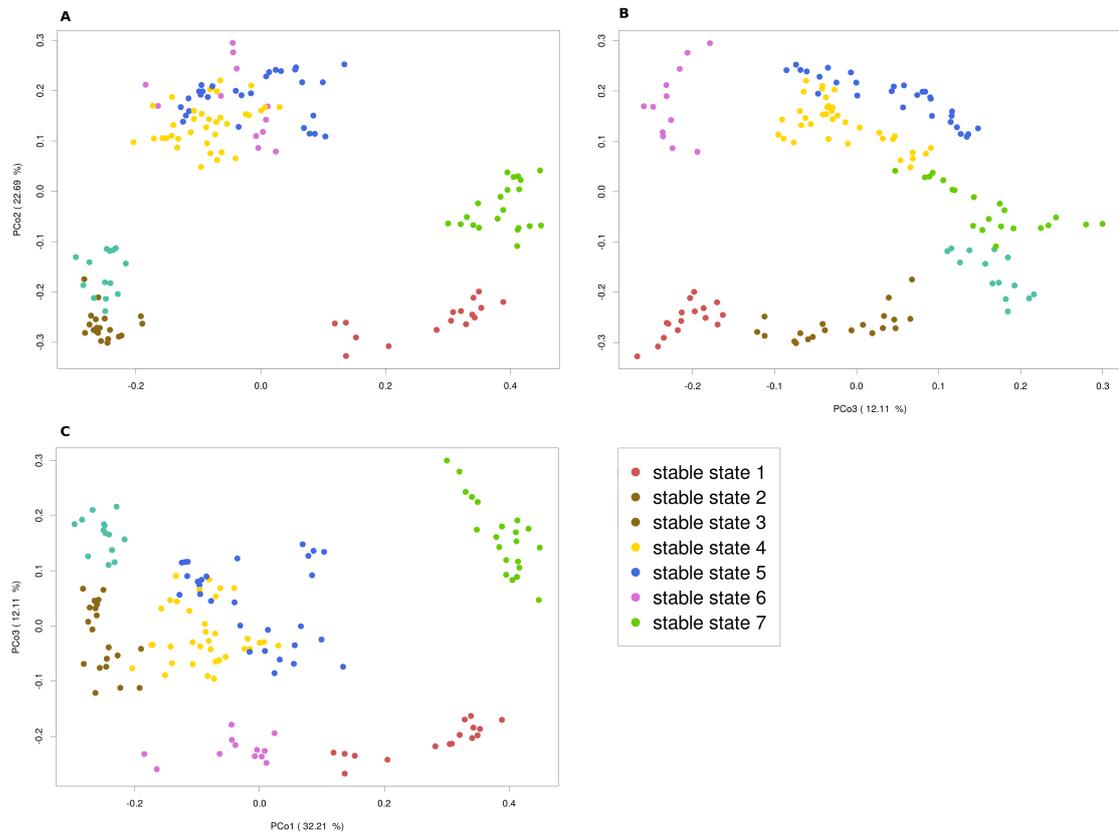


Figure A.6 – Principal component analysis of the bacterial communities (genus level) at stable states in the EPFL experiment. The two first coordinates Dim 1, Dim2 and Dim3 explain 32.21%, 22.69% and 12.11% of the variance, respectively.

Table A.13 – Recipes for the preparation of the different synthetic wastewater used in the EPFL experiment in g per 10 l.

denomination	simple			transition				complex monomeric			transition	complex polymeric
	1	1B	2	2B	3	3B	4A	4B	4C	5		
compound	formula											
C-medium												
Sodium acetate trihydrate	42.84	56.70	34.27	45.36	25.70	34.02	14.28	18.90	18.90	18.90	18.90	18.90
	C <sub>2</sub> H <sub>3</sub> O <sub>2</sub> Na *3H <sub>2</sub> O											
Sodium propionate	17.52	22.87	14.02	18.30	10.51	13.723	5.84	7.62	7.62	7.62	7.62	7.62
	C <sub>3</sub> H <sub>5</sub> O <sub>2</sub> Na											
Glucose monohydrate <sup>1</sup>	-	-	3.78	5.5	7.55	11.01	12.59	18.38	18.38	18.38	18.38	18.38
	C <sub>6</sub> H <sub>12</sub> O <sub>6</sub> * H <sub>2</sub> O											
Starch[Haybraud et al., 2017] <sup>0</sup>	-	-	-	-	-	-	-	-	-	-	2.25	4.50
	(C <sub>6</sub> H <sub>12</sub> O <sub>6</sub> ) <sub>n</sub>											
Magnesium sulfate heptahydrate	1.77	1.77	1.77	1.77	1.77	1.77	1.77	1.77	1.77	1.77	1.77	1.77
	MgSO <sub>4</sub> * 7H <sub>2</sub> O											
Calcium chloride dihydrate	1.65	1.65		1.65		1.65	1.65	1.65	1.65	1.65	1.65	1.65
	CaCl <sub>2</sub> * 2H <sub>2</sub> O											
Potassium chloride	3.575	3.575	3.575	3.575	3.575	3.575	3.575	3.575	3.575	3.575	3.575	3.575
	KCl											
NP-medium												
Alanine	-	-	0.485	0.707	0.970	1.414	1.617	2.360	2.360	2.000	1.650	1.178
	C <sub>3</sub> H <sub>7</sub> NO <sub>2</sub>											
Arginine	-	-	0.517	0.754	1.035	1.508	1.725	2.510	2.510	2.140	1.760	1.257
	C <sub>6</sub> H <sub>14</sub> N <sub>4</sub> O <sub>2</sub>											
Aspartic acid	-	-	0.725	1.056	1.450	2.113	2.416	3.520	3.520	2.990	2.465	1.761
	C <sub>4</sub> H <sub>7</sub> NO <sub>4</sub>											
Glutamic acid	-	-	0.534	0.778	1.068	1.557	1.781	2.590	2.590	2.210	1.816	1.297
	C <sub>5</sub> H <sub>9</sub> NO <sub>4</sub>											
Glycine	-	-	0.818	1.192	1.635	2.383	2.725	3.970	3.970	3.380	2.780	1.986
	C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub>											
Leucine	-	-	0.286	0.416	0.572	0.833	0.953	1.390	1.390	1.180	0.972	0.694
	C <sub>6</sub> H <sub>13</sub> NO <sub>2</sub>											
Proline	-	-	0.342	0.498	0.684	0.997	1.140	1.660	1.660	1.410	1.163	0.831
	C <sub>5</sub> H <sub>9</sub> NO <sub>2</sub>											
Peptone	-	-	-	-	-	-	-	-	-	1.90	3.81	6.35
	NH <sub>4</sub> Cl											
Ammonium chloride	18.93	18.93	18.93	16.315	18.93	13.702	18.93	7.91	7.91	7.91	7.91	7.91
	K <sub>2</sub> HPO <sub>4</sub>											
Potassium hydrogenophosphate	7.305	7.305	7.305	7.305	7.305	7.305	7.305	7.305	7.305	7.305	7.305	7.305
	KH <sub>2</sub> PO <sub>4</sub>											
Monobasic potassium phosphate	2.855	2.855	2.855	2.855	2.855	2.855	2.855	2.855	2.855	2.855	2.855	2.855
	KH <sub>2</sub> PO <sub>4</sub>											

#### A.3.1 Electronic supplementary material chapter IV

Table A.14 – Bray-Curtis distance matrix of the bacterial communities in the AGS samples of experiment 1 and 2. *'Bray\_dist\_all\_LBE.csv'*

Table A.15 – P-values of the t-tests comparing the mean abundance of each taxa detected in the AGS of E1 and E2 in the three wastewater types (simple, complex monomeric and complex polymeric). The divergent, abundant and discriminant taxa are indicated. *'t-test\_mediums.ods'*

Table A.16 – Mean relative abundance of the taxa (genus level) in the AGS samples treating the different wastewater types (simple, complex monomeric and complex polymeric). *'mean\_stable\_states\_per\_medium.csv'*

Table A.17 – P-values of the t-tests comparing the mean abundance of each taxa detected in the AGS of E1 and E2 in the seven bacterial stable states. *'all\_stable\_states\_t\_tests.ods'*

## A.4 Supplementary information chapter V

### A.4.1 Electronic supplementary material chapter V

Table A.18 – P-values of the t-tests comparing the mean abundance of each taxa detected in the AGS of E1 and E2 in the seven bacterial stable states. *'all\_stable\_states\_t\_tests.ods'*

Table A.19 – Complete table of the all the MAG obtained from grouping PacBio contigs, with the number of contigs, total length, N50, minimum and maximum contig length, number of coding sequences (CDS), estimated percentage of completeness, estimated percentage of duplication, number of complete (C), single copy (CSC), duplicated (D) and fragmented (F) BUSCO, the coverage with Illumina sequences from extraction A (cov A), the coverage with Illumina sequences from extraction B (cov B), the number of repeat region, tmRNA, and tRNA. *'MAG\_table\_thesis.csv'*

Table A.20 – Distance between the characteristic tetranucleotide frequency vectors of selected reference genomes. *'distance\_tetranucl\_ref\_genomes.csv'*

File A.1 : Alignment of phosphate low-affinity transporter Pit sequences from Ca. Accumulibacter and Tetrasphaera genomes. *'CAP-Tet\_Pit\_mafft\_alignment.fasta'*

Table A.21 – Number of selected P-related hmm profiles found in the MAG and not grouped PacBio contigs. *'table\_hmm\_summary\_P\_gather.csv'*

Table A.22 – Number of selected N-related hmm profiles found in the MAG and not grouped PacBio contigs. *'table\_hmm\_summary\_N\_gather.csv'*

Folder A.1: Prokka outputs of the main MAG including .gff and .fna files. *'Main\_MAG\_prokka\_reports'*

### A.4.2 Fragment analysis reports and evaluation of the assembly after polishing

#### A.4. Supplementary information chapter V

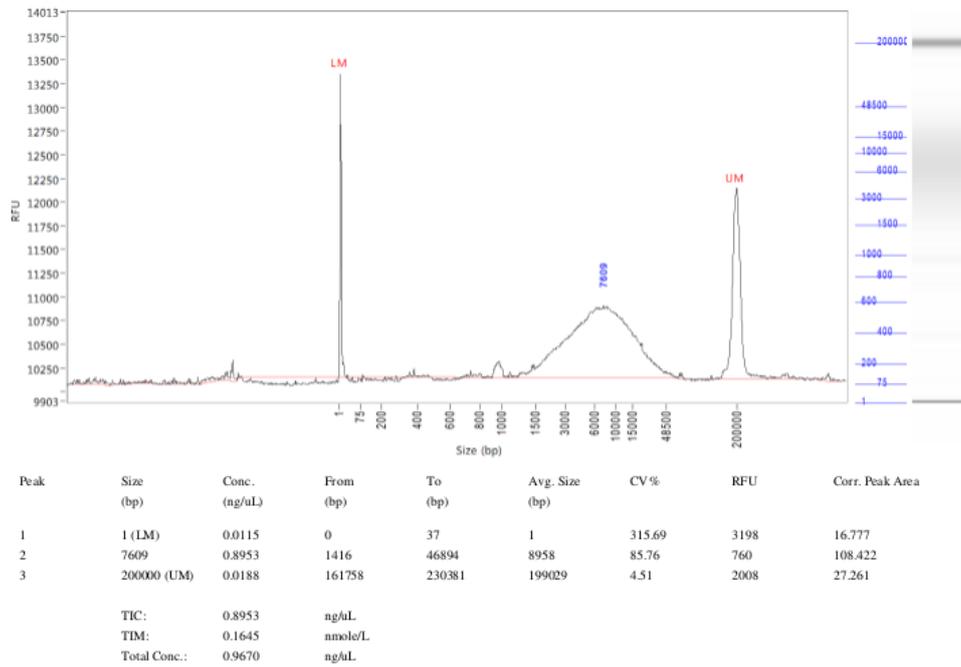


Figure A.7 – Fragment analysis report for the DNA extracted from sample d71 with the bacterial genomic DNA isolation CTAB protocol (extraction A)

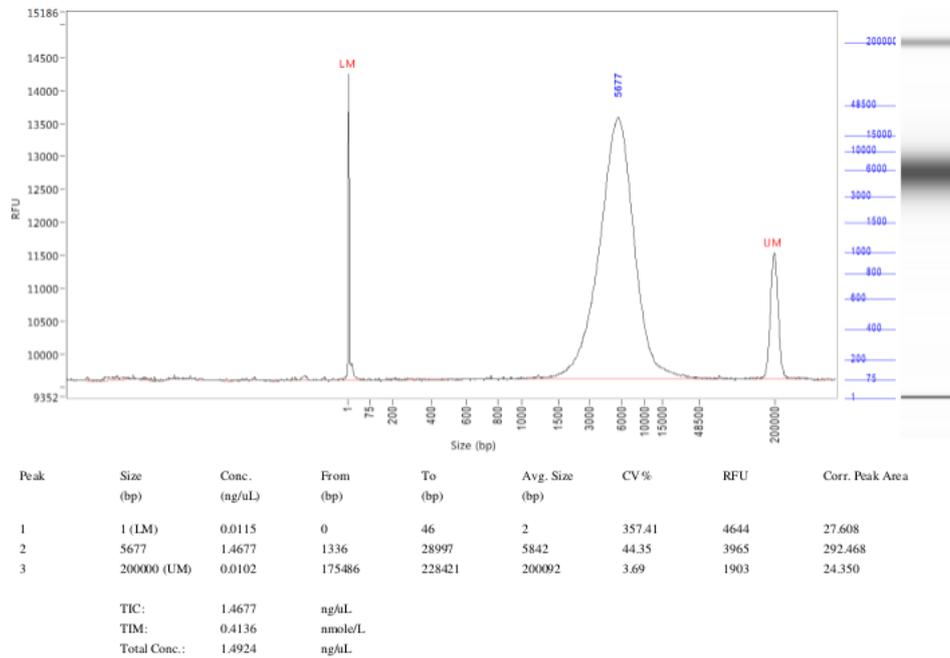


Figure A.8 – Fragment analysis report for the DNA extracted from sample d71 with the Maxwell® 16 Tissue DNA Purification Kit (extraction B)

**Appendix A. Supplementary material**

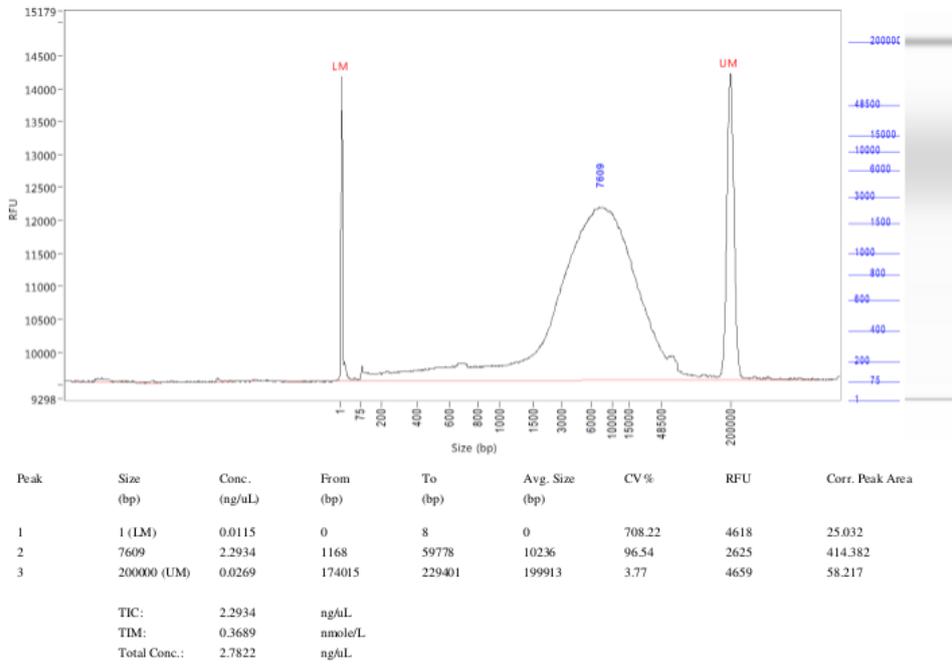


Figure A.9 – Fragment analysis report for the DNA extracted from sample d322 with the bacterial genomic DNA isolation CTAB protocol (extraction A)

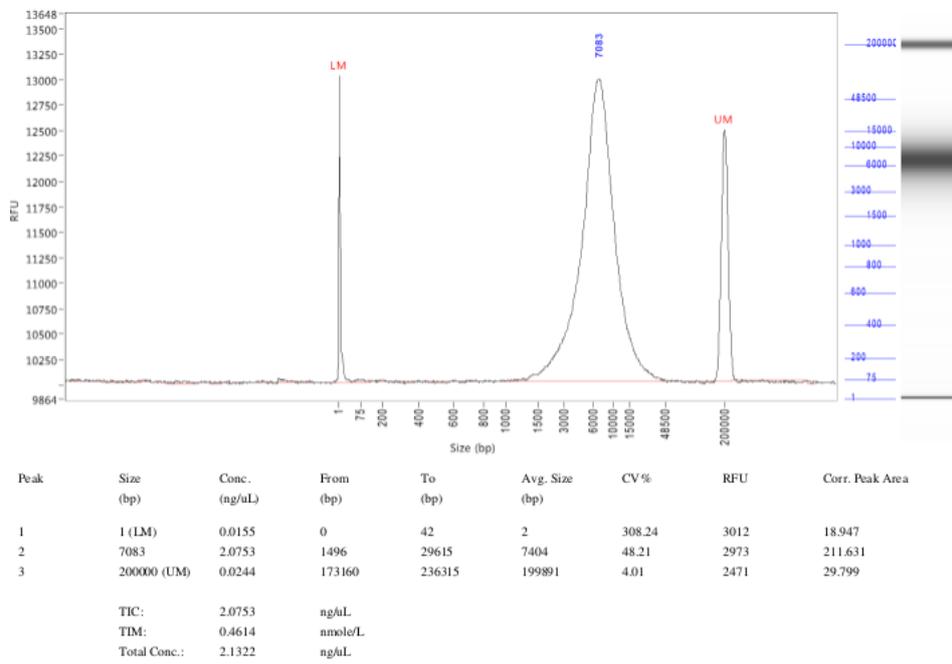


Figure A.10 – Fragment analysis report for the DNA extracted from sample d322 with the Maxwell® 16 Tissue DNA Purification Kit (extraction B)

#### A.4. Supplementary information chapter V

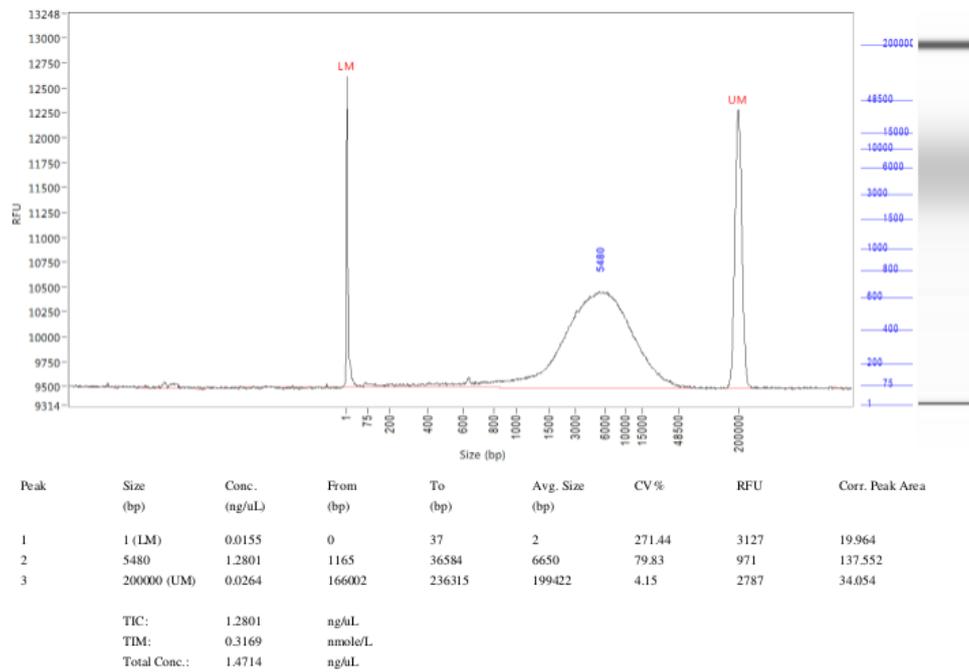


Figure A.11 – Fragment analysis report for the DNA extracted from sample d427 with the bacterial genomic DNA isolation CTAB protocol (extraction A)

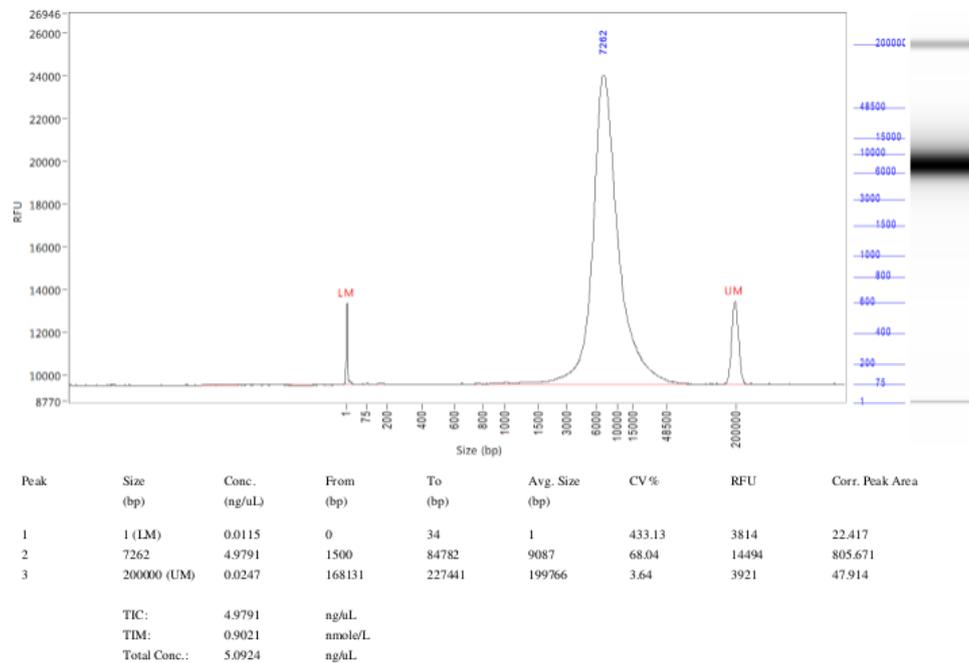


Figure A.12 – Fragment analysis report for the DNA extracted from sample d427 with the Maxwell® 16 Tissue DNA Purification Kit (extraction B)

**Appendix A. Supplementary material**

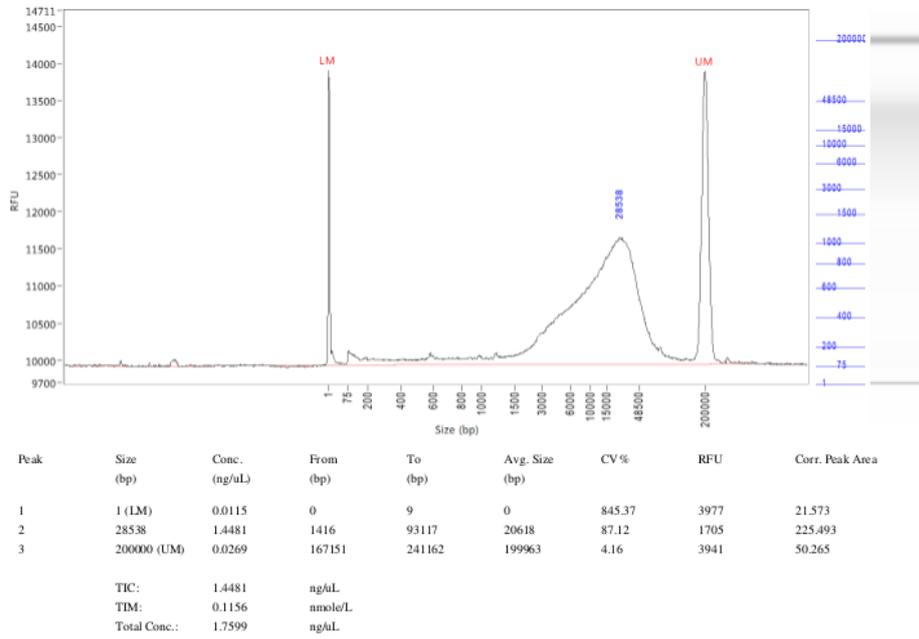


Figure A.13 – Fragment analysis report for the DNA extracted from sample d740 with the bacterial genomic DNA isolation CTAB protocol (extraction A)

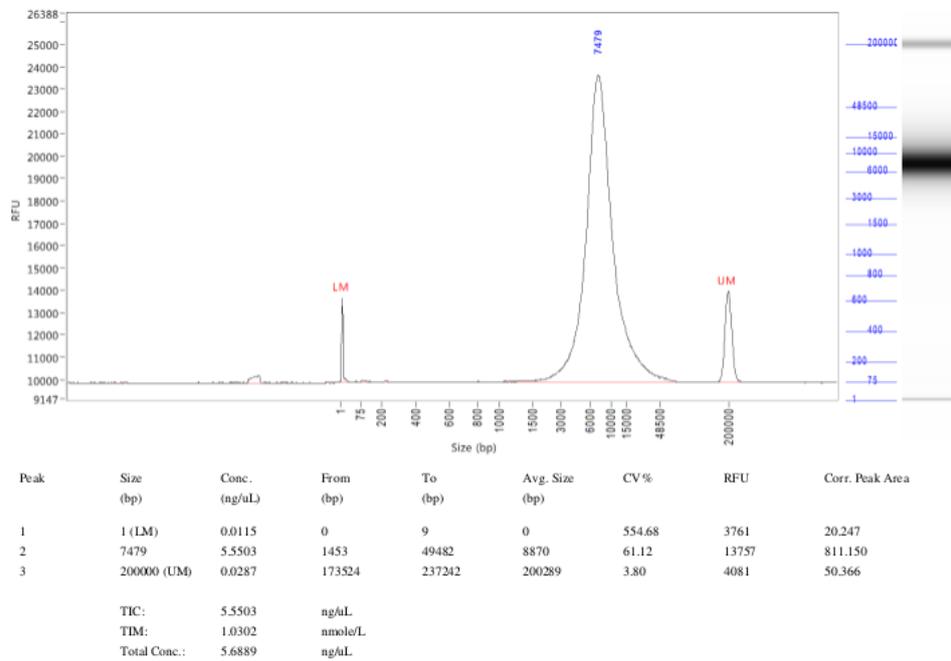


Figure A.14 – Fragment analysis report for the DNA extracted from sample d740 with the Maxwell® 16 Tissue DNA Purification Kit (extraction B)

## A.4. Supplementary information chapter V

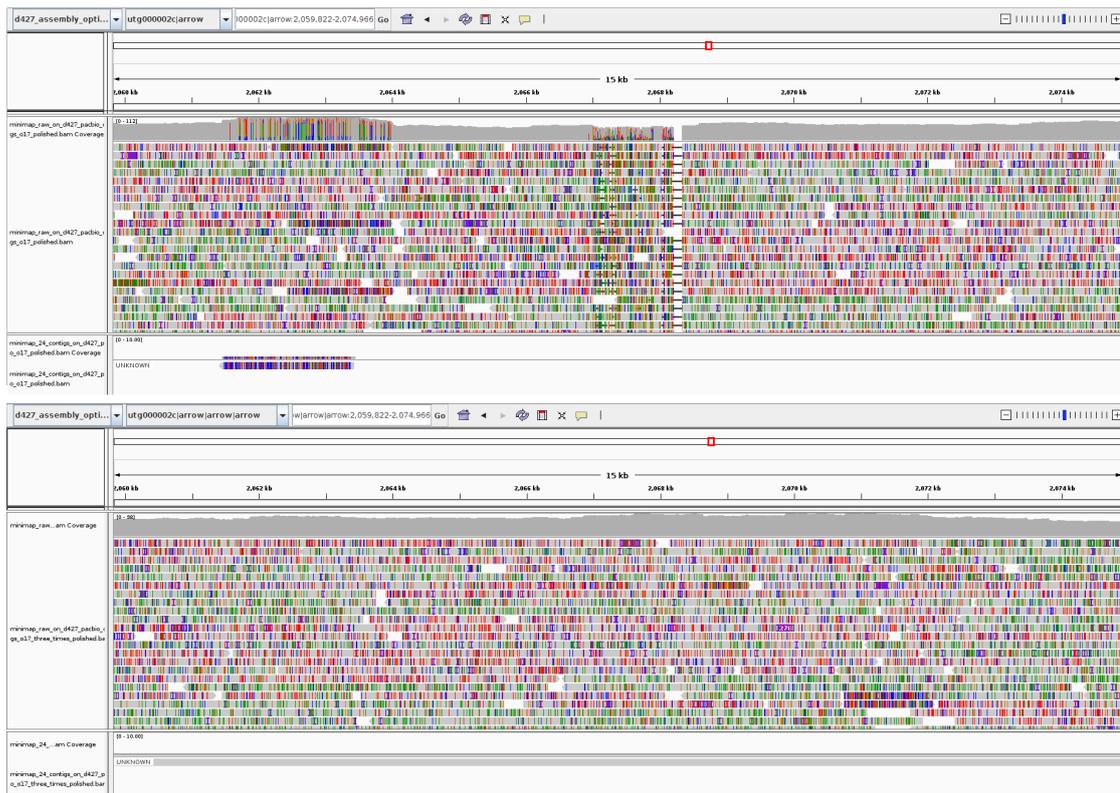


Figure A.15 – View on IGV of the mapping of PacBio long-reads and hybrid contigs from sample d427 on the PacBio contig utg000002c after 1 (upper image) and 3 (lower image) rounds of polishing.

## Appendix A. Supplementary material

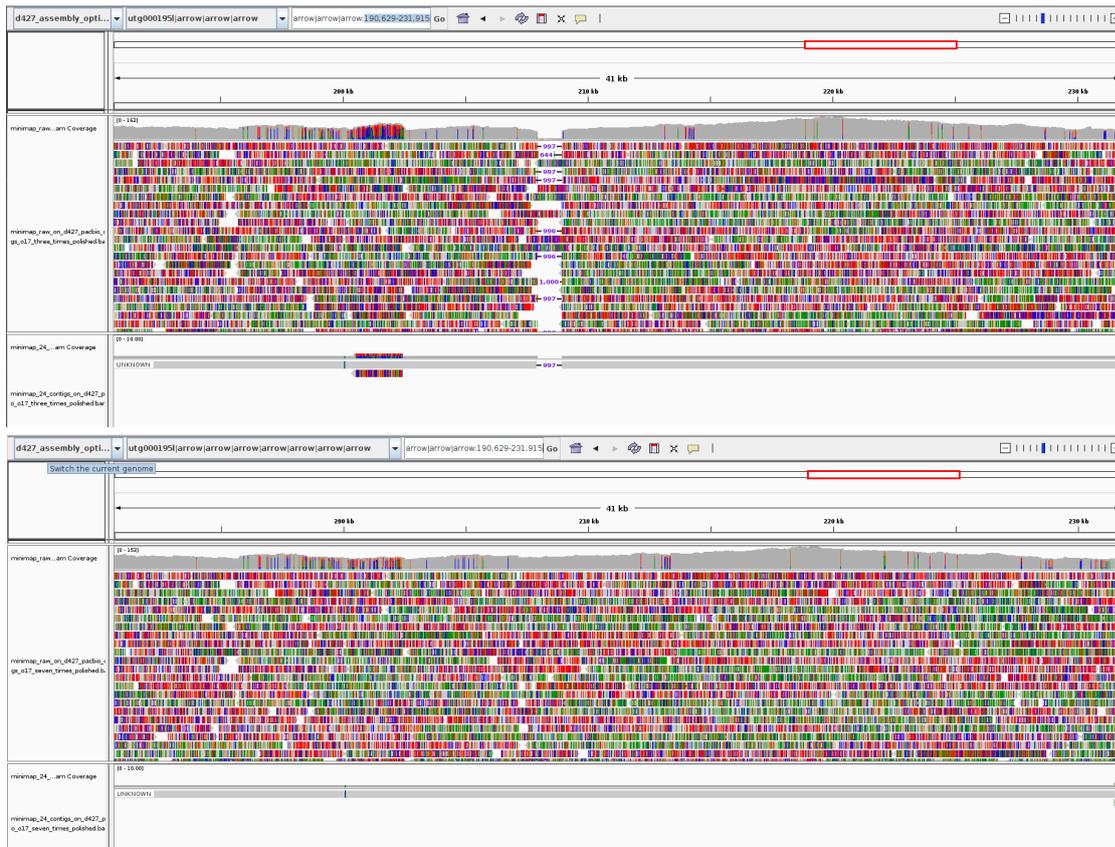


Figure A.16 – View on IGV of the mapping of PacBio long-reads and hybrid contigs from sample d427 on the PacBio contig utg000195l after 3 (upper image) and 7 (lower image) rounds of polishing.

## A.4. Supplementary information chapter V

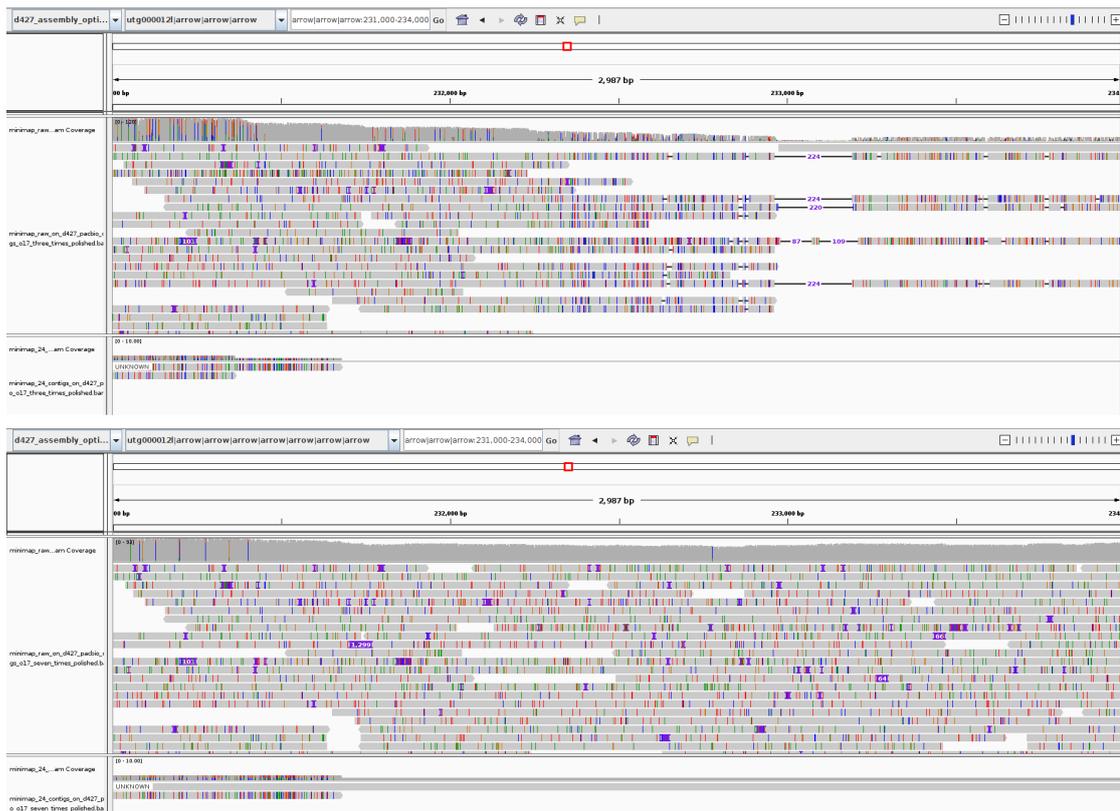


Figure A.17 – View on IGV of the mapping of PacBio long-reads and hybrid contigs from sample d427 on the PacBio contig utg000012l after 3 (upper image) and 7 (lower image) rounds of polishing.

## A.5 Supplementary information chapter VI

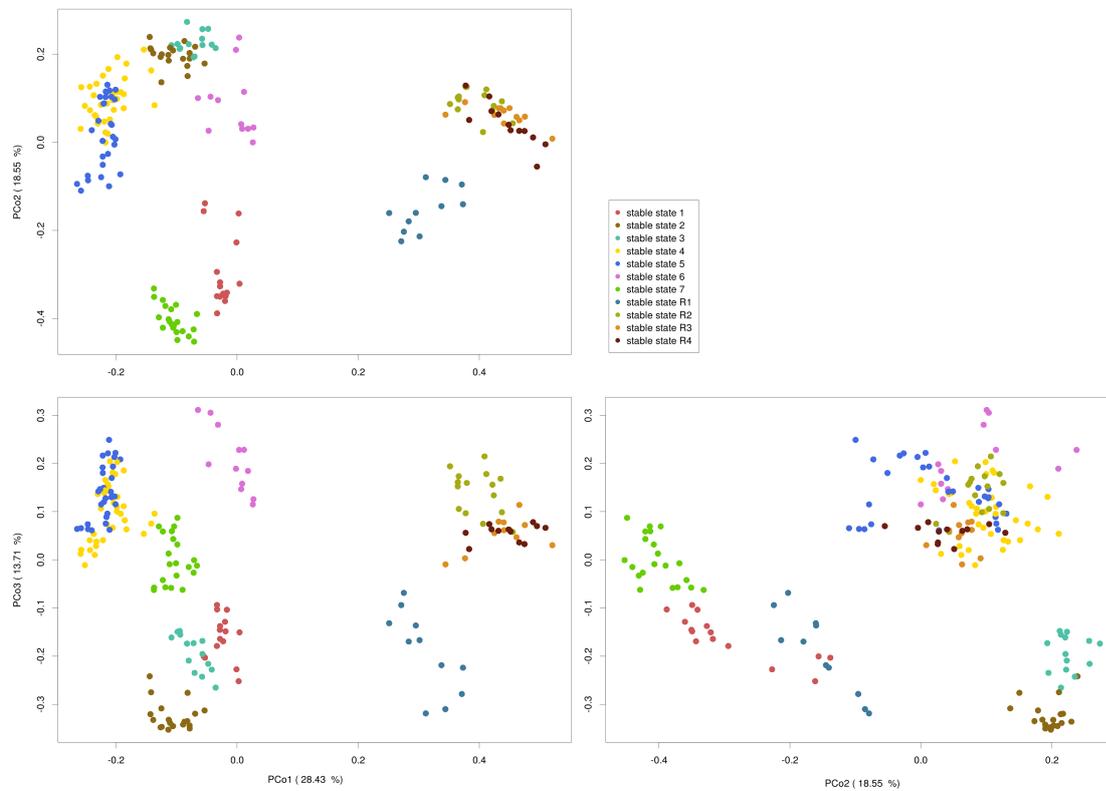


Figure A.18 – Principal coordinate analysis of the Bray-Curtis distance of the bacterial communities of the stable states of the two experiments on the 1st and 2nd axis (A), on the first and 3rd axis (B) and on the second and 3 axis (C).

# Bibliography

- Achenbach, L. and Coates, J. (2000). Disparity between Bacterial Phylogeny and Physiology - Comparing 16S rRNA sequences to assess relationships can be a powerful tool, but its limitations need to be considered. *ASM NEWS*, 66(12):714–715.
- Acinas, S., Marcelino, L., Klepac-Ceraj, V., and Polz, M. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *JOURNAL OF BACTERIOLOGY*, 186(9):2629–2635.
- Ahlgren, N. A., Chen, Y., Needham, D. M., Parada, A. E., Sachdeva, R., Trinh, V., Chen, T., and Fuhrman, J. A. (2017). Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. *ENVIRONMENTAL MICROBIOLOGY*, 19(6, SI):2434–2452.
- Akiyama, M., Crooke, E., and Kornberg, A. (1992). The polyphosphate kinase gene of *Escherichia coli* - isolation and sequence of the *ppk* gene and membrane location of the protein. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 267(31):22556–22561.
- Albertsen, M., Hansen, L. B. S., Saunders, A. M., Nielsen, P. H., and Nielsen, K. L. (2012). A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. *ISME JOURNAL*, 6(6):1094–1106.
- Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., and Nielsen, P. H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *NATURE BIOTECHNOLOGY*, 31(6):533+.
- Albertsen, M., Karst, S. M., Ziegler, A. S., Kirkegaard, R. H., and Nielsen, P. H. (2015). Back to Basics - The Influence of DNA Extraction and Primer Choice on Phylogenetic Analysis of Activated Sludge Communities. *PLOS ONE*, 10(7).
- Albi, T. and Serrano, A. (2016). Inorganic polyphosphate in the microbial world. Emerging roles for a multifaceted biopolymer. *WORLD JOURNAL OF MICROBIOLOGY & BIOTECHNOLOGY*, 32(2).
- Allison, S. D. and Martiny, J. B. H. (2008). Resistance, resilience, and redundancy in microbial communities. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 105(1):11512–11519.

## Bibliography

---

- Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. E., and Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *NATURE METHODS*, 11(11):1144–1146.
- Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. (1990). Basic local alignment search tool. *JOURNAL OF MOLECULAR BIOLOGY*, 215(3):403–410.
- Amann, R., Ludwig, W., and Scheifer, K. (1995). Phylogenetic identification and in-situ detection of individual microbial-cells without cultivation. *MICROBIOLOGICAL REVIEWS*, 59(1):143–169.
- Andersson, M. G. I., Berga, M., Lindstrom, E. S., and Langenheder, S. (2014). The spatial structure of bacterial communities is influenced by historical environmental conditions. *ECOLOGY*, 95(5):1134–1140.
- Andreadakis, A. (1993). Physical and chemical-properties of activated-sludge floc. *WATER RESEARCH*, 27(12):1707–1714.
- APHA (2012). *Standard methods for the examination of water and wastewater*. American Public Health Association (APHA), American Water Works Association (AWWA) and Water Pollution Control Federation (WPCF), New-York, 22<sup>nd</sup> edition.
- Ardern, E. and Lockett, W. T. (1914). Experiments on the oxidation of sewage without the aid of filters. *Journal of the Society of Chemical Industry*, 33(10):523–539.
- Auchincloss, A. H., Cuche, B. A., Rivoire, C., Baratin, D., de Castro, E., Coudert, E., Keller, G., Pedruzzi, I., Bougueleret, L., Redaschi, N., Poux, S., Bridge, A., and Xenarios, I. (2014). HAMAP in 2015: updates to the protein family classification and annotation system. *Nucleic Acids Research*, 43(D1):D1064–D1070.
- Bae, H.-S., Moe, W. M., Yan, J., Tiago, I., da Costa, M. S., and Rainey, F. A. (2006). *Propionicicella superfundia* gen. nov., sp nov., a chlorosolvent-tolerant propionate-forming, facultative anaerobic bacterium isolated from contaminated groundwater. *SYSTEMATIC AND APPLIED MICROBIOLOGY*, 29(5):404–413.
- Barnard, J. L., Dunlap, P., and Steichen, M. (2017). Rethinking the Mechanisms of Biological Phosphorus Removal. *WATER ENVIRONMENT RESEARCH*, 89(11):2043–2054.
- Barnett, D. W., Garrison, E. K., Quinlan, A. R., Stroemberg, M. P., and Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *BIOINFORMATICS*, 27(12):1691–1692.
- Barr, J. J., Cook, A. E., and Bond, P. L. (2010a). Granule Formation Mechanisms within an Aerobic Wastewater System for Phosphorus Removal. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 76(22):7588–7597.

- Barr, J. J., Dutilh, B. E., Skennerton, C. T., Fukushima, T., Hastie, M. L., Gorman, J. J., Tyson, G. W., and Bond, P. L. (2016). Metagenomic and metaproteomic analyses of *Accumulibacter phosphatis*-enriched floccular and granular biofilm. *ENVIRONMENTAL MICROBIOLOGY*, 18(1):273–287.
- Barr, J. J., Slater, F. R., Fukushima, T., and Bond, P. L. (2010b). Evidence for bacteriophage activity causing community and performance changes in a phosphorus-removal activated sludge. *FEMS MICROBIOLOGY ECOLOGY*, 74(3):631–642.
- Basheer, F. and Farooqi, I. (2014). Hydrodynamic properties of aerobic granules cultivated on phenol as carbon source. *APCBEE Procedia*, 10:126 – 130. 5th International Conference on Environmental Science and Development – ICESD 2014.
- Bassin, J. P., Kleerebezem, R., Dezotti, M., and van Loosdrecht, M. C. M. (2012). Measuring biomass specific ammonium, nitrite and phosphate uptake rates in aerobic granular sludge. *CHEMOSPHERE*, 89(10):1161–1168.
- Bateman, A., Kalvari, I., Argasinska, J., Finn, R. D., Petrov, A. I., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., and Eddy, S. R. (2017). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Research*, 46(D1):D335–D342.
- Bateman, A., Smart, A., Luciani, A., Salazar, G. A., Mistry, J., Richardson, L. J., Qureshi, M., El-Gebali, S., Potter, S. C., Finn, R. D., Eddy, S. R., Sonnhammer, E. L., Piovesan, D., Paladin, L., Tosatto, S. C., and Hirsh, L. (2018). The Pfam protein families database in 2019. *Nucleic Acids Research*, 47(D1):D427–D432.
- Bekkelund, A., Hansen, B., Stokken, H., Eriksen, B., and Kashulin, N. (2018). *Flavobacterium* sp. 140616w15, direct submission. accession CP033068.
- Benneouala, M., Bareha, Y., Mengelle, E., Bounouba, M., Sperandio, M., Bessiere, Y., and Paul, E. (2017). Hydrolysis of particulate settleable solids (PSS) in activated sludge is determined by the bacteria initially adsorbed in the sewage. *WATER RESEARCH*, 125:400–409.
- Beun, J., Hendriks, A., Van Loosdrecht, M., Morgenroth, E., Wilderer, P., and Heijnen, J. (1999). Aerobic granulation in a sequencing batch reactor. *WATER RESEARCH*, 33(10):2283–2290.
- Beun, J., van Loosdrecht, M., and Heijnen, J. (2000). Aerobic granulation. *WATER SCIENCE AND TECHNOLOGY*, 41(4-5):41–48. 4th International Conference on Biofilm Systems, NEW YORK, NEW YORK, OCT 17-20, 1999.
- Bin, Z., Bin, X., Zhigang, Q., Zhiqiang, C., Junwen, L., Taishi, G., Wenci, Z., and Jingfeng, W. (2015). Denitrifying capability and community dynamics of glycogen accumulating organisms during sludge granulation in an anaerobic-aerobic sequencing batch reactor. *SCIENTIFIC REPORTS*, 5.
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *BIOINFORMATICS*, 30(15):2114–2120.

## Bibliography

---

- Bor, B., Bedree, J. K., Shi, W., McLean, J. S., and He, X. (2019). Saccharibacteria (TM7) in the Human Oral Microbiome. *JOURNAL OF DENTAL RESEARCH*, 98(5):500–509.
- Botton, S., van Heusden, M., Parsons, J., Smidt, H., and van Straalen, N. (2006). Resilience of microbial systems towards disturbances. *CRITICAL REVIEWS IN MICROBIOLOGY*, 32(2):101–112.
- Briones, A. and Raskin, L. (2003). Diversity and dynamics of microbial communities in engineered environments and their implications for process stability. *CURRENT OPINION IN BIOTECHNOLOGY*, 14(3):270–276.
- Brockhurst, M. A., Fenton, A., Roulston, B., and Rainey, P. B. (2006). The impact of phages on interspecific competition in experimental populations of bacteria. *BMC Ecology*, 6(1):19.
- Burne, R. (1998). Oral streptococci... products of their environment. *Journal of Dental Research*, 77(3):445–452. PMID: 9496917.
- Burow, L. C., Kong, Y., Nielsen, J. L., Blackall, L. L., and Nielsen, P. H. (2007). Abundance and ecophysiology of *DeFluviicoccus* spp., glycogen-accumulating organisms in full-scale wastewater treatment processes. *MICROBIOLOGY-SGM*, 153(1):178–185.
- Bushnell, B. (2014). Bbmap short aligner, and other bioinformatic tools.
- Böhme, A., Risse-Buhl, U., and Küsel, K. (2009). Protists with different feeding modes change biofilm morphology. *FEMS Microbiology Ecology*, 69(2):158–169.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). BLAST plus : architecture and applications. *BMC BIOINFORMATICS*, 10.
- Campos, J. L., Valenzuela-Heredia, D., Pedrouso, A., Val del Rio, A., Belmonte, M., and Mosquera-Corral, A. (2016). Greenhouse Gases Emissions from Wastewater Treatment Plants: Minimization, Treatment, and Prevention. *JOURNAL OF CHEMISTRY*.
- Cech, J. and Hartman, P. (1993). Competition between polyphosphate and polysaccharide accumulating bacteria in enhanced biological phosphate removal systems. *WATER RESEARCH*, 27(7):1219–1225.
- Cetin, E., Karakas, E., Dulekgurgen, E., Ovez, S., Kolukirik, M., and Yilmaz, G. (2018). Effects of high-concentration influent suspended solids on aerobic granulation in pilot-scale sequencing batch reactors treating real domestic wastewater. *WATER RESEARCH*, 131:74–89.
- Chen, Y., Jiang, W., Liang, D. T., and Tay, J. H. (2007). Structure and stability of aerobic granules cultivated under different shear force in sequencing batch reactors. *APPLIED MICROBIOLOGY AND BIOTECHNOLOGY*, 76(5):1199–1208.
- Coates, J., Chakraborty, R., Lack, J., O'Connor, S., Cole, K., Bender, K., and Achenbach, L. (2001). Anaerobic benzene oxidation coupled to nitrate reduction in pure culture by two strains of *Dechloromonas*. *NATURE*, 411(6841):1039–1043.

- Coats, E. R., Brinkman, C. K., and Lee, S. (2017). Characterizing and contrasting the microbial ecology of laboratory and full-scale EBPR systems cultured on synthetic and real wastewaters. *WATER RESEARCH*, 108:124–136.
- Coats, E. R., Watkins, D. L., and Kranenburg, D. (2011). A comparative environmental life-cycle analysis for removing phosphorus from wastewater: Biological versus physical/chemical processes. *Water Environment Research*, 83(8):750–760.
- Coma, M., Verawaty, M., Pijuan, M., Yuan, Z., and Bond, P. L. (2012). Enhancing aerobic granulation for biological nutrient removal from domestic wastewater. *BIORESOURCE TECHNOLOGY*, 103(1):101–108.
- Confer, D. and Logan, B. (1998). Location of protein and polysaccharide hydrolytic activity in suspended and biofilm wastewater cultures. *WATER RESEARCH*, 32(1):31–38.
- Cvitkovitch, D., Liu, Y., and Ellen, R. (2003). Quorum sensing and biofilm formation in Streptococcal infections. *JOURNAL OF CLINICAL INVESTIGATION*, 112(11):1626–1632.
- Daims, H., Lebedeva, E. V., Pjevac, P., Han, P., Herbold, C., Albertsen, M., Jehmlich, N., Palatin-szky, M., Vierheilig, J., Bulaev, A., Kirkegaard, R. H., von Bergen, M., Rattei, T., Bendinger, B., Nielsen, P. H., and Wagner, M. (2015). Complete nitrification by Nitrospira bacteria. *NATURE*, 528(7583):504+.
- de Almeida, L. G., Ortiz, J. H., Schneider, R. P., and Spira, B. (2015). *phoU* Inactivation in *Pseudomonas aeruginosa* Enhances Accumulation of ppGpp and Polyphosphate. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 81(9):3006–3015.
- de Bruin, L., de Kreuk, M., van der Roest, H., Uijterlinde, C., and van Loosdrecht, M. (2004). Aerobic granular sludge technology: an alternative to activated sludge? *WATER SCIENCE AND TECHNOLOGY*, 49(11-12):1–7. 5th IWA International Conference on Biofilm Systems, Cape Town, SOUTH AFRICA, SEP 14-18, 2003.
- de Kreuk, M., Heijnen, J., and van Loosdrecht, M. (2005). Simultaneous COD, nitrogen, and phosphate removal by aerobic granular sludge. *BIOTECHNOLOGY AND BIOENGINEERING*, 90(6):761–769.
- de Kreuk, M. and van Loosdrecht, M. (2004). Selection of slow growing organisms as a means for improving aerobic granular sludge stability. *WATER SCIENCE AND TECHNOLOGY*, 49(11-12):9–17. 5th IWA International Conference on Biofilm Systems, Cape Town, SOUTH AFRICA, SEP 14-18, 2003.
- de Kreuk, M. K., Kishida, N., Tsuneda, S., and van Loosdrecht, M. C. M. (2010). Behavior of polymeric substrates in an aerobic granular sludge system. *WATER RESEARCH*, 44(20):5929–5938.
- de Kreuk, M. K., Kishida, N., and van Loosdrecht, M. C. M. (2007). Aerobic granular sludge - state of the art. *WATER SCIENCE AND TECHNOLOGY*, 55(8-9):75–81. 6th IWA International Specialty Conference on Biofilm Systems, Amsterdam, NETHERLANDS, SEP 24-27, 2006.

## Bibliography

---

- de Kreuk, M. K. and van Loosdrecht, M. C. M. (2006). Formation of aerobic granules with domestic sewage. *JOURNAL OF ENVIRONMENTAL ENGINEERING-ASCE*, 132(6):694–697.
- deBeer, D. and Stoodley, P. (1995). Relation between the structure of an aerobic biofilm and transport phenomena? *WATER SCIENCE AND TECHNOLOGY*, 32(8):11–18. IAWQ International Conference and Workshop on Biofilm Structure, Growth and Dynamics, NOORDWIJKERHOUT, NETHERLANDS, AUG 30-SEP 01, 1995.
- Derlon, N., Wagner, J., Ribeiro da Costa, R. H., and Morgenroth, E. (2016). Formation of aerobic granules for the treatment of real and low-strength municipal wastewater using a sequencing batch reactor operated at constant volume. *WATER RESEARCH*, 105:341–350.
- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G. L. (2006). Greengenes, a chimera-checked 16s rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, 72(7):5069–5072.
- Deschavanne, P., Giron, A., Vilain, J., Dufraigne, C., and Fertil, B. (2000). Genomic signature is preserved in short DNA fragments. In *IEEE INTERNATIONAL SYMPOSIUM ON BIO-INFORMATICS AND BIOMEDICAL ENGINEERING, PROCEEDINGS*, pages 161–167. IEEE, Comp Soc; IEEE, Comp Soc, Task Force Virtual Intelligence; IEEE, Engn Med & Biol Soc; EMB; Allis Inc; ICIIS, ICTAI, NIH, NCI, TIGR. IEEE International Symposium on Bio-Informatics and Biomedical Engineering, ARLINGTON, VA, NOV 08-10, 2000.
- Dethlefsen, L. and Relman, D. A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 108(1):4554–4561.
- DHV (2019). Nereda. <https://www.royalhaskoningdhv.com/en-gb/nereda>. [Online; accessed 17-May-2019].
- Dick, G. J., Andersson, A. F., Baker, B. J., Simmons, S. L., Yelton, A. P., and Banfield, J. F. (2009). Community-wide analysis of microbial genome sequence signatures. *GENOME BIOLOGY*, 10(8).
- Dimock, R. and Morgenroth, E. (2006). The influence of particle size on microbial hydrolysis of protein particles in activated sludge. *WATER RESEARCH*, 40(10):2064–2074.
- Drury, W., Stewart, P., and Characklis, W. (1993). Transport of 1- $\mu$ m latex-particles in *Pseudomonas-aeruginosa* biofilms. *BIOTECHNOLOGY AND BIOENGINEERING*, 42(1):111–117.
- Dueholm, T., Andreasen, K., and Nielsen, P. (2001). Transformation of lipids in activated sludge. *WATER SCIENCE AND TECHNOLOGY*, 43(1):165–172. 1st World Water Congress of the International-Water-Association (IWA), PARIS, FRANCE, JUL 03-07, 2000.

- Dulekgurgen, E., Ovez, S., Artan, N., and Orhon, D. (2003). Enhanced biological phosphate removal by granular sludge in a sequencing batch reactor. *BIOTECHNOLOGY LETTERS*, 25(9):687–693.
- Dupont, C. L., Rusch, D. B., Yooseph, S., Lombardo, M.-J., Richter, R. A., Valas, R., Novotny, M., Yee-Greenbaum, J., Selengut, J. D., Haft, D. H., Halpern, A. L., Lasken, R. S., Nealson, K., Friedman, R., and Venter, J. C. (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME JOURNAL*, 6(6):1186–1199.
- Eberhard Morgenroth (2008). *Biological wastewater treatment: principles, modelling and design*. IWA Publishing, London, UK, edition. Edited by M. Henze, M. C. M. van Loosdrecht, G. A. Ekama and D. Brdjanovic.
- Ebrahimi, S., Gabus, S., Rohrbach-Brandt, E., Hosseini, M., Rossi, P., Maillard, J., and Holliger, C. (2010). Performance and microbial community composition dynamics of aerobic granular sludge from sequencing batch bubble column reactors operated at 20 A degrees C, 30 A degrees C, and 35 A degrees C. *APPLIED MICROBIOLOGY AND BIOTECHNOLOGY*, 87(4):1555–1568.
- Edgar, R. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC BIOINFORMATICS*, 5:1–19.
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., deWinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Veceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korfach, J., and Turner, S. (2009). Real-Time DNA Sequencing from Single Polymerase Molecules. *SCIENCE*, 323(5910):133–138.
- Eisen, J., Huntemann, M., Han, J., Chen, A., Kyrpides, N., Mavromatis, K., Markowitz, V., Palaniappan, K., Ivanova, N., Schaumberg, A., Pati, A., Liolios, K., Nordberg, H., Cantor, M., Hua, S., and Woyke, T. (2014). Niabella soli dsm 19437, direct submission. accession CP007035.
- English, A. C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D. M., Reid, J. G., Worley, K. C., and Gibbs, R. A. (2012). Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *PLOS ONE*, 7(11).
- Etterer, T. and Wilderer, P. (2001). Generation and properties of aerobic granular sludge. *WATER SCIENCE AND TECHNOLOGY*, 43(3):19–26. 2nd IWA International Conference on Sequencing Batch Reactor Technology, NARBONNE, FRANCE, JUL 10-12, 2000.
- Faust, K., Lahti, L., Gonze, D., de Vos, W. M., and Raes, J. (2015). Metagenomics meets time series analysis: unraveling microbial community dynamics. *CURRENT OPINION IN MICROBIOLOGY*, 25:56–66.

## Bibliography

---

- Federle, M. and Bassler, B. (2003). Interspecies communication in bacteria. *JOURNAL OF CLINICAL INVESTIGATION*, 112(9):1291–1299.
- Feng, L., Huang, Y., and Wang, H. (2008). Solid-phase microextraction in combination with GC-FID for quantification of the volatile free fatty acids in wastewater from constructed wetlands. *JOURNAL OF CHROMATOGRAPHIC SCIENCE*, 46(7):577–584.
- Fernandez, A., Hashsham, S., Dollhopf, S., Raskin, L., Glagoleva, O., Dazzo, F., Hickey, R., Cridle, C., and Tiedje, J. (2000). Flexible community structure correlates with stable community function in methanogenic bioreactor communities perturbed by glucose. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 66(9):4058–4067.
- Flowers, J. J., Cadkin, T. A., and McMahan, K. D. (2013). Seasonal bacterial community dynamics in a full-scale enhanced biological phosphorus removal plant. *WATER RESEARCH*, 47(19):7019–7031.
- Flowers, J. J., He, S., Yilmaz, S., Noguera, D. R., and McMahan, K. D. (2009). Denitrification capabilities of two biological phosphorus removal sludges dominated by different ‘*Candidatus Accumulibacter*’ clades. *ENVIRONMENTAL MICROBIOLOGY REPORTS*, 1(6):583–588.
- Frank, J. A., Pan, Y., Tooming-Klunderud, A., Eijsink, V. G. H., McHardy, A. C., Nederbragt, A. J., and Pope, P. B. (2016). Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. *SCIENTIFIC REPORTS*, 6.
- Frolund, B., Griebe, T., and Nielsen, P. (1995). Enzymatic-activity in the activated-sludge floc matrix. *APPLIED MICROBIOLOGY AND BIOTECHNOLOGY*, 43(4):755–761.
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *BIOINFORMATICS*, 28(23):3150–3152.
- Galperin, M. Y., Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *NUCLEIC ACIDS RESEARCH*, 43(D1):D261–D269.
- Garnier, S. (2018). *viridis: Default Color Maps from 'matplotlib'*. R package version 0.5.1.
- Giesen, A., de Bruin, L. M. M., Niermans, R. P., and van der Roest, H. F. (2013). Advancements in the application of aerobic granular biomass technology for sustainable treatment of wastewater. *Water Practice and Technology*, 8(1):47–54.
- Goeker, M., Huntemann, M., Clum, A., Pillay, M., Palaniappan, K., Varghese, N., Mikhailova, N., Stamatis, D., Reddy, T., Daum, C., Shapiro, N., Ivanova, N., Kyrpides, N., and Woyke, T. (2018). *Defluviimonas denitrificans*, direct submission. accession PVEP01000001.
- Goel, R., Mino, T., Satoh, H., and Matsuo, T. (1997). Effect of electron acceptor conditions on hydrolytic enzyme synthesis in bacterial cultures. *WATER RESEARCH*, 31(10):2597–2603.

- Goel, R., Mino, T., Satoh, H., and Matsuo, T. (1998a). Comparison of hydrolytic enzyme systems in pure culture and activated sludge under different electron acceptor conditions. *WATER SCIENCE AND TECHNOLOGY*, 37(4-5):335–343. 2nd International Conference on Microorganisms in Activated Sludge and Biofilm Processes, BERKELEY, CALIFORNIA, JUL 21-23, 1997.
- Goel, R., Mino, T., Satoh, H., and Matsuo, T. (1998b). Enzyme activities under anaerobic and aerobic conditions inactivated sludge sequencing batch reactor. *WATER RESEARCH*, 32(7):2081–2088.
- Gonzalez-Gil, G. and Holliger, C. (2011). Dynamics of Microbial Community Structure of and Enhanced Biological Phosphorus Removal by Aerobic Granules Cultivated on Propionate or Acetate. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 77(22):8041–8051.
- Gonzalez-Gil, G. and Holliger, C. (2014). Aerobic Granules: Microbial Landscape and Architecture, Stages, and Practical Implications. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 80(11):3433–3441.
- Gonze, D., Lahti, L., Raes, J., and Faust, K. (2017). Multi-stability and the origin of microbial community types. *ISME JOURNAL*, 11(10):2159–2166.
- Guimaraes, L. B., Mezzari, M. P., Daudt, G. C., and da Costa, R. H. R. (2017). Microbial pathways of nitrogen removal in aerobic granular sludge treating domestic wastewater. *JOURNAL OF CHEMICAL TECHNOLOGY AND BIOTECHNOLOGY*, 92(7):1756–1765.
- Guimarães, L. B., Wagner, J., Akaboci, T. R. V., Daudt, G. C., Nielsen, P. H., van Loosdrecht, M. C. M., Weissbrodt, D. G., and da Costa, R. H. R. (2018). Elucidating performance failures in use of granular sludge for nutrient removal from domestic wastewater in a warm coastal climate region. *Environmental Technology*, 0(0):1–16. PMID: 30465694.
- Gujer, W. (2007). *Siedlungswasserwirtschaft*. Springer, Berlin.
- Haft, D., Loftus, B., Richardson, D., Yang, F., Eisen, J., Paulsen, I., and White, O. (2001). TIGRFAMs: a protein family resource for the functional identification of proteins. *NUCLEIC ACIDS RESEARCH*, 29(1):41–43.
- Hanada, S., Liu, W., Shintani, T., Kamagata, Y., and Nakamura, K. (2002). *Tetrasphaera elongata* sp nov., a polyphosphate-accumulating bacterium isolated from activated sludge. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*, 52(3):883–887.
- Hao, L., McIlroy, S. J., Kirkegaard, R. H., Karst, S. M., Fernando, W. E. Y., Aslan, H., Meyer, R. L., Albertsen, M., Nielsen, P. H., and Dueholm, M. S. (2018). Novel prosthecate bacteria from the candidate phylum Acetothermia. *ISME JOURNAL*, 12(9):2225–2237.
- Harrell Jr, F. E., with contributions from Charles Dupont, and many others. (2018). *Hmisc: Harrell Miscellaneous*. R package version 4.1-1.

## Bibliography

---

- Haybrard, J., Simon, N., Danel, C., Pincon, C., Barthelemy, C., Tessier, F. J., Decaudin, B., Boulanger, E., and Odou, P. (2017). Factors Generating Glucose Degradation Products In Sterile Glucose Solutions For Infusion: Statistical Relevance Determination Of Their Impacts. *SCIENTIFIC REPORTS*, 7.
- He, Q., Wang, H., Yang, X., Zhou, J., Ye, Y., Chen, D., and Yang, K. (2016a). Culture of denitrifying phosphorus removal granules with different influent wastewater. *DESALINATION AND WATER TREATMENT*, 57(37):17247–17254.
- He, Q., Zhou, J., Wang, H., Zhang, J., and Wei, L. (2016b). Microbial population dynamics during sludge granulation in an A/O/A sequencing batch reactor. *BIORESOURCE TECHNOLOGY*, 214:1–8.
- He, S., Gall, D. L., and McMahon, K. D. (2007). “Candidatus Accumulibacter” population structure in enhanced biological phosphorus removal Sludges as revealed by polyphosphate kinase genes. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 73(18):5865–5874.
- He, S. and McMahon, K. D. (2011). ‘Candidatus Accumulibacter’ gene expression in response to dynamic EBPR conditions. *ISME JOURNAL*, 5(2):329–340.
- Henriet, O., Meunier, C., Henry, P., and Mahillon, J. (2016). Improving phosphorus removal in aerobic granular sludge processes through selective microbial management. *BIORESOURCE TECHNOLOGY*, 211:298–306.
- Hesselmann, R., Werlen, C., Hahn, D., van der Meer, J., and Zehnder, A. (1999). Enrichment, phylogenetic analysis and detection of a bacterium that performs enhanced biological phosphate removal in activated sludge. *SYSTEMATIC AND APPLIED MICROBIOLOGY*, 22(3):454–465.
- Hille, F., Richter, H., Wong, S. P., Bratovic, M., Ressel, S., and Charpentier, E. (2018). The Biology of CRISPR-Cas: Backward and Forward. *CELL*, 172(6):1239–1259.
- Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, 4(1):1–23.
- Holling, C. S. (1996). *Engineering Resilience versus Ecological Resilience*. National Academy, Washinfon, D.C., USA.
- Hulot, F., Lacroix, G., Lescher-Moutoue, F., and Loreau, M. (2000). Functional diversity governs ecosystem response to nutrient enrichment. *NATURE*, 405(6784):340–344.
- Im, W. (2018). Draft genome sequence of ferruginibacter sp. bo-59. accession RJJR01000001.
- Infos-eau (2009). L’OCDE appelle à investir dans l’assainissement. <http://infos-eau.blogspot.com/2009/01/locde-appelle-investir-dans.html>. [Online; accessed 17-May-2019].
- Institute, H. H. M. (2018). Hmmer 3.2.1 (june 2018); hmmscan :: search sequence(s) against a profile database copyright (c). Freely distributed under the BSD open source license.

- IWA (2009). L'OCDE appelle à investir dans l'assainissement. <http://infos-eau.blogspot.com/2009/01/locde-appelle-investir-dans.html>. [Online; accessed 17-May-2019].
- Jabari, P., Yuan, Q., and Oleszkiewicz, J. A. (2016). Potential of Hydrolysis of Particulate COD in Extended Anaerobic Conditions to Enhance Biological Phosphorous Removal. *BIOTECHNOLOGY AND BIOENGINEERING*, 113(11):2377–2385.
- Jain, M., Fiddes, I. T., Miga, K. H., Olsen, H. E., Paten, B., and Akeson, M. (2015). Improved data analysis for the MinION nanopore sequencer. *NATURE METHODS*, 12(4):351–U115.
- James, B. T., Luczak, B. B., and Girgis, H. Z. (2018). MeShClust: an intelligent tool for clustering DNA sequences. *Nucleic Acids Research*, 46(14):e83–e83.
- Janning, K., Le Tallec, X., and Harremoës, P. (1998). Hydrolysis of organic wastewater particles in laboratory scale and pilot scale biofilm reactors under anoxic and aerobic conditions. *WATER SCIENCE AND TECHNOLOGY*, 38(8-9):179–188. 19th Biennial Conference of the International-Association-on-Water-Quality, VANCOUVER, CANADA, JUN 21-26, 1998.
- Jefferson, K. (2004). What drives bacteria to produce a biofilm? *FEMS MICROBIOLOGY LETTERS*, 236(2):163–173.
- Jefferson, K., Pier, D., Goldmann, D., and Pier, G. (2004). The teicoplanin-associated locus regulator (TcaR) and the intercellular adhesin locus regulator (IcaR) are transcriptional inhibitors of the ica locus in *Staphylococcus aureus*. *JOURNAL OF BACTERIOLOGY*, 186(8):2449–2456.
- Jiang, X., Ma, M., Li, J., Lu, A., and Zhong, Z. (2008). Bacterial diversity of active sludge in wastewater treatment plant. *Earth Science Frontiers*, 15(6):163 – 168.
- Jones, D., Taylor, W., and Thornton, J. (1992). The rapid generation of mutation data matrices from protein sequences. *COMPUTER APPLICATIONS IN THE BIOSCIENCES*, 8(3):275–282.
- Ju, F. and Zhang, T. (2015). Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. *ISME JOURNAL*, 9(3):683–695.
- Juretschko, S., Loy, A., Lehner, A., and Wagner, M. (2002). The microbial community composition of a nitrifying-denitrifying activated sludge from an industrial sewage treatment plant analyzed by the full-cycle rRNA approach. *SYSTEMATIC AND APPLIED MICROBIOLOGY*, 25(1):84–99.
- Kampschreur, M. J., Temmink, H., Kleerebezem, R., Jetten, M. S. M., and van Loosdrecht, M. C. M. (2009). Nitrous oxide emission during wastewater treatment. *WATER RESEARCH*, 43(17):4093–4103.
- Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., and Tanabe, M. (2019). New approach for understanding genome variations in KEGG. *NUCLEIC ACIDS RESEARCH*, 47(D1):D590–D595.

## Bibliography

---

- Kang, A. J., Brown, A. K., Wong, C. S., Huang, Z., and Yuan, Q. (2018). Variation in bacterial community structure of aerobic granular and suspended activated sludge in the presence of the antibiotic sulfamethoxazole. *BIORESOURCE TECHNOLOGY*, 261:322–328.
- Katherine Jungles, M., Figueroa, M., Morales, N., Val del Rio, A., Ribeiro da Costa, R. H., Luis Campos, J., Mosquera-Corral, A., and Mendez, R. (2011). Start up of a pilot scale aerobic granular reactor for organic matter and nitrogen removal. *JOURNAL OF CHEMICAL TECHNOLOGY AND BIOTECHNOLOGY*, 86(5):763–768.
- Khan, A. A., Ahmad, M., and Giesen, A. (2015). NEREDA (R): an emerging technology for sewage treatment. *WATER PRACTICE AND TECHNOLOGY*, 10(4):799–805.
- Kim, E. and Yi, H. (2016). *Rhodobacter* sp. lpb0142, direct submission. accession CP017781.
- Kim, J. M., Lee, H. J., Lee, D. S., and Jeon, C. O. (2013). Characterization of the Denitrification-Associated Phosphorus Uptake Properties of “Candidatus *Accumulibacter phosphatis*” Clades in Sludge Subjected to Enhanced Biological Phosphorus Removal. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 79(6):1969–1979.
- Klenk, H.-P., Huntemann, M., Clum, A., Pillay, M., Palaniappan, K., Varghese, N., Mikhailova, N., Stamatis, D., Reddy, T., Daum, C., Shapiro, N., I. N., Kyrpides, N., and Woyke, T. (2017). *Propionicimonas paludicola*, direct submission. accession PDJC01000001.
- Koch, G., Kuhni, M., Gujer, W., and Siegrist, H. (2000). Calibration and validation of Activated Sludge Model No. 3 for Swiss municipal wastewater. *WATER RESEARCH*, 34(14):3580–3590.
- Kong, Y., Nielsen, J., and Nielsen, P. (2004). Microautoradiographic study of Rhodocyclus-related polyphosphate accumulating bacteria in full-scale enhanced biological phosphorus removal plants. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 70(9):5383–5390.
- Kong, Y., Nielsen, J., and Nielsen, P. (2005). Identity and ecophysiology of uncultured actinobacterial polyphosphate-accumulating organisms in full-scale enhanced biological phosphorus removal plants. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 71(7):4076–4085.
- Kong, Y., Xia, Y., Nielsen, J., and Nielsen, P. (2006). Ecophysiology of a group of uncultured Gammaproteobacterial glycogen-accumulating organisms in full-scale enhanced biological phosphorus removal wastewater treatment plants. *ENVIRONMENTAL MICROBIOLOGY*, 8(3):479–489.
- Kong, Y., Xia, Y., Nielsen, J. L., and Nielsen, P. H. (2007). Structure and function of the microbial community in a full-scale enhanced biological phosphorus removal plant. *MICROBIOLOGY-SGM*, 153(12):4061–4073.
- Kong, Y., Xia, Y., and Nielsen, P. H. (2008). Activity and identity of fermenting microorganisms in full-scale biological nutrient removing wastewater treatment plants. *ENVIRONMENTAL MICROBIOLOGY*, 10(8):2008–2019.

- Koren, S. and Phillippy, A. M. (2015). One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *CURRENT OPINION IN MICROBIOLOGY*, 23:110–120.
- Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J. T., Ganapathy, G., Wang, Z., Rasko, D. A., McCombie, W. R., Jarvis, E. D., and Phillippy, A. M. (2012). Hybrid error correction and de novo assembly of single-molecule sequencing reads. *NATURE BIOTECHNOLOGY*, 30(7):692+.
- Kornberg, A., Rao, N., and Ault-Riche, D. (1999). Inorganic polyphosphate: A molecule of many functions. *ANNUAL REVIEW OF BIOCHEMISTRY*, 68:89–125.
- Koskella, B. and Brockhurst, M. A. (2014). Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS MICROBIOLOGY REVIEWS*, 38(5):916–931.
- Kristiansen, R., Nguyen, H. T. T., Saunders, A. M., Nielsen, J. L., Wimmer, R., Le, V. Q., McIlroy, S. J., Petrovski, S., Seviour, R. J., Calteau, A., Nielsen, K. L., and Nielsen, P. H. (2013). A metabolic model for members of the genus *Tetrasphaera* involved in enhanced biological phosphorus removal. *ISME JOURNAL*, 7(3):543–554.
- Kriventseva, E. V., Zdobnov, E. M., Simão, F. A., Ioannidis, P., and Waterhouse, R. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19):3210–3212.
- Kuba, T., VanLoosdrecht, M., and Heijnen, J. (1996). Phosphorus and nitrogen removal with minimal cod requirement by integration of denitrifying dephosphatation and nitrification in a two-sludge system. *WATER RESEARCH*, 30(7):1702–1710.
- Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., and Hugenholtz, P. (2008). A Bioinformatician's Guide to Metagenomics. *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS*, 72(4):557–578.
- Kuypers, M. M. M., Marchant, H. K., and Kartal, B. (2018). The microbial nitrogen-cycling network. *NATURE REVIEWS MICROBIOLOGY*, 16(5):263–276.
- Kyrpides, N., Huntemann, M., Han, J., Chen, A., Mavromatis, K., Markowitz, V., Palaniappan, K., Ivanova, N., Schaumberg, A., Pati, A., Liolios, K., Nordberg, H., Cantor, M., Hua, S., and Woyke, T. (2014). *Thiothrix lacustris* dsm 21227, direct submission. accession JHYQ01000001.
- Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *NATURE METHODS*, 9(4):357–U54.
- Larsen, P., Nielsen, J. L., Otzenj, D., and Nielsen, P. H. (2008). Amyloid-like adhesins produced by floc-forming and filamentous bacteria in activated sludge. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 74(5):1517–1526.

## Bibliography

---

- Larsen, T. and Harremoes, P. (1994). Degradation mechanisms of colloidal organic-matter in biofilm reactors. *WATER RESEARCH*, 28(6):1443–1452.
- Lawson, C. E., Wu, S., Bhattacharjee, A. S., Hamilton, J. J., McMahon, K. D., Goel, R., and Noguera, D. R. (2017). Metabolic network analysis reveals microbial community interactions in anammox granules. *NATURE COMMUNICATIONS*, 8.
- Le, S., Josse, J., and Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *JOURNAL OF STATISTICAL SOFTWARE*, 25(1):1–18.
- Lemaire, R., Webb, R. I., and Yuan, Z. (2008). Micro-scale observations of the structure of aerobic microbial granules used for the treatment of nutrient-rich industrial wastewater. *ISME JOURNAL*, 2(5):528–541.
- Leventhal, G. E., Boix, C., Kuechler, U., Enke, T. N., Sliwerska, E., Holliger, C., and Cordero, O. X. (2018). Strain-level diversity drives alternative community types in millimetre-scale granular biofilms. *NATURE MICROBIOLOGY*, 3(11):1295–1303.
- Levine, A., Tchobanoglous, G., and Asano, T. (1991). Size distributions of particulate contaminants in waste-water and their impact on treatability. *WATER RESEARCH*, 25(8):911–922.
- Levine, A. D., Tchobanoglous, G., and Asano, T. (1985). Characterization of the size distribution of contaminants in wastewater: Treatment and reuse implications. *Journal (Water Pollution Control Federation)*, 57(7):805–816.
- Lewis, K. (2001). Riddle of biofilm resistance. *ANTIMICROBIAL AGENTS AND CHEMOTHERAPY*, 45(4):999–1007.
- Li, A.-j., Yang, S.-f., Li, X.-y., and Gu, J.-d. (2008). Microbial population dynamics during aerobic sludge granulation at different organic loading rates. *WATER RESEARCH*, 42(13):3552–3560.
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *BIOINFORMATICS*, 31(10):1674–1676.
- Li, H. (2016). Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*, 32(14):2103–2110.
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18):3094–3100.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Proc. . G. P. D. (2009). The Sequence Alignment/Map format and SAMtools. *BIOINFORMATICS*, 25(16):2078–2079.
- Li, J., Ding, L.-B., Cai, A., Huang, G.-X., and Horn, H. (2014). Aerobic Sludge Granulation in a Full-Scale Sequencing Batch Reactor. *BIOMED RESEARCH INTERNATIONAL*.

- Lin, Y., Liu, Y., and Tay, J. (2003). Development and characteristics of phosphorus-accumulating microbial granules in sequencing batch reactors. *APPLIED MICROBIOLOGY AND BIOTECHNOLOGY*, 62(4):430–435.
- Lin, Y. M., Sharma, P. K., and van Loosdrecht, M. C. M. (2013). The chemical and mechanical differences between alginate-like exopolysaccharides isolated from aerobic flocculent sludge and aerobic granular sludge. *WATER RESEARCH*, 47(1):57–65.
- Liu, J., Li, J., Tao, Y., Sellamuthu, B., and Walsh, R. (2017). Analysis of bacterial, fungal and archaeal populations from a municipal wastewater treatment plant developing an innovative aerobic granular sludge process. *WORLD JOURNAL OF MICROBIOLOGY & BIOTECHNOLOGY*, 33(1).
- Liu, Y. and Liu, Q. (2006). Causes and control of filamentous growth in aerobic granular sludge sequencing batch reactors. *BIOTECHNOLOGY ADVANCES*, 24(1):115–127.
- Liu, Y., Liu, Y., and Tay, J. (2004). The effects of extracellular polymeric substances on the formation and stability of biogranules. *APPLIED MICROBIOLOGY AND BIOTECHNOLOGY*, 65(2):143–148.
- Liu, Y., Wang, Z., Qin, L., Liu, Y., and Tay, J. (2005). Selection pressure-driven aerobic granulation in a sequencing batch reactor. *APPLIED MICROBIOLOGY AND BIOTECHNOLOGY*, 67(1):26–32.
- Liu, Y.-Q., Kong, Y., Tay, J.-H., and Zhu, J. (2011). Enhancement of start-up of pilot-scale granular SBR fed with real wastewater. *SEPARATION AND PURIFICATION TECHNOLOGY*, 82:190–196.
- Liu, Y.-Q., Moy, B., Kong, Y.-H., and Tay, J.-H. (2010). Formation, physical characteristics and microbial community structure of aerobic granules in a pilot-scale sequencing batch reactor for real wastewater treatment. *ENZYME AND MICROBIAL TECHNOLOGY*, 46(6):520–525.
- Liu, Y.-Q. and Tay, J.-H. (2012). The competition between flocculent sludge and aerobic granules during the long-term operation period of granular sludge sequencing batch reactor. *Environmental Technology*, 33(23):2619–2626. PMID: 23437662.
- Lochmatter, S. and Holliger, C. (2013). *Optimization of Reactor Startup and Nitrogen Removal of Aerobic Granular Sludge Systems*. PhD thesis, EPFL, 1015 Lausanne. supervised by Christof Holliger.
- Lochmatter, S. and Holliger, C. (2014). Optimization of operation conditions for the startup of aerobic granular sludge reactors biologically removing carbon, nitrogen, and phosphorous. *WATER RESEARCH*, 59:58–70.
- Lochmatter, S., Maillard, J., and Holliger, C. (2014). Nitrogen Removal over Nitrite by Aeration Control in Aerobic Granular Sludge Sequencing Batch Reactors. *INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH*, 11(7):6955–6978.

## Bibliography

---

- Lu, H., Oehmen, A., Viridis, B., Keller, J., and Yuan, Z. (2006). Obtaining highly enriched cultures of *Candidatus Accumulibacter phosphatis* through alternating carbon sources. *WATER RESEARCH*, 40(20):3838–3848.
- Luecker, S., Wagner, M., Maixner, F., Pelletier, E., Koch, H., Vacherie, B., Rattei, T., Damste, J. S. S., Spieck, E., Le Paslier, D., and Daims, H. (2010). A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 107(30):13479–13484.
- Lv, J., Wang, Y., Zhong, C., Li, Y., Hao, W., and Zhu, J. (2014). The microbial attachment potential and quorum sensing measurement of aerobic granular activated sludge and flocculent activated sludge. *BIORESOURCE TECHNOLOGY*, 151:291–296.
- Ma, B., Wang, S., Cao, S., Miao, Y., Jia, F., Du, R., and Peng, Y. (2016). Biological nitrogen removal from sewage via anammox: Recent advances. *BIORESOURCE TECHNOLOGY*, 200:981–990.
- Mao, Y., Yu, K., Xia, Y., Chao, Y., and Zhang, T. (2014). Genome Reconstruction and Gene Expression of “*Candidatus Accumulibacter phosphatis*” Clade IB Performing Biological Phosphorus Removal. *ENVIRONMENTAL SCIENCE & TECHNOLOGY*, 48(17):10363–10371.
- Margot, J., Magnet, A., Thonney, D., Chèvre, N., Alencastro, D., Felipe, L., and Rossi, L. (2011). Traitement des micropolluants dans les eaux usées - rapport final sur les essais pilotes à la step de vidy (lausanne). page 106.
- Marques, R., Santos, J., Nguyen, H., Carvalho, G., Noronha, J. P., Nielsen, P. H., Reis, M. A. M., and Oehmen, A. (2017). Metabolism and ecological niche of *Tetrasphaera* and *Ca. Accumulibacter* in enhanced biological phosphorus removal. *WATER RESEARCH*, 122:159–171.
- Martin, H., Ivanova, N., Kunin, V., Warnecke, F., Barry, K. and He, S., Salamov, A., Szeto, E., Dalin, E., Pangilinan, J., Lapidus, A., Lowry, S., Kyrpides, N., McMahon, K., and Hugenholtz, P. (2014). Complete sequence of chromosome of *Candidatus Accumulibacter phosphatis* clade iia str. uw-1. accession CP001715.
- Martin, H. G., Ivanova, N., Kunin, V., Warnecke, F., Barry, K. W., McHardy, A. C., Yeates, C., He, S., Salamov, A. A., Szeto, E., Dalin, E., Putnam, N. H., Shapiro, H. J., Pangilinan, J. L., Rigoutsos, I., Kyrpides, N. C., Blackall, L. L., McMahon, K. D., and Hugenholtz, P. (2006). Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *NATURE BIOTECHNOLOGY*, 24(10):1263–1269.
- Martins, A., Picioreanu, C., Heijnen, J., and van Loosdrecht, M. (2004). Three-dimensional dual-morphotype species Modeling of activated sludge flocs. *ENVIRONMENTAL SCIENCE & TECHNOLOGY*, 38(21):5632–5641.
- Martins, A. M. P., Karahan, O., and van Loosdrecht, M. C. M. (2011). Effect of polymeric substrate on sludge settleability. *WATER RESEARCH*, 45(1):263–273.

- Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing dna. *Proceedings of the National Academy of Sciences*, 74(2):560–564.
- McCann, K. (2000). The diversity-stability debate. *NATURE*, 405(6783):228–233.
- McHardy, A. C., Garcia Martin, H., Tsirigos, A., Hugenholtz, P., and Rigoutsos, I. (2007). Accurate phylogenetic classification of variable-length DNA fragments. *NATURE METHODS*, 4(1):63–72.
- McIlroy, J., Karst M, S., Albertsen, M., Mcilroy, S., and Nielsen, H. (2018). *Micropruina glyco- genica*, direct submission. accession LT985188.
- McIlroy, S., Saunders, A., Albertsen, M., Nierychlo, M., McIlroy, B., Hansen, A., Karst, S., Nielsen, J., and Nielsen, P. (2015a). MiDAS: the field guide to the microbes of activated sludge. Database.
- McIlroy, S. J., Albertsen, M., Andresen, E. K., Saunders, A. M., Kristiansen, R., Stokholm- Bjerregaard, M., Nielsen, K. L., and Nielsen, P. H. (2014). ‘Candidatus Competibacter’- lineage genomes retrieved from metagenomes reveal functional metabolic diversity. *ISME JOURNAL*, 8(3):613–624.
- McIlroy, S. J., Kirkegaard, R. H., McIlroy, B., Nierychlo, M., Kristensen, J. M., Karst, S. M., Albertsen, M., and Nielsen, P. H. (2017). MiDAS 2.0: an ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups. *DATABASE-THE JOURNAL OF BIOLOGICAL DATABASES AND CURATION*.
- McIlroy, S. J., Nittami, T., Kanai, E., Fukuda, J., Saunders, A. M., and Nielsen, P. H. (2015b). Re-appraisal of the phylogeny and fluorescence in situ hybridization probes for the analysis of the Competibacteraceae in wastewater treatment systems. *ENVIRONMENTAL MICROBIOLOGY REPORTS*, 7(2):166–174.
- McIlroy, S. J., Onetto, C. A., McIlroy, B., Herbst, F.-A., Dueholm, M. S., Kirkegaard, R. H., Fernando, E., Karst, S. M., Nierychlo, M., Kristensen, J. M., Eales, K. L., Grbin, P. R., Wimmer, R., and Nielsen, P. H. (2018). Genomic and in Situ Analyses Reveal the *Micropruina* spp. as Abundant Fermentative Glycogen Accumulating Organisms in Enhanced Biological Phosphorus Removal Systems. *FRONTIERS IN MICROBIOLOGY*, 9.
- McLean, J. S., Bor, B., To, T. T., Liu, Q., Kerns, K. A., Solden, L., Wrighton, K., He, X., and Shi, W. (2018). Evidence of independent acquisition and adaption of ultra-small bacteria to human hosts across the highly diverse yet reduced genomes of the phylum saccharibacteria. *bioRxiv*.
- McIlroy, S. J., Kirkegaard, R. H., Dueholm, M. S., Fernando, E., Karst, S. M., Albertsen, M., and Nielsen, P. H. (2017). Culture-Independent Analyses Reveal Novel Anaerolineaceae as Abundant Primary Fermenters in Anaerobic Digesters Treating Waste Activated Sludge. *FRONTIERS IN MICROBIOLOGY*, 8.

## Bibliography

---

- McMahon, K., Dojka, M., Pace, N., Jenkins, D., and Keasling, J. (2002). Polyphosphate kinase from activated sludge performing enhanced biological phosphorus removal. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 68(10):4971–4978.
- McMahon, K. D., Yilmaz, S., He, S., Gall, D. L., Jenkins, D., and Keasling, J. D. (2007). Polyphosphate kinase genes from full-scale activated sludge plants. *APPLIED MICROBIOLOGY AND BIOTECHNOLOGY*, 77(1):167–173.
- Merritt, J., Qi, F., Goodman, S., Anderson, M., and Shi, W. (2003). Mutation of luxS affects biofilm formation in *Streptococcus mutans*. *INFECTION AND IMMUNITY*, 71(4):1972–1979.
- Metcalf & Eddy (2014). *Wastewater engineering: treatment and resource recovery*. McGraw Hill Education, Boston, 5<sup>th</sup> edition. revised by Abu-Orf, M., Tchobanoglous, G., Stensel, H. D., Tsuchihashi, R., Burton, F., Bowden, G., Pfrang, W.
- MiDAS (2015). Sampling and DNA extraction from wastewater activated sludge. [http://www.midasfieldguide.org/download/protocols/dna\\_extraction/sampling\\_and\\_dna\\_extraction\\_from\\_wastewater\\_activated\\_sludge.pdf](http://www.midasfieldguide.org/download/protocols/dna_extraction/sampling_and_dna_extraction_from_wastewater_activated_sludge.pdf). [version 7.1].
- Middelboe, M. (2000). Bacterial growth rate and marine virus-host dynamics. *MICROBIAL ECOLOGY*, 40(2):114–124.
- Mielczarek, A. T., Nguyen, H. T. T., Nielsen, J. L., and Nielsen, P. H. (2013). Population dynamics of bacteria involved in enhanced biological phosphorus removal in Danish wastewater treatment plants. *WATER RESEARCH*, 47(4):1529–1544.
- Mihciokur, H. and Oguz, M. (2016). Removal of oxytetracycline and determining its biosorption properties on aerobic granular sludge. *ENVIRONMENTAL TOXICOLOGY AND PHARMACOLOGY*, 46:174–182.
- Mino, T., Van Loosdrecht, M., and Heijnen, J. (1998). Microbiology and biochemistry of the enhanced biological phosphate removal process. *WATER RESEARCH*, 32(11):3193–3207.
- Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D., Chang, H.-Y., El-Gebali, S., Fraser, M. I., Gough, J., Haft, D. R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., Natale, D. A., Necci, M., Nuka, G., Orengo, C., Pandurangan, A. P., Paysan-Lafosse, T., Pesseat, S., Potter, S. C., Qureshi, M. A., Rawlings, N. D., Redaschi, N., Richardson, L. J., Rivoire, C., Salazar, G. A., Sangrador-Vegas, A., Sigrist, C. J. A., Sillitoe, I., Sutton, G. G., Thanki, N., Thomas, P. D., Tosatto, S. C. E., Yong, S.-Y., and Finn, R. D. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *NUCLEIC ACIDS RESEARCH*, 47(D1):D351–D360.
- M.N., P., Spanjers, H., Baetens, D., Nowak, O., Gillot, S., Nouwen, J., and Schuttina, N. (2004). Wastewater characteristics in Europe - A survey. *European Water Management Online*, 4:10 p.

- Moreno-Vivian, C., Cabello, P., Martinez-Luque, M., Blasco, R., and Castillo, F. (1999). Prokaryotic nitrate reduction: Molecular properties and functional distinction among bacterial nitrate reductases. *JOURNAL OF BACTERIOLOGY*, 181(21):6573–6584.
- Morgenroth, E., Kommedal, R., and Harremoes, P. (2002). Processes and modeling of hydrolysis of particulate organic matter in aerobic wastewater treatment - a review. *WATER SCIENCE AND TECHNOLOGY*, 45(6):25–40. 5th Seminar on Activated Sludge Modelling, KOLLEKOLLE, DENMARK, SEP 10-12, 2001.
- Morgenroth, E., Sherden, T., van Loosdrecht, M., Heijnen, J., and Wilderer, P. (1997). Aerobic granular sludge in a sequencing batch reactor. *WATER RESEARCH*, 31(12):3191–3194.
- Morohoshi, T., Maruo, T., Shirai, Y., Kato, J., Ikeda, T., Takiguchi, N., Ohtake, H., and Kuroda, A. (2002). Accumulation of inorganic polyphosphate in *phoU* mutants of *Escherichia coli* and *Synechocystis* sp strain PCC6803. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 68(8):4107–4110.
- Mosquera-Corral, A., Arrojo, B., Figueroa, M., Campos, J. L., and Mendez, R. (2011). Aerobic granulation in a mechanical stirred SBR: treatment of low organic loads. *WATER SCIENCE AND TECHNOLOGY*, 64(1):155–161.
- Mosquera-Corral, A., Montras, A., Heijnen, J., and van Loosdrecht, M. (2003). Degradation of polymers in a biofilm airlift suspension reactor. *WATER RESEARCH*, 37(3):485–492.
- Mulder, A., Vandergraaf, A., Robertson, L., and Kuenen, J. (1995). Anaerobic ammonium oxidation discovered in a denitrifying fluidized-bed reactor. *FEMS MICROBIOLOGY ECOLOGY*, 16(3):177–183.
- Muller, E., Narayanasamy, S., Zeimes, M., Laczny, C., Lebrun, L., Herold, M., Hicks, N., Gillece, J., Schupp, J., Keim, P., and Wilmes, P. (2017). First draft genome sequence of a novel species strain belonging to the zoogloea genus and its gene expression in situ. BioSample: SAMN06480675.
- Murray, R. and Stackbrandt, E. (1995). Taxonomic note - implementation of the provisional status Candidatus for incompletely described prokaryotes. *INTERNATIONAL JOURNAL OF SYSTEMATIC BACTERIOLOGY*, 45(1):186–187.
- Muszynski, A. and Zaleska-Radziwill, M. (2015). Polyphosphate accumulating organisms in treatment plants with different wastewater composition. *ARCHITECTURE CIVIL ENGINEERING ENVIRONMENT*, 8(4):99–105.
- Nancharaiyah, Y. V. and Reddy, G. K. K. (2018). Aerobic granular sludge technology: Mechanisms of granulation and biotechnological applications. *BIORESOURCE TECHNOLOGY*, 247:1128–1143.
- Nawrocki, E. P. and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *BIOINFORMATICS*, 29(22):2933–2935.

## Bibliography

---

- Nguyen, H. T. T., Le, V. Q., Hansen, A. A., Nielsen, J. L., and Nielsen, P. H. (2011). High diversity and abundance of putative polyphosphate-accumulating Tetrasphaera-related bacteria in activated sludge systems. *FEMS MICROBIOLOGY ECOLOGY*, 76(2):256–267.
- Nguyen, H. T. T., Nielsen, J. L., and Nielsen, P. H. (2012). Candidatus Halomonas phosphatis, a novel polyphosphate-accumulating organism in full-scale enhanced biological phosphorus removal plants. *ENVIRONMENTAL MICROBIOLOGY*, 14(10, SI):2826–2837.
- Ni, B.-J., Xie, W.-M., Liu, S.-G., Yu, H.-Q., Wang, Y.-Z., Wang, G., and Dai, X.-L. (2009). Granulation of activated sludge in a pilot-scale sequencing batch reactor for the treatment of low-strength municipal wastewater. *WATER RESEARCH*, 43(3):751–761.
- Nielsen, H. B., Almeida, M., Juncker, A. S., Rasmussen, S., Li, J., Sunagawa, S., Plichta, D. R., Gautier, L., Pedersen, A. G., Le Chatelier, E., Pelletier, E., Bonde, I., Nielsen, T., Manichanh, C., Arumugam, M., Batto, J.-M., dos Santos, M. B. Q., Blom, N., Borruel, N., Burgdorf, K. S., Boumezbeur, F., Casellas, F., Dore, J., Dworzynski, P., Guarner, F., Hansen, T., Hildebrand, E., Kaas, R. S., Kennedy, S., Kristiansen, K., Kultima, J. R., Leonard, P., Levenez, F., Lund, O., Moumen, B., Le Paslier, D., Pons, N., Pedersen, O., Prifti, E., Qin, J., Raes, J., Sorensen, S., Tap, J., Tims, S., Ussery, D. W., Yamada, T., Renault, P., Sicheritz-Ponten, T., Bork, P., Wang, J., Brunak, S., Ehrlich, S. D., and Consortium, M. (2014). Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *NATURE BIOTECHNOLOGY*, 32(8):822–828.
- Nielsen, J. L., Nguyen, H., Meyer, R. L., and Nielsen, P. H. (2012a). Identification of glucose-fermenting bacteria in a full-scale enhanced biological phosphorus removal plant by stable isotope probing. *MICROBIOLOGY-SGM*, 158(7):1818–1825.
- Nielsen, P., de Muro, M., and Nielsen, J. (2000). Studies on the in situ physiology of Thiothrix spp. present in activated sludge. *ENVIRONMENTAL MICROBIOLOGY*, 2(4):389–398.
- Nielsen, P. H. (2017). Microbial biotechnology and circular economy in wastewater treatment. *MICROBIAL BIOTECHNOLOGY*, 10(5, SI):1102–1105.
- Nielsen, P. H., Mielczarek, A. T., Kragelund, C., Nielsen, J. L., Saunders, A. M., Kong, Y., Hansen, A. A., and Vollertsen, J. (2010). A conceptual ecosystem model of microbial communities in enhanced biological phosphorus removal plants. *WATER RESEARCH*, 44(17, SI):5070–5088.
- Nielsen, P. H., Saunders, A. M., Hansen, A. A., Larsen, P., and Nielsen, J. L. (2012b). Microbial communities involved in enhanced biological phosphorus removal from wastewater - a model system in environmental biotechnology. *CURRENT OPINION IN BIOTECHNOLOGY*, 23(3):452–459.
- Novak, L., Larrea, L., Wanner, J., and Garciaheras, J. (1993). Non-filamentous activated-sludge bulking in a laboratory-scale system. *WATER RESEARCH*, 27(8):1339–1346.
- Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., Prjibelsky, A., Pyshkin, A., Sirotkin, A., Sirotkin, Y., Stepanauskas, R., McLean, J., Lasken, R., Clingenpeel,

- S. R., Woyke, T., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. (2013). Assembling genomes and mini-metagenomes from highly chimeric reads. In Deng, M., Jiang, R., Sun, F., and Zhang, X., editors, *Research in Computational Molecular Biology*, pages 158–170, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *GENOME RESEARCH*, 27(5):824–834.
- Oehmen, A., Carvalho, G., Freitas, F., and Reis, M. A. M. (2010). Assessing the abundance and activity of denitrifying polyphosphate accumulating organisms through molecular and chemical techniques. *WATER SCIENCE AND TECHNOLOGY*, 61(8):2061–2068.
- Oehmen, A., Lemos, P. C., Carvalho, G., Yuan, Z., Keller, J., Blackall, L. L., and Reis, M. A. M. (2007). Advances in enhanced biological phosphorus removal: From micro to macro scale. *WATER RESEARCH*, 41(11):2271–2300.
- OFEV, editor (2011). *OEaux*. Le Conseil fédéral suisse. [du 28 octobre 1998 (Etat le 1er août 2011)].
- Okabe, S., Kindaichi, T., and Ito, T. (2004). Mar-fish&mdash;an ecophysiological approach to link phylogenetic affiliation and *in situ* metabolic activity of microorganisms at a single-cell resolution. *Microbes and Environments*, 19(2):83–98.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P. R., O’Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H. (2018). *vegan: Community Ecology Package*. R package version 2.5-2.
- Olson, N. D., Treangen, T. J., Hill, C. M., Cepeda-Espinoza, V., Ghurye, J., Koren, S., and Pop, M. (2017). Metagenomic assembly through the lens of validation: recent advances in assessing and improving the quality of genomes assembled from metagenomes. *Briefings in Bioinformatics*.
- Ono, Y., Asai, K., and Hamada, M. (2013). PBSIM: PacBio reads simulator-toward accurate genome assembly. *BIOINFORMATICS*, 29(1):119–121.
- Oshiki, M., Satoh, H., and Okabe, S. (2016). Ecology and physiology of anaerobic ammonium oxidizing bacteria. *ENVIRONMENTAL MICROBIOLOGY*, 18(9, SI):2784–2796.
- Oyserman, B. O., Noguera, D. R., del Rio, T. G., Tringe, S. G., and McMahon, K. D. (2016). Metatranscriptomic insights on gene expression and regulatory controls in *Candidatus Accumulibacter phosphatis*. *ISME JOURNAL*, 10(4):810–822.
- Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *NATURE BIOTECHNOLOGY*, 36(10):996+.

## Bibliography

---

- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *GENOME RESEARCH*, 25(7):1043–1055.
- Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B. J., Evans, P. N., Hugenholtz, P., and Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *NATURE MICROBIOLOGY*, 2(11):1533–1542.
- Peterson, S. B., Warnecke, F., Madejska, J., McMahon, K. D., and Hugenholtz, P. (2008). Environmental distribution and population biology of *Candidatus Accumulibacter*, a primary agent of biological phosphorus removal. *ENVIRONMENTAL MICROBIOLOGY*, 10(10):2692–2703.
- Petratits, P. and Dudgeon, S. (2004). Detection of alternative stable states in marine communities. *JOURNAL OF EXPERIMENTAL MARINE BIOLOGY AND ECOLOGY*, 300(1-2):343–371.
- Pochana, K. and Keller, J. (1999). Study of factors affecting simultaneous nitrification and denitrification (SND). *WATER SCIENCE AND TECHNOLOGY*, 39(6):61–68. 3rd AWWA/IAWQ Regional Conference on Biological Nutrient Removal (BNR3), BRISBANE, AUSTRALIA, NOV 30-DEC 04, 1997.
- Pretty, J., Mason, C., Nedwell, D., Hine, R., Leaf, S., and Dils, R. (2003). Environmental costs of freshwater eutrophication in England and Wales. *ENVIRONMENTAL SCIENCE & TECHNOLOGY*, 37(2):201–208.
- Pride, D., Meinersmann, R., Wassenaar, T., and Blaser, M. (2003). Evolutionary implications of microbial genome tetranucleotide frequency biases. *GENOME RESEARCH*, 13(2):145–158.
- Pronk, M., Abbas, B., Al-zuhairy, S. H. K., Kraan, R., Kleerebezem, R., and van Loosdrecht, M. C. M. (2015a). Effect and behaviour of different substrates in relation to the formation of aerobic granular sludge. *APPLIED MICROBIOLOGY AND BIOTECHNOLOGY*, 99(12):5257–5268.
- Pronk, M., de Kreuk, M. K., de Bruin, B., Kamminga, P., Kleerebezem, R., and van Loosdrecht, M. C. M. (2015b). Full scale performance of the aerobic granular sludge process for sewage treatment. *WATER RESEARCH*, 84:207–217.
- Pronk, M., Giesen, A., Thompson, A., Robertson, S., and van Loosdrecht, M. (2017). Aerobic granular biomass technology: advancements in design, applications and further developments. *WATER PRACTICE AND TECHNOLOGY*, 12(4):987–996.
- Qiu, G., Zuniga-Montanez, R., Law, Y., Thi, S. S., Nguyen, T. Q. N., Eganathan, K., Liu, X., Nielsen, P. H., Williams, R. B. H., and Wuertz, S. (2019). Polyphosphate-accumulating organisms in full-scale tropical wastewater treatment plants use diverse carbon sources. *WATER RESEARCH*, 149:496–510.
- Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., and Gu, Y. (2012). A tale of three next generation sequencing platforms:

- comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC GENOMICS*, 13.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1):D590–D596.
- Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rao, N. N., Gomez-Garcia, M. R., and Kornberg, A. (2009). Inorganic Polyphosphate: Essential for Growth and Survival. *ANNUAL REVIEW OF BIOCHEMISTRY*, 78:605–647.
- Ricker, N., Shen, S. Y., Goordial, J., Jin, S., and Fulthorpe, R. R. (2016). PacBio SMRT assembly of a complex multi-replicon genome reveals chlorocatechol degradative operon in a region of genome plasticity. *GENE*, 586(2):239–247.
- Rittmann, B. and McCarty, P. (1981). Substrate flux into biofilms of any thickness. *Journal of Environmental Engineering, ASCE*, 107(4):831–849.
- Robertson, Struan and van der Roest, Helle and van Bentem, André (2015). Achieving sustainable wastewater treatment through innovation: an update on the Nereda technology. *Water21*, pages 39–41.
- Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. (2011). Integrative genomics viewer. *NATURE BIOTECHNOLOGY*, 29(1):24–26.
- Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., Buchanan, J., Desnues, C., Dinsdale, E., Edwards, R., Felts, B., Haynes, M., Liu, H., Lipson, D., Mahaffy, J., Belen Martin-Cuadrado, A., Mira, A., Nulton, J., Pasic, L., Rayhawk, S., Rodriguez-Mueller, J., Rodriguez-Valera, F., Salamon, P., Srinagesh, S., Thingstad, T. E., Tran, T., Thurber, R. V., Willner, D., Youle, M., and Rohwer, F. (2010). Viral and microbial community dynamics in four aquatic environments. *ISME JOURNAL*, 4(6):739–751.
- Rusanowska, P., Cydzik-Kwiatkowska, A., Swiatczak, P., and Wojnowska-Baryla, I. (2019). Changes in extracellular polymeric substances (EPS) content and composition in aerobic granule size-fractions during reactor cycles at different organic loads. *BIORESOURCE TECHNOLOGY*, 272:188–193.
- Rusten, B., Odegaard, H., and Lundar, A. (1992). Treatment of dairy waste-water in a novel moving bed biofilm reactor. *WATER SCIENCE AND TECHNOLOGY*, 26(3-4):703–711. 16TH BIENNIAL CONF OF THE INTERNATIONAL ASSOC ON WATER POLLUTION RESEARCH AND CONTROL : WATER QUALITY INTERNATIONAL 92, WASHINGTON, DC, MAY 24-30, 1992.

## Bibliography

---

- Sakuragi, Y. and Kolter, R. (2007). Quorum-sensing regulation of the biofilm matrix genes (*pel*) of *Pseudomonas aeruginosa*. *JOURNAL OF BACTERIOLOGY*, 189(14):5383–5386.
- San Pedro, D., Mino, T., and Matsuo, T. (1994). Evaluation of the rate of hydrolysis of slowly biodegradable COD (SBCOD) using starch as substrate under anaerobic, anoxic and aerobic conditions. *WATER SCIENCE AND TECHNOLOGY*, 30(11):191–199.
- Sanger, F., Nicklen, S., and Coulson, A. (1977). DNA sequencing with chain-terminating inhibitors. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 74(12):5463–5467.
- Santos, M., Lemos, P., Reis, M., and Santos, H. (1999). Glucose metabolism and kinetics of phosphorus removal by the fermentative bacterium *Micrococcus phosphovorans*. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 65(9):3920–3928.
- Sarma, S. J., Tay, J. H., and Chu, A. (2017). Finding knowledge Gaps Aerobic Granulation Technology. *TRENDS IN BIOTECHNOLOGY*, 35(1):66–78.
- Saunders, A., Oehmen, A., Blackall, L., Yuan, Z., and Keller, J. (2003). The effect of GAOs (glycogen accumulating organisms) on anaerobic carbon requirements in full-scale Australian EBPR (enhanced biological phosphorus removal) plants. *WATER SCIENCE AND TECHNOLOGY*, 47(11):37–43. 3rd World Water Congress of the International-Water-Association, MELBOURNE, AUSTRALIA, APR 07-12, 2002.
- Saunders, A. M., Albertsen, M., Vollertsen, J., and Nielsen, P. H. (2016). The activated sludge ecosystem contains a core community of abundant organisms. *ISME JOURNAL*, 10(1):11–20.
- Schaefer, A., Val, D., Hanzelka, B., Cronan, J., and Greenberg, E. (1996). Generation of cell-to-cell signals in quorum sensing: Acyl homoserine lactone synthase activity of a purified *Vibrio fischeri* LuxI protein. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 93(18):9505–9509.
- Schmid, M., Frei, D., Patrignani, A., Schlapbach, R., Frey, J. E., Remus-Emsermann, M. N. P., and Ahrens, C. H. (2018). Pushing the limits of de novo genome assembly for complex prokaryotic genomes harboring very long, near identical repeats. *NUCLEIC ACIDS RESEARCH*, 46(17):8953–8965.
- Schwarzenbeck, N., Erley, R., and Wilderer, P. (2004). Aerobic granular sludge in an SBR-system treating wastewater rich in particulate matter. *WATER SCIENCE AND TECHNOLOGY*, 49(11-12):41–46. 5th IWA International Conference on Biofilm Systems, Cape Town, SOUTH AFRICA, SEP 14-18, 2003.
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14):2068–2069.
- Seviour, T., Pijuan, M., Nicholson, T., Keller, J., and Yuan, Z. (2009). Gel-forming exopolysaccharides explain basic differences between structures of aerobic sludge granules and floccular sludges. *WATER RESEARCH*, 43(18):4469–4478.

- Seviour, T., Yuan, Z., van Loosdrecht, M. C. M., and Lin, Y. (2012). Aerobic sludge granulation: A tale of two polysaccharides? *WATER RESEARCH*, 46(15):4803–4813.
- Seviour, T. W., Lambert, L. K., Pijuan, M., and Yuan, Z. (2011). Selectively inducing the synthesis of a key structural exopolysaccharide in aerobic granules by enriching for *Candidatus "Competibacter phosphatis"*. *APPLIED MICROBIOLOGY AND BIOTECHNOLOGY*, 92(6):1297–1305.
- Shade, A., Read, J. S., Youngblut, N. D., Fierer, N., Knight, R., Kratz, T. K., Lottig, N. R., Roden, E. E., Stanley, E. H., Stombaugh, J., Whitaker, R. J., Wu, C. H., and McMahon, K. D. (2012). Lake microbial communities are resilient after a whole-ecosystem disturbance. *ISME JOURNAL*, 6(12):2153–2167.
- Sharon, I., Morowitz, M. J., Thomas, B. C., Costello, E. K., Relman, D. A., and Banfield, J. F. (2013). Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *GENOME RESEARCH*, 23(1):111–120.
- Siripong, S. and Rittmann, B. E. (2007). Diversity study of nitrifying bacteria in full-scale municipal wastewater treatment plants. *WATER RESEARCH*, 41(5):1110–1120.
- Skennerton, C. (2012). Metagenome of two novel *accumulibacter* clades.
- Skennerton, C. T., Angly, F. E., Breitbart, M., Bragg, L., He, S., McMahon, K. D., Hugenholtz, P., and Tyson, G. W. (2011). Phage Encoded H-NS: A Potential Achilles Heel in the Bacterial Defence System. *PLOS ONE*, 6(5).
- Skennerton, C. T., Barr, J. J., Slater, F. R., Bond, P. L., and Tyson, G. W. (2015). Expanding our view of genomic diversity in *Candidatus Accumulibacter* clades. *ENVIRONMENTAL MICROBIOLOGY*, 17(5):1574–1585.
- Skennerton, C. T., Imelfort, M., and Tyson, G. W. (2013). Crass: identification and reconstruction of CRISPR from unassembled metagenomic data. *NUCLEIC ACIDS RESEARCH*, 41(10).
- Slaby, B. M., Hackl, T., Horn, H., Bayer, K., and Hentschel, U. (2017). Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. *ISME JOURNAL*, 11(11):2465–2478.
- Sloan, D. B., Bennett, G. M., Engel, P., Williams, D., and Ochman, H. (2013). Disentangling Associated Genomes. In DeLong, EF, editor, *MICROBIAL METAGENOMICS, METATRANSCRIPTOMICS, AND METAPROTEOMICS*, volume 531 of *Methods in Enzymology*, pages 445–464.
- Soliman, M. and Eldyasti, A. (2018). Ammonia-Oxidizing Bacteria (AOB): opportunities and applications—a review. *REVIEWS IN ENVIRONMENTAL SCIENCE AND BIO-TECHNOLOGY*, 17(2):285–321.

## Bibliography

---

- Soo, R. M., Skennerton, C. T., Sekiguchi, Y., Imelfort, M., Paech, S. J., Dennis, P. G., Steen, J. A., Parks, D. H., Tyson, G. W., and Hugenholtz, P. (2014). An Expanded Genomic Representation of the Phylum Cyanobacteria. *Genome Biology and Evolution*, 6(5):1031–1045.
- Soucy, S. M., Huang, J., and Gogarten, J. P. (2015). Horizontal gene transfer: building the web of life. *NATURE REVIEWS GENETICS*, 16(8):472–482.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9):1312–1313.
- Stern, A., Mick, E., Tirosh, I., Sagy, O., and Sorek, R. (2012). CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *GENOME RESEARCH*, 22(10):1985–1994.
- Stokholm-Bjerregaard, M., McIlroy, S. J., Nierychlo, M., Karst, S. M., Albertsen, M., and Nielsen, P. H. (2017). A Critical Assessment of the Microorganisms Proposed to be Important to Enhanced Biological Phosphorus Removal in Full-Scale Wastewater Treatment Systems. *FRONTIERS IN MICROBIOLOGY*, 8.
- Strous, M., Kuenen, J., and Jetten, M. (1999). Key physiology of anaerobic ammonium oxidation. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 65(7):3248–3250.
- Suresh, A., Grygolowicz-Pawlak, E., Pathak, S., Poh, L. S., Majid, M. b. A., Dominiak, D., Bugge, T. V., Gao, X., and Ng, W. J. (2018). Understanding and optimization of the flocculation process in biological wastewater treatment processes: A review. *CHEMOSPHERE*, 210:401–416.
- Sutherland, I. (2001). Exopolysaccharides in biofilms, flocs and related structures. *WATER SCIENCE AND TECHNOLOGY*, 43(6):77–86. 1st IWA International Conference on Microbial Extracellular Polymeric Substances, MULHEIM, GERMANY, SEP 18-20, 2000.
- Sutherland, I., Hughes, K., Skillman, L., and Tait, K. (2004). The interaction of phage and biofilms. *FEMS MICROBIOLOGY LETTERS*, 232(1):1–6.
- Sutherland, I. W. (1995). Polysaccharide lyases. *FEMS Microbiology Reviews*, 16(4):323–347.
- Swiatczak, P. and Cydzik-Kwiatkowska, A. (2018). Performance and microbial characteristics of biomass in a full-scale aerobic granular sludge wastewater treatment plant. *ENVIRONMENTAL SCIENCE AND POLLUTION RESEARCH*, 25(2):1655–1669. International Conference on Conservation Agriculture and Sustainable Land Use (CASLU), Budapest, HUNGARY, MAY 31-JUN 02, 2016.
- Szabo, E., Liebana, R., Hermansson, M., Modin, O., Persson, F., and Wilen, B.-M. (2017a). Comparison of the bacterial community composition in the granular and the suspended phase of sequencing batch reactors. *AMB EXPRESS*, 7.

- Szabo, E., Liebana, R., Hermansson, M., Modin, O., Persson, F., and Wilen, B.-M. (2017b). Microbial Population Dynamics and Ecosystem Functions of Anoxic/Aerobic Granular Sludge in Sequencing Batch Reactors Operated at Different Organic Loading Rates. *FRONTIERS IN MICROBIOLOGY*, 8.
- Tan, C. H., Koh, K. S., Xie, C., Tay, M., Zhou, Y., Williams, R., Ng, W. J., Rice, S. A., and Kjelleberg, S. (2014). The role of quorum sensing signalling in EPS production and the assembly of a sludge community into aerobic granules. *ISME JOURNAL*, 8(6):1186–1197.
- Teeling, H., Waldmann, J., Lombardot, T., Bauer, M., and Glockner, F. (2004). TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. *BMC BIOINFORMATICS*, 5.
- Teitzel, G. and Parsek, M. (2003). Heavy metal resistance of biofilm and planktonic *Pseudomonas aeruginosa*. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 69(4):2313–2320.
- Teschler, J. K., Zamorano-Sanchez, D., Utada, A. S., Warner, C. J. A., Wong, G. C. L., Linington, R. G., and Yildiz, F. H. (2015). Living in the matrix: assembly and control of *Vibrio cholerae* biofilms. *NATURE REVIEWS MICROBIOLOGY*, 13(5):255–268.
- Thingstad, T. and Lignell, R. (1997). Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *AQUATIC MICROBIAL ECOLOGY*, 13(1):19–27. 5th European Marine Microbiology Symposium, BERGEN, NORWAY, AUG 11-15, 1996.
- Third, K., Burnett, N., and Cord-Ruwisch, R. (2003). Simultaneous nitrification and denitrification using stored substrate (PHB) as the electron donor in an SBR. *BIOTECHNOLOGY AND BIOENGINEERING*, 83(6):706–720.
- Togna, A. and Singh, M. (1994). Biological vapor-phase treatment using biofilter and biotrickling filter reactors - practical operating regimes. *ENVIRONMENTAL PROGRESS*, 13(2):94–97.
- Tsai, Y.-C., Conlan, S., Deming, C., Segre, J. A., Kong, H. H., Korlach, J., Oh, J., and Progra, N. C. S. (2016). Resolving the Complexity of Human Skin Metagenomes Using Single-Molecule Sequencing. *MBIO*, 7(1).
- van der Roest, H., de Bruin, B., van Loosdrecht, M., and Pronk, M. (2011). Dutch sustainable nereda(r) technology takes an impressive step towards maturity. In Wanner, J. and Somlyódy, L., editors, *Design, operation and economics of large wastewater treatment plants*, pages 22–23. International Water Association.
- van Dijk, E. J. H., Pronk, M., and van Loosdrecht, M. C. M. (2018a). Controlling effluent suspended solids in the aerobic granular sludge process. *WATER RESEARCH*, 147:50–59.
- van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018b). The Third Revolution in Sequencing Technology. *TRENDS IN GENETICS*, 34(9):666–681.

## Bibliography

---

- van Kessel, M. A. H. J., Speth, D. R., Albertsen, M., Nielsen, P. H., Op den Camp, H. J. M., Kartal, B., Jetten, M. S. M., and Lucker, S. (2015). Complete nitrification by a single microorganism. *NATURE*, 528(7583):555+.
- van Nieuwenhuijzen, A., van der Graaf, J., Kampschreur, M., and Mels, A. (2004). Particle related fractionation and characterisation of municipal wastewater. *WATER SCIENCE AND TECHNOLOGY*, 50(12):125–132. 2nd International Specialised Conference on Nano and Micro Particles in Water and Wastewater Treatment, Zurich, SWITZERLAND, SEP 22-24, 2003.
- Vanveen, H., Abee, T., Kortstee, G., Pereira, H., Konings, W., and Zehnder, A. (1994). Generation of a proton motive force by the excretion of metal-phosphate in the polyphosphate-accumulating *Acinetobacter-Johnsonii* strain 210A. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 269(47):29509–29514.
- Vinh, L. V., Lang, T. V., Binh, L. T., and Hoai, T. V. (2015). A two-phase binning algorithm using l-mer frequency on groups of non-overlapping reads. *ALGORITHMS FOR MOLECULAR BIOLOGY*, 10.
- Vjayan, T. and Vadivelu, V. M. (2017). Effect of famine-phase reduced aeration on polyhydroxyalkanoate accumulation in aerobic granules. *BIORESOURCE TECHNOLOGY*, 245(A):970–976.
- von Canstein, H., Kelly, S., Li, Y., and Wagner-Dobler, I. (2002). Species diversity improves the efficiency of mercury-reducing biofilms under changing environmental conditions. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 68(6):2829–2837.
- von Sperling, M. (2007). *Wastewater characteristics, treatment and disposal*, volume 1. IWA Publishing, Alliance House, 12 Caxton Street, London SW1H 0QS, UK. Biological wastewater treatment series.
- Wagner, J., Guimaraes, L. B., Akaboci, T. R. V., and Costa, R. H. R. (2015a). Aerobic granular sludge technology and nitrogen removal for domestic wastewater treatment. *WATER SCIENCE AND TECHNOLOGY*, 71(7):1040–1046.
- Wagner, J., Weissbrodt, D. G., Manguin, V., Ribeiro da Costa, R. H., Morgenroth, E., and Derlon, N. (2015b). Effect of particulate organic substrate on aerobic granulation and operating conditions of sequencing batch reactors. *WATER RESEARCH*, 85:158–166.
- Wagner, M., Erhart, R., Manz, W., Amann, R., Lemmer, H., Wedi, D., and Schleifer, K. (1994). Development of a ribosomal-RNA-targeted oligonucleotide probe specific for the genus *Acinetobacter* and its application for in-situ monitoring in activated-sludge. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 60(3):792–800.
- Walter, G. (1980). Stability and structure of compartmental-models of ecosystems. *MATHEMATICAL BIOSCIENCES*, 51(1-2):1–10.

- Wang, F., Lu, S., Wei, Y., and Ji, M. (2009). Characteristics of aerobic granule and nitrogen and phosphorus removal in a SBR. *JOURNAL OF HAZARDOUS MATERIALS*, 164(2-3):1223–1227.
- Wang, M.-z., Zheng, X., He, H.-z., Shen, D.-s., and Feng, H.-j. (2012). Ecological roles and release patterns of acylated homoserine lactones in *Pseudomonas* sp HF-1 and their implications in bacterial bioaugmentation. *BIORESOURCE TECHNOLOGY*, 125:119–126.
- Wang, Q., Yao, R., Yuan, Q., Gong, H., Xu, H., Ali, N., Jin, Z., Zuo, J., and Wang, K. (2018). Aerobic granules cultivated with simultaneous feeding/draw mode and low-strength wastewater: Performance and bacterial community analysis. *BIORESOURCE TECHNOLOGY*, 261:232–239.
- Wang, X., Wen, X., Criddle, C., Yan, H., Zhang, Y., and Ding, K. (2010). Bacterial community dynamics in two full-scale wastewater treatment systems with functional stability. *JOURNAL OF APPLIED MICROBIOLOGY*, 109(4):1218–1226.
- Wang, Z., Guo, F., Mao, Y., Xia, Y., and Zhang, T. (2014a). Metabolic Characteristics of a Glycogen-Accumulating Organism in *Defluviicoccus* Cluster II Revealed by Comparative Genomics. *MICROBIAL ECOLOGY*, 68(4):716–728.
- Wang, Z., Zhang, X.-X., Lu, X., Liu, B., Li, Y., Long, C., and Li, A. (2014b). Abundance and Diversity of Bacterial Nitrifiers and Denitrifiers and Their Functional Genes in Tannery Wastewater Treatment Plants Revealed by High-Throughput Sequencing. *PLOS ONE*, 9(11).
- Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., and Venables, B. (2016). *gplots: Various R Programming Tools for Plotting Data*. R package version 3.0.1.
- Weissbrodt, D. G., Lochmatter, S., Ebrahimi, S., Rossi, P., Maillard, J., and Holliger, C. (2012). Bacterial selection during the formation of early-stage aerobic granules in wastewater treatment systems operated under wash-out dynamics. *FRONTIERS IN MICROBIOLOGY*, 3.
- Weissbrodt, D. G., Maillard, J., Brovelli, A., Chabreliè, A., May, J., and Holliger, C. (2014a). Multi-level Correlations in the Biological Phosphorus Removal Process: from Bacterial Enrichment to Conductivity-Based Metabolic Batch Tests and Polyphosphatase Assays. *BIOTECHNOLOGY AND BIOENGINEERING*, 111(12):2421–2435.
- Weissbrodt, D. G., Neu, T. R., Kuhlicke, U., Rappaz, Y., and Holliger, C. (2013). Assessment of bacterial and structural dynamics in aerobic granular biofilms. *FRONTIERS IN MICROBIOLOGY*, 4.
- Weissbrodt, D. G., Shani, N., and Holliger, C. (2014b). Linking bacterial population dynamics and nutrient removal in the granular sludge biofilm ecosystem engineered for wastewater treatment. *FEMS MICROBIOLOGY ECOLOGY*, 88(3):579–595.
- Wells, G. F., Park, H.-D., Eggleston, B., Francis, C. A., and Criddle, C. S. (2011). Fine-scale bacterial community dynamics and the taxa-time relationship within a full-scale activated sludge bioreactor. *WATER RESEARCH*, 45(17):5476–5488.

## Bibliography

---

- Wey, J. K., Jürgens, K., and Weitere, M. (2012). Seasonal and successional influences on bacterial community composition exceed that of protozoan grazing in river biofilms. *Applied and Environmental Microbiology*, 78(6):2013–2024.
- WHO (2019). Sanitation. <https://www.who.int/en/news-room/fact-sheets/detail/sanitation>. [Online; accessed 17-May-2019].
- Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12).
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wilmes, P., Wexler, M., and Bond, P. L. (2008). Metaproteomics Provides Functional Insight into Activated Sludge Wastewater Treatment. *PLOS ONE*, 3(3).
- Winkler, M.-K. H., Kleerebezem, R., de Bruin, L. M. M., Verheijen, P. J. T., Abbas, B., Habermacher, J., and van Loosdrecht, M. C. M. (2013). Microbial diversity differences within aerobic granular sludge and activated sludge flocs. *APPLIED MICROBIOLOGY AND BIOTECHNOLOGY*, 97(16):7447–7458.
- Winkler, M. K. H., Kleerebezem, R., Khunjar, W. O., de Bruin, B., and van Loosdrecht, M. C. M. (2012). Evaluating the solid retention time of bacteria in flocculent and granular sludge. *WATER RESEARCH*, 46(16):4973–4980.
- Wittebolle, L., Marzorati, M., Clement, L., Balloi, A., Daffonchio, D., Heylen, K., De Vos, P., Verstraete, W., and Boon, N. (2009). Initial community evenness favours functionality under selective stress. *NATURE*, 458(7238):623–626.
- Wittebolle, L., Vervaeren, H., Verstraete, W., and Boon, N. (2008). Quantifying community dynamics of nitrifiers in functionally stable reactors. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 74(1):286–293.
- Wong, M., Mino, T., Seviour, R., Onuki, M., and Liu, W. (2005). In situ identification and characterization of the microbial community structure of full-scale enhanced biological phosphorous removal plants in Japan. *WATER RESEARCH*, 39(13):2901–2914.
- Wooley, J. C., Godzik, A., and Friedberg, I. (2010). A Primer on Metagenomics. *PLOS COMPUTATIONAL BIOLOGY*, 6(2).
- Wrighton, K. C., Thomas, B. C., Sharon, I., Miller, C. S., Castelle, C. J., VerBerkmoes, N. C., Wilkins, M. J., Hettich, R. L., Lipton, M. S., Williams, K. H., Long, P. E., and Banfield, J. F. (2012). Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla. *SCIENCE*, 337(6102):1661–1665.
- Xavier, J. B., De Kreuk, M. K., Picioreanu, C., and Van Loosdrecht, M. C. M. (2007). Multi-scale individual-based model of microbial and bioconversion dynamics in aerobic granular sludge. *ENVIRONMENTAL SCIENCE & TECHNOLOGY*, 41(18):6410–6417.

- Xia, S., Zhou, L., Zhang, Z., and Li, J. (2012). Influence and mechanism of N-(3-oxooxotanoyl)-L-homoserine lactone (C-8-oxo-HSL) on biofilm behaviors at early stage. *JOURNAL OF ENVIRONMENTAL SCIENCES-CHINA*, 24(12):2035–2040.
- Xia, Y., Kong, Y., and Nielsen, P. H. (2007). In situ detection of protein-hydrolyzing microorganisms in activated sludge. *FEMS MICROBIOLOGY ECOLOGY*, 60(1):156–165.
- Xia, Y., Kong, Y., and Nielsen, P. H. (2008a). In situ detection of starch-hydrolyzing microorganisms in activated sludge. *FEMS MICROBIOLOGY ECOLOGY*, 66(2):462–471.
- Xia, Y., Kong, Y., Thomsen, T. R., and Nielsen, P. H. (2008b). Identification and ecophysiological characterization of epiphytic protein-hydrolyzing Saprospiraceae ("Candidatus epiflobacter" spp.) in activated sludge. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 74(7):2229–2238.
- Xia, Y., Wang, X., Wen, X., Ding, K., Zhou, J., Yang, Y., and Zhang, Y. (2014). Overall functional gene diversity of microbial communities in three full-scale activated sludge bioreactors. *APPLIED MICROBIOLOGY AND BIOTECHNOLOGY*, 98(16):7233–7242.
- Yarwood, J., Bartels, D., Volper, E., and Greenberg, E. (2004). Quorum sensing in *Staphylococcus aureus* biofilms. *JOURNAL OF BACTERIOLOGY*, 186(6):1838–1850.
- Yarza, P., Yilmaz, P., Pruesse, E., Gloeckner, F. O., Ludwig, W., Schleifer, K.-H., Whitman, W. B., Euzéby, J., Amann, R., and Rossello-Mora, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *NATURE REVIEWS MICROBIOLOGY*, 12(9):635–645.
- Yoon, J., Kang, S., and Oh, T. (2006). *Dokdonella koreensis* gen. nov., sp. nov., isolated from soil. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*, 56(1):145–150.
- Yu, G. (2018). *treeio: Base Classes and Functions for Phylogenetic Tree Input and Output*. R package version 1.4.1.
- Yu, G., Smith, D., Zhu, H., Guan, Y., and Lam, T. T.-Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8:28–36.
- Zeng, R., Lemaire, R., Yuan, Z., and Keller, J. (2003a). Simultaneous nitrification, denitrification, and phosphorus removal in a lab-scale sequencing batch reactor. *BIOTECHNOLOGY AND BIOENGINEERING*, 84(2):170–178.
- Zeng, R., van Loosdrecht, M., Yuan, Z., and Keller, J. (2003b). Metabolic model for glycogen-accumulating organisms in anaerobic/aerobic activated sludge systems. *BIOTECHNOLOGY AND BIOENGINEERING*, 81(1):92–105.

## Bibliography

---

- Zeng, R., Yuan, Z., and Keller, J. (2003c). Enrichment of denitrifying glycogen-accumulating organisms in anaerobic/anoxic activated sludge system. *BIOTECHNOLOGY AND BIOENGINEERING*, 81(4):397–404.
- Zhang, D., Li, W., Hou, C., Shen, J., Jiang, X., Sun, X., Li, J., Han, W., Wang, L., and Liu, X. (2017a). Aerobic granulation accelerated by biochar for the treatment of refractory wastewater. *CHEMICAL ENGINEERING JOURNAL*, 314:88–97.
- Zhang, H., Sekiguchi, Y., Hanada, S., Hugenholtz, P., Kim, H., Kamagata, Y., and Nakamura, K. (2003). *Gemmatimonas aurantiaca* gen. nov., sp nov., a gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum Gemmatimonadetes phyl. nov. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*, 53(4):1155–1163.
- Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *BIOINFORMATICS*, 30(5):614–620.
- Zhang, L., Li, Q., Chen, C., Li, X., Li, M., Hu, J., and Shen, X. (2017b). *Propioniciclava sinopodophylli* sp nov., isolated from leaves of *Sinopodophyllum hexandrum* (Royle) Ying. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*, 67(10):4111–4115.
- Zhou, D., Niu, S., Xiong, Y., Yang, Y., and Dong, S. (2014). Microbial selection pressure is not a prerequisite for granulation: Dynamic granulation and microbial community study in a complete mixing bioreactor. *BIORESOURCE TECHNOLOGY*, 161:102–108.
- Zilles, J., Peccia, J., Kim, M., Hung, C., and Noguera, D. (2002). Involvement of Rhodocyclus-related organisms in phosphorus removal in full-scale wastewater treatment plants. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 68(6):2763–2769.

## Aline ADLER

alineadler@bluewin.ch

Swiss, from Geneva,  
7<sup>th</sup> of August 1979

### PROFESSIONNEL INTERESTS

---

Microbiology, environment, bioremediation, wastewater treatment, metagenomics, bioinformatic, statistic, programming.

### EDUCATION

---

- From 2014 to 2019** Ph.-D. student at the Laboratory for Environmental Biotechnology (LBE) at EPFL, project on the influence of wastewater composition on the microbial communities in aerobic granular sludge for wastewater treatment.
- From 2011 to 2014** University of Geneva – Master in environmental sciences (MUSE)
- From 2008 to 2011** University of Geneva– Bachelor in biology
- From 1998 to 2003** EPFL – mathematic – Mathematic engineer diploma

### PUBLICATIONS

---

- 2019** Layer M., Adler A., Reynaert E., Hernandez A., Pagni M., Morgenroth E., Holliger C. and Derlon N.  
***Organic substrate diffusibility governs microbial community composition, nutrient removal performance and kinetics of granulation of aerobic granular sludge***, Water research X, [<https://doi.org/10.1016/j.wroa.2019.100033>]
- 2016** Adler A., Devarajan N., Wildi W. and Pote J.  
***Metal distribution and characterization of cultivable lead-resistant bacteria in shooting range soils***, Soil & sediment contamination, p. 378-394.

### EMPLOYMENT EXPERIENCE

---

- From 2009 to 2014** Agricultural-assistant and IT engineer for the company **Directdujardin**
- From 2004 to 2008** Mathematics teacher at the C.O. des Colombières and C.O. Montbrillant  
Teaching aptitude certificate for secondary school
- From 2002 to 2004** Mathematics teacher substitute in Geneva

### LANGUAGES

---

- French**, first language
- English**, spoken and written (level B2, in progress)
- German**, spoken and written (level B1, in progress)

### INFORMATIC

---

Good command of **R**, **Matlab**, **Latex**, **Python**, **Pascal**, **Racket**, **C** and **C++** programming

### HOBBY

---

- From 2007** **Viet Vo Dao** (Vietnamese martial art) four hours a week.

