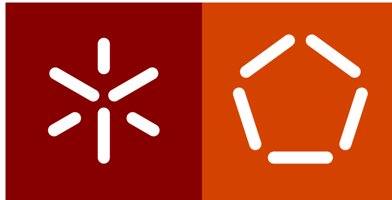


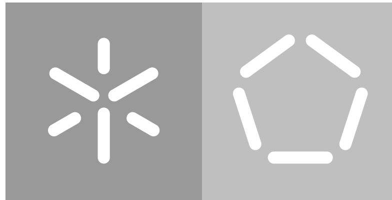
Universidade do Minho
Escola de Engenharia
Departamento de Informática

Bruna Daniela Azevedo da Silva

Metagenomic analysis of a Nitritation/Anammox reactor

Community members and processes

October 2017



Universidade do Minho
Escola de Engenharia
Departamento de Informática

Bruna Daniela Azevedo da Silva

Metagenomic analysis of a Nitritation/Anammox reactor

Community members and processes

Master dissertation
Master Degree in Bioinformatics

Dissertation supervised by
Barth F. Smets
Isabel Rocha

October 2017

ACKNOWLEDGEMENTS

First, I would like to thank my supervisors, Barth Smets and Isabel Rocha for the opportunity, the valuable feedback and all the knowledge that I acquired in this last year. I would like to also thank Alex Palomo for teaching me the required skills and being by my side thought this time, helping me overcome the obstacles I would encounter doing my project.

I would like to thank all the members of the Metlab research group, for the sympathy, help and fun times we had during our breaks. I would like to specially thank Bastian, Carlos, Jan, Alex and Yunjie for the enjoyable lunches, fun conversations, and the help in understanding the biological background of my project.

I would also like to thank the MRM2 conference organization for accepting my work for an oral presentation and letting me know new people from this field of work.

Thanks to all my friends that have made me enjoy this last two years in this Master, and for the fun moments we shared and all the difficulties we overcame. Thank you, Pedro and Daniela, for sharing this journey with me, Z and Fernando for adventuring through Denmark with me and lastly Asbjørn for teaching me Danish language, showing me Denmark and without you this experience would not have been the same.

Finally, I would like to thank my family, for supporting me and fighting for a better future for me. A bid special thanks to my sister Raquel Azevedo, for the advice, guidance and love.

ABSTRACT

Nitrogen is a fundamental element for all organisms. It is nevertheless predominantly found in the atmosphere, in the form of unreactive nitrogen. In the last century, a man-made method for nitrogen fixation improved the crops yield, fuelling a populational growth. The exponential increase of anthropogenic nitrogen in soils and water bodies has, however, affected the environment and deregulated the natural biogeochemical nitrogen cycle. Currently, the costs of repairing the damage caused by the reactive nitrogen load from human activities have overcome the profits of the agricultural improvement, derived from the application of fertilizers. Wastewater treatment plants remove the excessive amounts of nutrients such as carbon, nitrogen and phosphorus from wastewater to prevent environmental impacts derived from excessive nitrogen in the biosphere, like eutrophication.

The current conventional wastewater treatment applied is nitrification coupled with denitrification. However, the requirement for an external carbon source and aeration render this process costly. Furthermore, one of the intermediates of denitrification is nitrous oxide, a greenhouse gas with an effect three hundred times worse than carbon dioxide and with a lifespan of one hundred and twenty years in the ozone layer.

The Partial Nitrification/Anammox (PNA) process combines aerobic ammonium oxidation with anaerobic ammonium oxidation while suppressing the activity of nitrite oxidizing bacteria. This efficient process of nitrogen removal from wastewater reduces the aeration cost and the need for external carbon with zero nitrous oxide emissions.

Unknown microbial interactions may, on the other hand, impair this process, resulting in suboptimal performance such as, excessive nitrate and nitrous oxide emissions. To better understand the microbial community and its interactions and to find the causes of the treatments instability, metabolic analysis and genomic annotation was performed, using two complementary binning methods. The biological samples used in this study were retrieved from a high-rate PNA sequencing batch reactor, fed with carbon-free ammonium-rich synthetic wastewater.

Fifty-seven draft genomes making up about eighty percent of the total community metagenome were recovered. In addition to the three genomes each from Nitrosomonas and Candidatus Brocadia, several genomes belonged to *Proteobacteria*, *Chloroflexi*, *Planctomycetes*, *Bac-*

teroidetes, *Armatimonadetes*, *Ignavibacteriae*, *Acidobacteria*, *Chlorobi*, *Verrucomicrobia*, *Actinobacteria* and *Gemmatimonadetes* phyla.

In this study, the heterotrophic organisms encoding partial denitrification could be divided into niches accordingly to their role in this pathway, describing their interactions as a community. The complexity of the community was also ascertained with the discovery of putative heterotrophic hydroxylamine oxidizing bacteria and putative heterotrophic nitrite oxidizing bacteria.

Overall, high quality genomes that constitute a high fraction of the metagenome were recovered, allowing for a precise description of the PNA reactors community and the flow of nitrogen oxides. A complex community with high redundancy was uncovered basing the main interactions on the partitioning of the nitrogen oxides respiratory pathway.

RESUMO

O azoto é um elemento essencial para todos os organismos. No entanto, este é predominantemente encontrado na atmosfera, sob a forma de azoto não-reativo. No século passado, um método artificial para a fixação de azoto melhorou o rendimento das culturas, alimentando um crescimento populacional. O aumento exponencial do azoto antropogénico nos solos e massas de água afetou o meio ambiente e desregulou o ciclo natural de azoto. Atualmente, os custos de reparação dos danos causados pela inserção excessiva de azoto reativo, fruto das atividades humanas, superam os lucros da melhoria na produção agrícola, derivada da aplicação de fertilizantes. As centrais de tratamento de águas residuais eliminam quantidades excessivas de nutrientes, como o carbono, azoto e fosforo das águas residuais, para prevenir impactos ambientais derivados do excesso de azoto na biosfera, como a eutrofização.

O tratamento convencional de águas residuais aplicado é a nitrificação juntamente com a desnitrificação. No entanto, a necessidade de uma fonte externa de carbono e de aeração tornam esse processo bastante caro. Além disso, um dos intermediários da desnitrificação é protóxido de azoto, ou óxido nitroso, um gás de efeito de estufa com um efeito trezentas vezes pior que o dióxido de carbono, e uma vida útil na camada de ozono de cento e vinte anos.

O processo de Nitrificação parcial /Anammox (PNA) comina a oxidação aeróbica de amónia com a oxidação anaeróbica de amónia enquanto a atividade de bactéria que oxidam nitrito é suprimida. Este eficiente processo de remoção de azoto de águas residuais reduz o custo de aeração, não requer uma fonte de carbono e gera zero emissões de protóxido de azoto.

As interações microbianas desconhecidas podem, por outro lado, prejudicar este processo, resultando num desempenho inferior, como produção excessiva de nitrato e emissão de protóxido de azoto. Para entender melhor a comunidade microbiana, as suas interações e as causas da instabilidade do tratamento, foi realizada uma análise metagenómica seguida de anotação genómica, utilizando dois métodos complementares de *binning*. As amostras biológicas utilizadas como dados neste estudo foram obtidas de um reator de PNA de alta performance, alimentado com águas residuais sintéticas ricas em amónio sem adição de

carbono.

Foram recuperados cinquenta e sete genomas que compõem cerca de oitenta por cento do metagenoma. Além dos três genomas de cada de *Nitrosomonas* e *Candidatus* Brocadia, vários genomas pertenciam aos filos de *Proteobacteria*, *Chloroflexi*, *Planctomycetes*, *Bacteroidetes*, *Armatimonadetes*, *Ignavibacteriae*, *Acidobacteria*, *Chlorobi*, *Verrucomicrobia*, *Actinobacteria* e *Gemmatimonadetes*.

Neste estudo, os organismos heterotróficos que codificam a desnitrificação parcial podem ser divididos em grupos de acordo com o seu papel nesta via metabólica, descrevendo as suas interações na comunidade. A complexidade da comunidade também foi verificada com a descoberta das bactérias heterotróficas putativas que oxidam hidroxilamina e as bactérias heterotróficas putativas que oxidam nitrito.

Em geral, foram recuperados genomas de alta qualidade que constituem uma grande fração do metagenoma, permitindo uma descrição precisa da comunidade do reator e o fluxo de óxidos de azoto. Uma comunidade complexa com enorme redundância e com as suas principais interações baseadas na partição da via respiratória de óxidos de azoto.

CONTENTS

1 INTRODUCTION 2
2 STATE OF THE ART 5
2.1 Nitrogen Cycle and Wastewater Treatment Plants 5
2.1.1 Nitrogen cycle 5
2.1.2 New discoveries to the nitrogen cycle 8
2.2 Impact of human activity in the biogeochemical nitrogen cycle 10
2.2.1 Consequences of nitrogen cycle unbalance 10
2.2.2 Biological wastewater treatments 12
2.3 Metagenomics 16
2.3.1 History of microbial community studies 17
2.4 Marker gene amplification metagenomics 19
2.5 Shotgun metagenomics 19
2.6 Metagenomics Procedure Overview: 20
2.6.1 Pre-processing 20
2.6.2 Assembly 21
2.6.3 Binning 22
2.6.4 Quality Assessment 24
2.6.5 Taxonomical annotation 25
2.6.6 Genome annotation 25
3 CASE STUDIES 26
4 CONCLUSION 70
4.1 Conclusions 70
4.2 Prospect for future work 71

A APPENDIX I 87

LIST OF FIGURES

Figure 1 Major processes of the nitrogen cycle. 6
Figure 2 Major processes of the nitrogen cycle, including the new discoveries highlighted in red. 9
Figure 3 A graphical representation of the variance of abundance throughout the time, in total six month, of the genomes displaying most fluctuation in abundance. 87
Figure 4 A graphical representation of the genomes, organized by their variation of abundance throughout the samples. From genomes with decrease in abundance (left) to the most abundant(middle), ending with the MAGs with increase in abundance(right) Eren et al. (2015). 88
Figure 5 A graphical representation of the decrease of certain organisms in abundance, while the two predators, highlighted by the blue rectangle, display growth. 89

LIST OF TABLES

Table 1 Comparison of the most used metagenomic assemblers, adaptated from Gurevich et al. (2013); Vollmers et al. (2017) 23

ACRONYMS

A

ANAMMOX Anaerobic Ammonium Oxidation. 3, 8, 15, 16

AOB Ammonia Oxidizing Bacteria. 3, 6, 7, 9, 15

B

BOD Biochemical Oxygen Demand. 12, 13

C

CG Composite Genome. 23

COD Chemical Oxygen Demand. 3, 14

COMAMMOX Complete Ammonia Oxidiser. 9

D

DGGE Denaturing Gradient Gel Electrophoresis. 17

DNRA Dissimilatory Nitrate Reduction to Ammonium. 5, 8

DO Dissolved Oxygen. 14, 15

E

EU European Union. 10

H

HB Haber-Bosch. 10

HTS High Throughput Screening. 17

M

MAGS Metagenome-Assembled Genomes. 18, 24, 70

N

NGS Next Generation Sequencing. 17

NOB Nitrite Oxidizing Bacteria. 6, 7, 9, 14, 15, 70

O

OD Oxidation Ditch. 13

OTU Operational Taxonomic Unit. 18

P

PCR Polymerase Chain Reaction. 17, 18

PHMM profile Hidden Markov Models. 22

PNA Partial-Nitrification/Anammox. 3, 4, 15, 16, 70, 71

S

SBR Sequencing Batch Reactor. 13

SHARON High Activity Ammonium Removal Over. 15

SNAD Simultaneous nitritation, anammox and denitrification. 16

SND Simultaneous nitrification and denitrification. 14, 15

T

T-RFLP Terminal Restriction-Fragment Length Polymorphism. 17

W

WWTPS Wastewater Treatment Plants. 3, 4, 11

1

INTRODUCTION

Freshwater is an irreplaceable natural resource essential for human life and ecosystem quality Pfister et al. (2009). Therefore, with the increase of the population, freshwater distribution is becoming scarce, leading to human disease, international conflicts, hindering of agriculture and industrial production and harsh degradation of ecosystems Postel, S.L.; Daily, G.C.; Ehrlich (2008). Besides with the shortening of the freshwater supply, the organic waste produced by anthropogenic activities compromise the water quality Foley (2005).

There are two types of water pollutants, conservative and nonconservative, which differ in their interactions with other natural components in the water stream. Although conservative pollutants pose some danger to the environment due to its toxicity, nonconservative have a more adverse effect in these ecosystems. These highly unstable pollutants are organic molecules that lixiviate from waste. Later, these molecules are reduced or oxidized by bacteria into inorganic materials, such as bicarbonates, nitrates, sulphates and phosphates. Kneese (2015).

Nonconservative pollutants are nutrients such as nitrogen and phosphorus, that enters freshwater from anthropogenic wastewater discharges Kivaisi (2001). Nutrient pollution surpasses other environmental pollutants such as, chemicals, heavy metals or oil discharges according to the level of devastation to the habitat Schindler and Vallentyne (2008). The disruptive increase of nutrients modifies the ecosystems homeostasis, altering the balance of the food chain and leading to the proliferation of specific organisms, named producers M. Nasir Khan (2014).

Wastewater treatment is a method of removing pollutants and excessive nutrients in water before it is disposed into receiving waters, preventing environmental impacts Smith (2003). Nutrients like nitrogen, phosphorus and silicon are removed to balance the nutrients stoichiometry in water outflow Rabalais (2002).

Initially thought as the only element involved in this process, phosphorus an essential ingredient for the eutrophication of water bodies and can trigger it at very low concentrations.

Therefore, several countries only medium of controlling it, is by reducing its concentration below 10μg/L. However, recent studies indicate nitrogen as the limiting factor for this process, so its removal is often preferred Yin and Kong (2014b).

Several methods are implemented in wastewater treatment for nitrogen removal, such as air stripping, chemical treatment, biological nitrification-denitrification, selective ion exchange and partial-nitritation-anammox procedure Yin and Kong (2014b).

Since 1960, in *Wastewater Treatment Plants* (*WWTPs*), Nitrification-Denitrification became defined as the procedure to biologically remove nitrogen (Isaacs and Henze, 1995). This is a two-step process, in which ammonia is oxidized to nitrate(Nitrification) and later reduced to nitrogen gas (Denitrification) Yin and Kong (2014a). Biological denitrification is performed by heterotrophic bacteria, organisms that require organic carbon as energy source, in anoxic conditions Windey et al. (2005). Therefore, carbon availability is a limiting factor in nitrogen removal, as it requires on average 5-10 *COD*/g N (*Chemical Oxygen Demand*) to achieve complete denitrification Isaacs and Henze (1995). In wastewater, the amount of organic carbon accessible is insufficient, requiring the addition of external carbon, making it costly zhen PENG et al. (2007).

Besides, during denitrification, nitrous oxide, a greenhouse gas with 300 times more ozone depleting potential than carbon dioxide is produced Tsuneda et al. (2005). In *WWTPs*, nitrous oxide is emitted, accounting for 26% of total emissions in water bodies, that corresponds to 1.3% of global release Kampschreur et al. (2009). Overall, nitrification-denitrification is economically at a disadvantage, releases an impactful greenhouse gas and denitrification efficiency depends on the ratio of (C/N) Sobieszuk and Szewczyk (2006); Zhao et al. (2017).

Partial-nitritation combined with *Anaerobic Ammonium Oxidation* (*Anammox*) is a recent biotechnological process for nitrogen removal in wastewater. *Anammox* bacteria oxidize ammonia and nitrite into nitrogen gas, in a 3-step pathway, without the emission of nitrous oxide Fux et al. (2002). Besides, these bacteria are autotrophic, not requiring external carbon source, reducing the cost of its application in *WWTPs* Strous et al. (1998). Since nitrite is a basic requirement in this pathway, *Anammox* bacteria are coupled with *Ammonia Oxidizing Bacteria* (*AOB*), who perform partial nitritation, supplying nitrite Hu et al. (2013).

Throughout a period of six month, seven samples were collected in a one-month interval from a high-rate *Partial-Nitrification/Anammox* (*PNA*) sequencing batch reactor, and were subjected to shotgun sequencing. The reactor ran through a period of three years with regular aeration periods and minimal nitrous oxide emissions, while being fed with

ammonium-rich synthetic wastewater without added carbon source Domingo-Félez et al. (2014).

Although several studies have investigated the microbial communities from several biological nitrogen removal processes in *WWTPs*, little is known about the organisms and their interactions Speth et al. (2016); Bhattacharjee et al. (2017). The aim of this work is to identify the organisms present in *PNA* reactors, clarifying the role of each one in this microbial community in the prospects of understanding their organization and exchanges. Therefore, metagenomic analysis, genome reconstruction and functional annotation of the microbial community were performed, elucidating on the complexity of this community. Understanding and optimization of this process can lead to a great reduction in operational energy costs in *WWTPs* and a decrease in the nitrous oxide emissions.

The thesis is divided in the state-of-the-art, case studies and conclusion/future perspectives. In the state of the art, a review is contrived to familiarize the reader to the topic. This chapter is divided into two parts, the nitrogen cycle and *WWTPs* and metagenomics overview.

In the case study, the results obtained throughout this master thesis are presented and discussed in a paper format. The manuscript here addressed is still in preparation for future publication in a peer-reviewed journal.

Lastly, the conclusion and future perspectives conclude the work here introduced and the new roads this study can go through.

2

STATE OF THE ART

2.1 NITROGEN CYCLE AND WASTEWATER TREATMENT PLANTS

2.1.1 *Nitrogen cycle*

Nitrogen is an essential element in life, being the building block of several macromolecules, such as proteins, heterocyclic N-compounds, RNA and DNA and energy transfer molecules, with organisms containing on average six percent of it in their dry mass Bothe et al. (2007). In nature, nitrogen can be found in two diverse ways, depending on its reactiveness Socolow (1999). Unreactive nitrogen, or dinitrogen gas, the main source of nitrogen represents approximately 79% of the atmosphere composition Sprent (1987). As most organisms require reactive forms of nitrogen, nitrogen fixation is crucial to prevent N-starvation Borges et al. (2008). After fixation, nitrogen becomes reactive and can be bound to carbon, producing organic nitrogen, or forms small molecules or ions with oxygen and hydrogen as nutrient nitrogen, or inorganic nitrogen Socolow (1999).

The biogeochemical nitrogen cycle is made of several and different processes that vary the nitrogen oxidation state from -3 (in ammonia) to +5 (in nitrate) Bothe et al. (2007). These pathways are Fixation, Nitrification, Denitrification, Assimilation, Mineralization, and *Dissimilatory Nitrate Reduction to Ammonium* (*DNRA*) (Figure 1).

Nitrogen Fixation

Nitrogen is required by all organisms but the majority of it is present in the atmosphere in the form of dinitrogen gas. This gas appears in a diatomic triply bonded form, assuring that the molecule is stable, thermodynamically and kinetically, making it necessary to dispend enormous amounts of energy to break it Moir (2011). Since dinitrogen gas is unreactive, a process like nitrogen fixation is essential to generate reactive nitrogen so organisms can uptake it and grow Fowler et al. (2013).

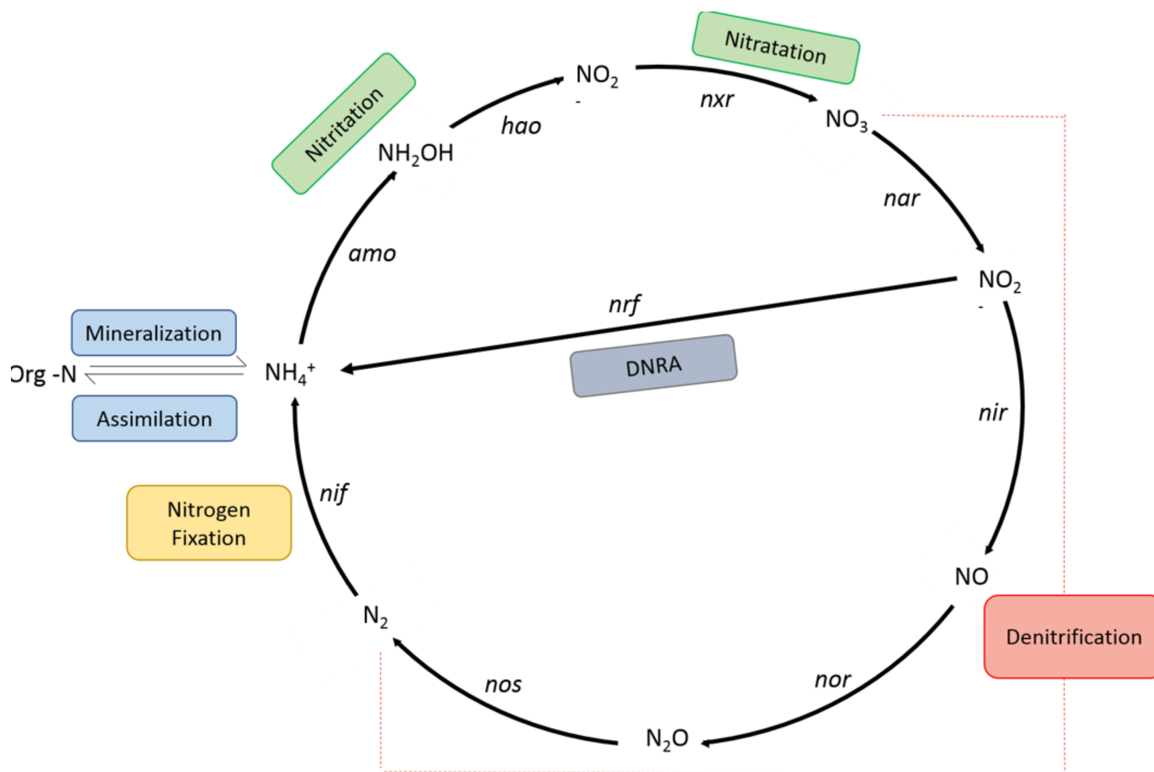


Figure 1.: Major processes of the nitrogen cycle.

Biologically, other than lightning Fields (2004), nitrogen fixation in soils and waters is performed by a restrict group of nitrogen-fixing bacteria, the diazotrophs Vicente and Dean (2017). Diazotrophs are free-living and symbiotic bacteria and archaea that utilize nitrogenase, the enzyme that catalyses nitrogen fixation, so it can break the nitrogen-nitrogen triple bond. In this process, at normal temperature and pressure , dinitrogen suffers reduction and protonation, to produce ammonia and hydrogen Hoffman et al. (2013). Since nitrogenase enzyme is sensible to oxygen this process mostly occurs in anaerobic conditions Vicente and Dean (2017).

The generated ammonia is either oxidized or assimilated.

Nitrification

Nitrification is a two-step process that oxidizes nitrate from ammonia via nitrite in the presence of oxygen . First, ammonia oxidation to nitrite is carried by *Ammonia Oxidizing Bacteria* (*AOB*), with the production of one intermediate compound, hydroxylamine. Two key enzymes are involved, ammonia monooxygenase (AMO) and hydroxylamine oxidoreductase (HAO). On the second step, *Nitrite Oxidizing Bacteria* (*NOB*) utilize nitrite oxidoreductase

(NXR) to oxidize nitrite to nitrate Ward et al. (2011).

AOB and *NOB* are chemolithotrophic bacterial organisms that fixate carbon dioxide and use ammonia and nitrite as an energy source, respectively Kampschreur et al. (2009). Also known as nitrifying bacteria, *AOB* belong to the *Proteobacteria* phylum, specifically to Beta- (*Nitrosomonas* and *Nitrosospira*) and Gamma- (*Nitrosococcus*) classes, while *NOB* phylogenetically fits with *Proteobacteria* (*Citrobacter*, *Nitrospina* and *Nitrococcus*) and *Nitrospirae* (*Nitrospira*) phylum Gerardi (2003).

This process produces oxidized forms of nitrogen, that later will be consumed by other organisms as respiratory substrates Ward and Jensen (2014).

Denitrification

Denitrification is a four-step process performed by facultative anaerobic bacteria in oxygen exhausted environments. Nitrate is reduced into several nitrogen oxides, nitrite, nitric oxide, nitrous oxide with dinitrogen gas as the final product, by nitrate reductase (NAR), nitrite reductase (NIR), nitric oxide reductase (NOR) and nitrous oxide synthase (NOS) enzymes respectively Revsbech and Sørensen (2013). This is a respiratory process in which denitrifiers utilize nitrate as electron acceptor, instead of oxygen, to oxidize organic matter Galloway (2005). Therefore, these heterotrophic bacteria, require organic matter as carbon source, necessitating considerable amounts of carbon to perform complete denitrification Sobieszuk and Szewczyk (2006). In cases of low of organic matter availability, incomplete denitrification will happen, altering the final product to nitric oxide or nitrous oxide Galloway (2005).

There are a diverse group of organisms that is involved in this pathway that are represented in several phyla, such as *Proteobacteria*, *Chloroflexi*, *Bacteroidetes*, *Armatimonadetes*, *Ignavibacteriae*, *Acidobacteria*, *Chlorobi*, *Verrucomicrobia*, *Omnitophica*, *Actinobacteria* and *Gemmatimonadetes* Speth et al. (2016)(this work). Although denitrification helps deplete the pool of reactive nitrogen in the environment, by decreasing the fertility levels, it completes the nitrogen cycle, maintain the balance Socolow (1999).

Assimilation and Mineralization

Nitrogen assimilation is the process which bounds carbon to nitrogen, forming organic nitrogen molecules indispensable to build macromolecules. Plants, fungi and some bacteria intake ammonium and nitrate in its ion form, while animals obtain the necessary nitrogen

compounds through digestion Masclaux-Daubresse et al. (2010); Rufino et al. (2016).

Mineralization then corresponds to the conversion of organic matter into inorganic compounds. This process happens during decomposition of organic matter and increase the bioavailability of nutrients to the environment Socolow (1999).

Dissimilatory Nitrate Reduction to Ammonium

DNRA or nitrate/nitrite ammonification is a competing process with denitrification, recycling unused nitrate into ammonia. This is an anaerobic process, and like denitrification, nitrate is used as electron acceptor to oxidize organic matter. The key enzyme in cytochrome c nitrite reductase, reducing nitrite back to ammonia Lam and Kuypers (2011).

2.1.2 *New discoveries to the nitrogen cycle*

The better understanding of the global nitrogen cycle, as well as the new discoveries, has been possible with the new methods of analysis available in the last few decades Fowler et al. (2013). For exemple, two microorganisms related to the nitrogen cycle were proposed in 1977, by Broda. Both organisms were autotrophs, with one synthetizing dinitrogen by ammonium oxidation with nitrite or nitrate as electron acceptor, and the other by utilizing inorganic nitrogen as electron acceptor for carbon dioxide fixation, in the presence of light Broda (1977). *Anaerobic Ammonium Oxidation* (*Anammox*) was discovered two decades after Brodas prevision Mulder et al. (1995).

Anammox bacteria belongs to the Planctomycetes phylum, and forms an unique clade of six genera, and a total of nineteen species, distributed in diverse aquatic environments, where they are responsible for 50% of the nitrogen turnover in water systems Ali and Okabe (2015); Schubert et al. (2006). The *Anammox* pathway is a three-step process, in which ammonium and nitrite are converted to dinitrogen with nitric oxide and hydrazine as intermediates Strous et al. (2006); Kartal et al. (2013). One unique feature of these organisms is an organelle designed anammoxosome Kartal et al. (2010). The anammoxosome is fully closed to contain hydrazine, an intermediate of the anammox pathway, extremely reactive and toxic Dietl et al. (2015).

One decade after, a similar photosynthetic organism similar to Brodas prediction performing anoxygenic photosynthesis with nitrite as an electron acceptor Broda (1977). This phototrophic carbon fixation pathway oxidizes nitrite to nitrate in the absence of oxygen.

This is the first oxidation process of nitrogen cycle that is photosynthetically driven that is reported Griffin et al. (2007).

Moreover, a mechanism of nitric oxide reduction to nitrogen with the production of oxygen was found in anaerobic methane oxidizing bacteria, Candidatus Methylomirabilis oxyfera Ettwig et al. (2010). Nitric oxide dismutase (NOD) belongs to the quinol-dependent nitric oxide reductase (qNOR) family. The formed oxygen is then available for microorganisms to aerobically catabolize pathways when in anoxic habitats Sanford et al. (2012).

For more than a century the nitrogen cycle processes have been known, but modifications to well-studied reactions have arisen, showing the need to adjust the nitrogen cycle Galloway et al. (2013). Nitrification is a two-step process carried out by two groups of organisms, *AOB* and *NOB*. Recently, a new clade of Nitrospira species codifying the complete ammonium oxidation to nitrate was discovered, and appointed as *Complete Ammonia Oxidiser* (*Comammox*) bacteria van Kessel et al. (2015); Daims et al. (2015); Palomo et al. (2016).

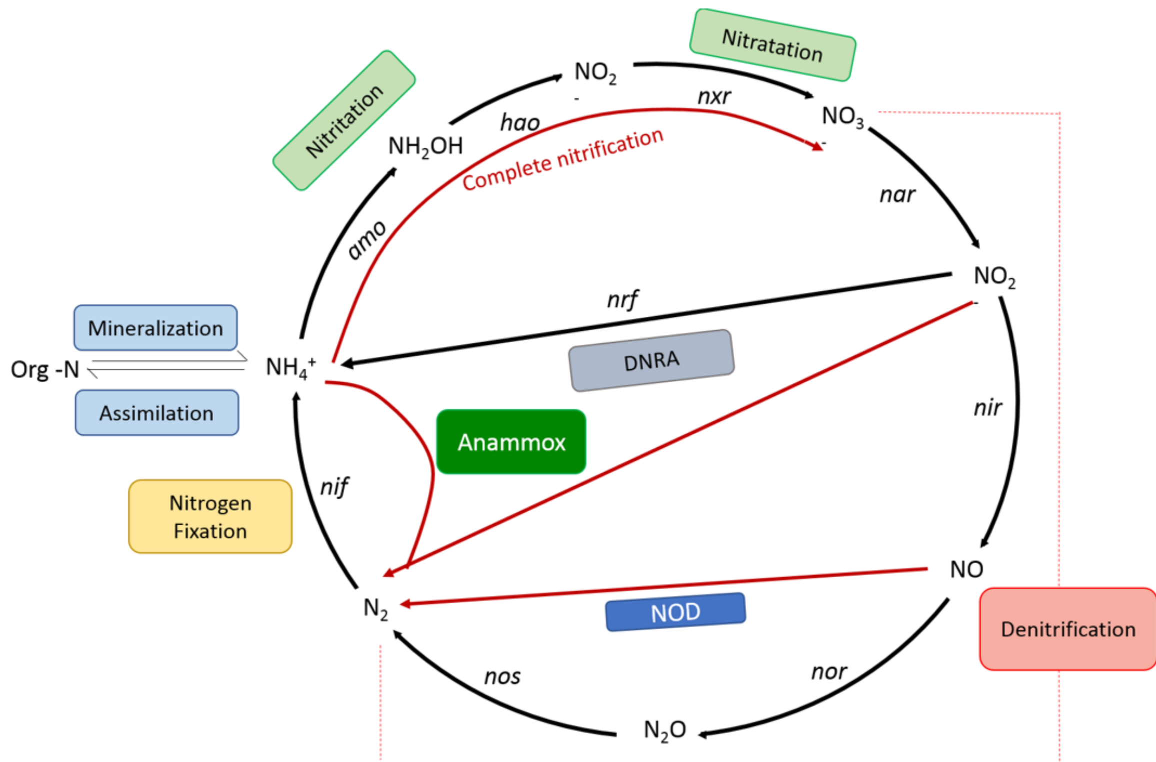


Figure 2.: Major processes of the nitrogen cycle, including the new discoveries highlighted in red.

2.2 IMPACT OF HUMAN ACTIVITY IN THE BIOGEOCHEMICAL NITROGEN CYCLE

The Reactive nitrogen bioavailability in natural environments is limited by biological nitrogen fixation, lightning and anthropogenic fixation through biomass combustion for energy supply Fowler et al. (2013). Therefore, the increase of reactive nitrogen availability has been exploited for application as fertilizer in food production, increasing its yield Fowler et al. (2013). In the last century, Fritz Haber and Carl Bosch have created a process to artificially fixate nitrogen, through the combination of hydrogen and nitrogen gas at high temperatures and pressures, with the help of a catalyst, producing ammonia Fowler et al. (2013); by Juan Ilerbaig (2002). Nowadays, the *Haber-Bosch* (*HB*) process is mainly used in fertilizers manufacturing, revolutionizing the food industry and supporting the exponential growth in human population Vicente and Dean (2017). In merely a century, the application of anthropogenic fixed nitrogen allowed the world population to grow form 1.6 billion habitants, in 1900, to approximately 6 billion Smil (1999).

With the integration of *HB* nitrogen into the global nitrogen cycle, in the beginning of this decade, its production (120 Tg N yr-1) doubled the natural amount biologically fixed (63 Tg N yr-1), disrupting the overall cycle Fowler et al. (2013). Human input into the cycle brought an increment in food production, specially meat and milk, but it is arguable that the negative impacts have surpassed the benefits Bouwman et al. (2013).

2.2.1 *Consequences of nitrogen cycle unbalance*

To the *European Union* (*EU*) the profits derived from application of fertilizers to crops and animal feeding are less than half the investment necessary to remove excess nitrogen from the environment. Since nitrogen as an element cascades through several chemical forms, 70 billion to 320 billion euros are required to alleviate the consequences Sutton et al. (2011). Additionally, the discovery and combustion of fossil fuel lead to an increase of the amount of dinitrogen gas in the atmosphere in about 10% Socolow (1999). Of all the biological cycles, the nitrogen cycle suffered the greatest impact / modification, implicating the quality of water bodies, air and soils, human health, biodiversity and climate Suddick et al. (2013).

Besides the environmental repercussions, some economic and politic disputes have arisen, such as cost and legislation of fertilizers distribution and application, socio-political issues related to uncontrolled population growth and international conflict derived from lack of freshwater Vicente and Dean (2017); Postel, S.L.; Daily, G.C.; Ehrlich (2008).

Atmosphere

Three of the most anthropogenic greenhouse gases, nitrous oxide, carbon dioxide and methane, are influenced by the nitrogen cycle. Nitrogen is involved in the sequestration of carbon in soils and forests, while ozone and nitrogen oxides affect the concentration of methane in the atmosphere. Therefore, mitigation of the nitrogen cycling throughout the next century should alleviate climate change Suddick et al. (2013).

Nitrous oxide, an intermediate of denitrification, is responsible for the regulation of the stratospheric ozone Galloway (2005). This gas is unreactive in the atmosphere, until it reacheds the stratosphere, there its photolyzed into nitric oxide Portmann et al. (2012). Then, the nitric oxide interacts with ozone creating new reactive oxygen and nitrogen species, breaking down ozone into oxygen gas Malik et al. (2016). Unfortunately, with the deregulation of the nitrogen cycle, the rate of nitrous oxide is increasing like carbon dioxide. However, nitrous oxide ozone potential as an ozone depleting gas is 300 times higher than carbon dioxide, with a life expectancy in stratosphere of 120 years Kampschreur et al. (2009); Tsuneda et al. (2005).

Water and soils

Of the Earths water volume, 0.77% is freshwater and available for life Postel, S.L.; Daily, G.C.; Ehrlich (2008). The hydrological cycle, like nitrogen cycle, has been anthropologically altered, due to excessive nutrients from fertilizer run-off, compromising water quality Foley (2005). Agricultural continuous land use with application of aeration techniques for better water and nutrient permeation destroys the anoxic environments required for denitrification. This leads to the accumulation of nitrate in soils, increasing the possibility of nutrient run-off or penetration to groundwater Socolow (1999).

Nutrient run-off through precipitation, and its consequentially accumulation in water bodies, leads to eutrophication. Thise disruptive increase of nutrients, changes the ecosystems biological production and nutrient cycles M. Nasir Khan (2014). This variation in nutrient levels leads to a proliferation of organisms, the producers. Producers that occupy this trophic level of the food chain, other than inorganic nutrients, also need sunlight, that limit their area of occupation. Therefore, the growth on the surface of the water body results in a wall of biomass Smith (2003). The proliferation of biomass blocks sunlight penetration in the water body, leading to the death of vegetal organisms that reside in the bottom. When these organisms die, the bacterial matter conversion from organic to inorganic drains the oxygen. Oxygen exhaustion, also known as hypoxia, suffocate of higher life forms and

leads to infertility of the water system Rabalais (2002).

With eutrophication creating an anoxic zone in water bodies and nutrient enrichment promoting the spreading and growth of fast-growing, non-native species, biodiversity is starting to be lost Suddick et al. (2013). Thus, nitrates, nitrites and ammonium are removed in drinking water and *WWTPs* using costly treatments. In Europe, 80% of freshwater bodies has a higher risk level of nitrogen, above 1.5 mg/L. The detection of nitrate in drinking water poses serious risks to human health, because the presence of nitrate in the blood stream inactivates the hemoglobin Socolow (1999). Climate changes affects the consistency of precipitation, varying the nitrogen accessibility through nutrients run-off during storms, causing an upsurge of algal blooms Suddick et al. (2013).

2.2.2 *Biological wastewater treatments*

Freshwater is essential for all living organisms, but human activity has degraded water quality, posing a serious risk to health and ecosystem quality. Population growth increases the demand for freshwater, whose resources have been decreased by pollution. Until 2003, around 2 million tons of anthropogenic and agricultural waste is was discharged into receiving waters per day UN WWAP (2003). In developing countries, the lack of freshwater is linked to the shortage of wastewater treatment facilities. Consequently, wastewater treatment becomes decisive to restock freshwater supply and prevent further pollution of water bodies Pfister et al. (2009).

Wastewater can originate from domestic households, agricultural or industrial discharges. Domestic wastewater composition does not fluctuate from the normal nutrients, like carbon, nitrogen and phosphorus, detergents and some compounds present in everyday products Ramalho (2012). Sewage water prevenient of agricultural, industrial or even domestic waters mixed with storm water discharges, besides the expected elevated percentage of nutrients, are also polluted with siltation, oxygen-depleting substances, metals, suspended solids, pesticides and organic toxic chemicals Burton and Pitt (2002).

The wastewater treatment process depends mainly on regulations applied to effluents quality, but overall, they are divided into pre-treatment, followed by a sludge producing primary and secondary treatment, and finished with the tertiary treatment Pfafflin and Ziegler (2006).

Pre-treatment or preliminary, sometimes combined with the primary treatment, screens out and removes the solids and the dense materials, like: papers, rags, plastics, sands and silts, that are capable of damaging or reducing the efficiency of the process Parr et al. (1999). It is also possible to do some Additional maintenance that consist of odour control, septage handling and flow equalization U.S. EPA (2000).

In primary treatment, wastewater suffers primary sedimentation, a layered physical treatment for the removal of light organic matter with dimensions above 50 μm. Although the methods is not sufficient to remove enough pollution in order to meet the necessary quality standards, primary treatment removes 50% of the suspended solids and up to 50% of the *Biochemical Oxygen Demand* (*BOD*). Nutrients like nitrogen and phosphorus are lost on average of 10%, as well as some hydrophobic constituents and metals that are eliminated. Improvements to this process can be the inclusion of coagulation or flocculation upstream of the sedimentation tank or the application of a filtration system downstream Asano (1998).

The biological wastewater treatment process, or secondary treatment, uses an array of processes coupled with secondary sedimentation for nutrients and *BOD* removal Asano (1998). Microorganism degrade the organic matter, organic phosphorus and nitrogen molecules and dissolve minerals, in the presence of oxygen. Microorganisms in suspended growth or fixed-film, like biofilms, are mixed with wastewater, with a continuous aeration for biological oxidation Parr et al. (1999). The degradable organic matter is responsible for the supply of nutrients and energy to microbial growth, this will produce carbon dioxide, water and other end products Asano (1998).

Removal of organic ions can be performed using biological agents or chemicals. the Biological removal of ammonia, nitrates and phosphates has a better performance than the chemical processes, although the cost is greatly increased Abdel-Raouf et al. (2012). In this process and in order to further improve water quality, removal of heavy metals and other additional treatments can be employed Asano (1998).

The disinfection step is one the final procedures in wastewater treatment that usesd oxidizing chemical, ozone, bromide, chlorine and hydrogen peroxide, to remove colour, odour and to improve the degradability of organic matter. In order toTo remove excessive microorganisms or pathogens, wastewater is then submitted to alkaline environments or granular/membrane filtration systems Asano (1998).

Advanced treatments are specialized treatments based on complex technology, like chemical precipitation, ozonation, reverse osmosis or carbon absorption. To remove compounds

such as inorganic nitrogen and phosphorus to legislation acceptable levels, it requires advanced treatments. These treatments are based on natural occurring aerobic and anaerobic processes in the environment Abdel-Raouf et al. (2012). For biological nitrogen removal, in the 1960s, nitrification/denitrification over nitrate was established Isaacs and Henze (1995).

Standard Processes of Biological Removal of Nitrogen

Throughout the years, several processes based on nitrogen cycles nitrification and denitrification have been applied to wastewater treatment plants as a way to improve nitrogen removal. Such processes are pre-denitrification (Anoxic/Oxic), modified Bardenpho, Bio-denitro, *Sequencing Batch Reactor* (*SBR*), *Oxidation Ditch* (*OD*), step feeding and anaerobic/anoxic/aerobic (A2/O) Whitacre (2013).

Conventional treatment for nitrogen removal uses autotrophic nitrification to oxidize ammonia to nitrate, via nitrite, and heterotrophic denitrification, reducing nitrate to dinitrogen Isaacs and Henze (1995). Denitrification requires a considerable amounts of carbon sources, but if the *COD*:N ration in wastewater is insufficient, or if further degradation is impossible, it is necessary an external supply of carbon (Sobieszuk and Szewczyk 2006). This process can be applied to ammonium-rich influents, if sufficient oxygen and carbon is supplied. Conventional nitrification/denitrification via nitrate is carried by different organisms and environments in two different spaces, increasing the costs, retention time and volume to obtain a satisfactory removal efficiency. Moreover, efficiency is also dependent of the high oxygen and carbon requirements to achieve complete nitrification and denitrification Whitacre (2013); Volcke et al. (2006).

However, water quality standards for wastewater treatment plants effluents have become stricter, of less than 10mg of total nitrogen per litre, requiring new methods to comply with these values Whitacre (2013).

Simultaneous nitrification and denitrification via nitrate

Simultaneous nitrification and denitrification (*SND*) via nitrate occurs when nitrification and denitrification are simultaneous in the same reactor. Contrary to the conventional process, both autotrophic nitrifiers and denitrifiers physically rearrange themselves in granular flocs or biofilms. Nitrifiers occupy the aerobic regions of the flocs, where the *Dissolved Oxygen* (*DO*) concentration is between 1 and 2 mg/L, and denitrifiers stay alive in the anoxic zone, with *DO* less than 0.5mg/L Hibiya et al. (2003). Biological *SND* via nitrate does not involve the traditional autotrophic nitrification and heterotrophic denitrification, and instead nitri-

fication is carried by heterotrophic nitrifiers while denitrification happens in the presence of oxygen. With the process taking place in the same area, construction and maintenance costs are reduced, simplifying the design. But, this process is limited by wastewater with low C:N ratio, less than 5, or the amount of external carbon will overshadow the benefits Whitacre (2013).

Simultaneous nitrification and denitrification via nitrite

SND via nitrite, or shortcut nitrification and denitrification, inhibits the growth of *NOB*, resulting in a shortcut. In this process, nitrite is the end product of nitrification, and denitrification starts with the reduction of nitrite instead of nitrate. Therefore, there is a reduction of 25% of the oxygen demand, 40% of the external carbon supply and 40% reduction of the produced biomass Peng and Zhu (2006). Suppression of *NOB* growth is obtained from high concentrations of free ammonia, temperature, lag-time between aerobic and anaerobic stages, low *DO* concentration during aeration or free-hydroxylamine concentration. High ammonia concentration accompanied by nitrous acid inhibit both *AOB* and *NOB*, with *NOB* requiring minor concentrations for inhibition. Normally, a shortcut of the nitrification is carried while creating a *DO* ration with free ammonia inferior to 5, by supplying high ammonia wastewater and low concentration of *DO* and maintaining the temperature around 25C Yoo et al. (1999).

Single reactor system for *High Activity Ammonium Removal Over* (*SHARON*) is operated at elevated temperatures, between 30 to 40C, and neutral pH. These conditions promote the inhibition of *NOB* resulting in their washout when setting a retention time of one day Volcke et al. (2006).

Either *SND* via nitrite and *SHARON*-denitrification are mainly applied to treatment of high ammonium wastewater at high temperatures. However it cannot be applied, when its required strict effluent standards or pre-heating of the receiving waters Whitacre (2013); Volcke et al. (2006).

Partial Nitritaton combined with Anammox

Complete autotrophic nitrogen removal, combines *AOB* and *Anammox* bacteria to fully convert ammonia and nitrite into dinitrogen gas without the production of nitrous oxide. In *PNA*, *AOB* are inhibited by low pH resulting from the oxidation of half of the reactor ammonia Schmidt et al. (2003). After, *Anammox* bacteria utilize the resulting nitrite to oxidize the remaining ammonia. Application of this process in a full-scale wastewater treatment

plants needs high amounts of biomass, warm temperatures (25C) and is limited by both organisms low growth rate Whitacre (2013). Therefore, in biofilms or granular flocs, thick biofilms organization provide anaerobic conditions for *Anammox* bacteria, while the *AOB* reside in the outer layer of these, or small flocks. The coexistence of both organisms in biofilms enhances the process performance, the suppression of the *NOB* bacteria and provides resistance to temperature changes. Application of this process is ideal for ammonia-rich, low-carbon receiving waters, and can decrease cost from aeration and external carbon supply and treatment of sludge production Laureni et al. (2016). Although *PNA* effluent waters contain nitrate, a product of anammox metabolism, and nitrogen removal efficiency is below standards, the benefits still make up for these undesirable consequences, making it viable Whitacre (2013). The first full-scale anammox reactor, for treatment of reject water, was set up in Rotterdam, Netherlands, in 2002. Since then, 114 new installation were reported, for the treatment of reject waters Ali and Okabe (2015).

Simultaneous nitritation, anammox and denitrification

Simultaneous nitritation, anammox and denitrification (*SNAD*) process derives from the *PNA* system, with the inclusion of heterotrophic denitrifiers, to improve effluent quality and removal of high concentration of organic matter. In mainstream wastewater treatment, the receiving waters have high C/N ratio, that promotes the growth of heterotrophic bacteria, so *PNA* cannot be applied, since heterotrophs outcompete *Anammox* bacteria. However, the addition of heterotrophic bacteria to *PNA*, when the amount of organic matter is at low concentrations, improves the overall efficiency Ali and Okabe (2015). Moreover, the coexistence of denitrifiers and *Anammox* makes the system shock and toxic loading resistant, preventing inhibition Langone et al. (2014). The production of nitrate by *Anammox* bacteria, around 11% of the converted nitrogen, and the coupling of this organisms with denitrifiers constitutes an innovative approach for complete nitrogen removal, cutting the excessive amounts of nitrate in the effluent of *PNA* systems. Overall, this process consumes 63% less oxygen, requires smaller amounts of biodegradable carbon molecules to nitrogen ratio, emits less quantities of greenhouse gases and generates less waste sludge Li and Tao (2017). Some difficulties associated with *SNAD* lies in the simultaneous maintenance of these three groups of organisms with so much variables in one treatment system Shannon (2016).

2.3 METAGENOMICS

Microorganisms are present in every environment on Earth, playing a fundamental role in the ecosystem and supporting most of the processes of the biosphere. They are responsible

for maintaining the biogeochemical cycles of essential elements, such as carbon, oxygen, nitrogen, sulphur, phosphorus and several metals Zhou et al. (2015a). These communities can survive extreme conditions, such as, temperature, pH and pressure. For example, they are found in the intestinal flora in the human digestive system, fighting infectious microbes and removing toxins, and inhospitable environments, eliminating pollutants, such as oils and chemicals Handelsman et al. (2007).

Furthermore, microbes are important to produce antibiotics fundamental for human health, ferment foods and drinks, produce biofuels and remediate soils and water bodies Timmis and Pieper (1999); Martínez (2008); Nigam and Singh (2011). Further characterization of the structure and community dynamics is important to understand their role in these processes to manipulate them to our benefit. Nevertheless, several limitations in sampling and cultivation pose considerable challenges Zhou et al. (2015a); Quince et al. (2008), since only around 1% of the microorganisms from environmental samples can be analysed with conventional microbiology techniques, leaving 99% of these organisms unstudied Kumar et al. (2015); Ranjan et al. (2016).

Recent technology advancements have allowed scientist to study these complex communities altogether, obtaining information on their capabilities and interactions Kumar et al. (2015). Through analysis of the genetic information in an environmental sample by employing genomic, bioinformatics and system biology techniques, without the need to more troublesome methodologies (e.g. isolation and cultivation individually) Handelsman et al. (1998).

2.3.1 *History of microbial community studies*

Begon et al. (1986) defined microbial communities as a group of organisms coexisting in the same time and space. Throughout 300 years, microbial communities were studied using microscopy and cultivation techniques to gather microbials morphology, growth and some biochemical profiles Franz (1988). The application of these methodologies gave birth to a new field of ecology, the microbial ecology, studying the microorganism and their environmental roles Ackert (2012).

In 1977, with the creation of the Sanger automated sequencing and the use of ribosomal RNA as a molecular marker to classify microorganisms revolutionized the field. The few decades following this discovery, several methods such as *Polymerase Chain Reaction* (*PCR*), rRNA genes cloning and sequencing, *Denaturing Gradient Gel Electrophoresis* (*DGGE*),

restriction-fragment length polymorphism and *Terminal Restriction-Fragment Length Polymorphism* (*T-RFLP*) were applied to further characterize these complex communities Escobar-Zepeda et al. (2015). Nonetheless, it was only with the development of gene expression techniques and gene cloning that was possible to discover new genes, functions and metabolic pathways, creating metagenomic analysis. This term was introduced in 1998, by Handelsman, to define the collection of genomes, from a microbial community of a specific environment Handelsman et al. (1998).

Although Sanger sequencing technology revolutionized the field, its sequencing length and yield only allowed the study of metagenomics by phylogenetic marker analysis Escobar-Zepeda et al. (2015). Nowadays, inexpensive technologies like *Next Generation Sequencing* (*NGS*), *High Throughput Screening* (*HTS*) methods and advanced bioinformatics tools triggered research in metagenomics yielding an enormous growth in microbial sequences and metabolic functions, expanding the tree of life and biogeochemical cycles Kumar et al. (2015); Quince et al. (2017).

For the last two decades, complex microbiomes have been investigated, bringing forth numerous discoveries. For example, the characterization of ammonia oxidizing archaea and species that carry out complete nitrification that play an important role in the nitrogen cycle Stahl and de la Torre (2012); Daims et al. (2015); Palomo et al. (2016). Moreover, application of these techniques in the study of the human gut microbiome found genes associated with antibiotic production, new pathogens and the linking between the microbial community and inflammatory bowel disease. These findings allowed a fine monitorization of changes at a strain-level after perturbations, such as fecal microbiome transplant Ranjan et al. (2016); Quince et al. (2017).

Recent analysis of the 1500 publicly available metagenomes retrieved around 8000 *Metagenome-Assembled Genomes* (*MAGs*). 17 bacterial and 3 archaeal new phyla were firstly described using the data retrieved in this study. Besides, close to 300 genomes belong to the Patescibacteria superphylum, that is made up of 35 candidate phyla. This work expanded the tree of life and provides sufficient material to further explore the unknown world of microbial communities and its diversity Parks et al. (2017).

Overall, metagenomic analysis has advanced several fields. Earth sciences, life sciences, biomedical sciences, bioenergy, bioremediation, biotechnology, agriculture and biodefense and microbial forensics, are some of the fields that received the biggest contribute with the advancement of microbial community analysis Handelsman et al. (2007).

2.4 MARKER GENE AMPLIFICATION METAGENOMICS

In this particular methodology, *PCR* amplification and sequencing of marker genes is used to make phylogenetic profiles to obtain the taxonomic distribution of the community. Marker genes, like 16 rRNA, are evolutionary conserved genes, with hypervariable regions, that allow to distinguish between taxonomic lineages and therefore, identify the organisms present in that environmental sample. The 16S rRNA is ubiquitous in bacteria, being considered an universal marker gene and has been employed the most to characterize microorganisms Kumar et al. (2015); Oulas et al. (2015). Other application of this technique is the use of functional genes as marker genes to target specific pathways, such as ammonium monooxygenase subunit A (*amoA*) and nitrogenase reductase (*nifH*) to detect nitrification and nitrogen fixation, respectively Kumar et al. (2015); Levy-Booth et al. (2014).

Although this technique is low cost and allows an initial screening of the community before further analysis, it has a few limitations. During *PCR*, the information that can be obtained is limited by the primers, loosing functional information or the complete lineage if primers mismatch, it requires previous knowledge about the community and the overall results only provide the relative abundance Zhou et al. (2015a). Since the taxonomic result is based on a putative association, an *Operational Taxonomic Unit* (*OTU*) defines the taxa. *OTU*s are less precise to identify the species, being mostly analysed by phylum to genera level, and therefore phylogenetic analysis by comparison limits the retrieved information Ranjan et al. (2016). In large metagenomic projects, marker gene amplification metagenomics evaluates the diversity of potential samples to choose the appropriate one for shotgun metagenomics Handelsman et al. (2007).

2.5 SHOTGUN METAGENOMICS

Shotgun metagenomics uses high-throughput sequencing to retrieve the complete DNA sequence present in the environmental sample. This method is carried out by the shredding of the extracted DNA that is then sequenced with a target gene. Since all information contained in the environmental sample is recovered with shotgun metagenomics, besides the taxonomical identification that the marker gene amplification methods provides, functional information can be obtained with further analysis Ranjan et al. (2016); Oulas et al. (2015).

In metagenomics, there are two main sequencing methods, the 454 Life Sciences and Illumina systems. The 454 Life sciences performs sequencing with a pyrosequencer, with which retrieves the genetic information similarly to the Sanger method, but by using light

to distinguish nucleotides rather than chain termination with dideoxynucleotides. On the other hand, Illumina systems (Illumina dye sequencing), used primers followed by three steps of amplification, sequencing and analysis. Between each round, the computer measures the wavelength of the fluorescent tag, recording the base that was added Oulas et al. (2015); Thomas et al. (2012).

Sequencing produces short reads of raw data that require several processes such as pre-processing, assembly, binning (for metagenomic samples) and annotation to yield the desirable information for functional and taxonomic analysis.

2.6 METAGENOMICS PROCEDURE OVERVIEW:

2.6.1 *Pre-processing*

The first step in a metagenomic analysis is the quality control to remove low quality reads and contaminants. The removal of low quality reads, although it sacrifices some desirable information, will later improve the accuracy of the analysis downstream. Impure samples or errors in sample preparation can lead to contaminant DNA, such as eukaryotic DNA, amongst the metagenome. Therefore, some trimming is necessary to ensure the quality of the reads Escobar-Zepeda et al. (2015); Zhou et al. (2015b).

Fastqc and Trimmomatic

Fastqc is easy to use modular software that report all quality control problems from the sequencer or the starting library and can be used interactively for a small number of files or through the command line Andrews (2010). The output can easily be integrated into larger pipelines, however it lacks trimming functionalities for quality and duplication filtration Zhou et al. (2015b). Consequently, it is normally coupled with Trimmomatic software to perform the trimming of the reads. Trimmomatic is a pair-aware, flexible and efficient pre-processing tool, optimized to work on paired-end data from Illumina system. It incorporates a series of steps for trimming and filtering, focusing on quality trimming instead of technical trimming Bolger et al. (2014).

Fastx-Toolkit

This particular software is a command line easy to install modular programs with read length trimming and format converting. The latest update released was in 2010 and is miss-

ing a functionality for contamination screening Lab (2009).

PRINSEQ

PRINSEQ is one of the most complete and user-friendly pre-processing software available. The output data can easily be integrated in pre-existing pipelines. Though, some essential functions such as, tag sequence removal, are not incorporated Schmieder and Edwards (2011).

NGS QC Toolkit

This versatile and adaptable software is suitable for input of raw data from Roche 454 and Illumina system. Besides, it has incorporated a format conversion tool for other sequencing methods, like Sanger and Solexa, and offers statistical analysis of the raw data. However, Fastx-Toolkit does not have a contamination screening functionality and it is not user friendly when performing multiple analysis Patel and Jain (2012).

Meta-QC-Chain

A fast quality control software that can perform status checking, sequencing-quality trimming and *de novo* identification of contamination. The pipeline is divided into four essential steps: (1) the technical test, retrieving a summary of the input data, (2) sequencing-quality trimming, to assertion the raw data quality and remove tag, duplication and errors, (3) contamination screening, recovering and alignment of 18S rRNA (universal marker gene of eukaryotes) and finally (4) contamination removal through employment of mapping tools. Though a very effective tool it is unable to identify virus contamination and lacks compatibility with other analysis pipelines Zhou et al. (2014).

2.6.2 *Assembly*

Assembly is the process of merging short reads obtained during sequencing, into larger fragments, named contigs, to reconstruct the original sequence. Contigs are built by aligning the reads and obtain a consensus sequence based on the overlaps. The development of bioinformatic tools allowed the creation of assemblers designed for samples holding multiple genomes, opposing to the previously used assemblers for whole genomes sequencing

Thomas et al. (2012).

There are two processes of assembling the short reads into contigs, the reference-based assembly and the *de novo* assembly. In the reference-based assembly method, several reference genomes are used to map against the short reads, to create the contigs accordingly to the reference. Since larger contigs can be obtained when mapping with reference genomes, the new sequences that not present similarity will not be considered, therefore resulting in fewer contigs. For this reason, this method is applied to already well known microbial communities or when targeting specific organisms Oulas et al. (2015); Thomas et al. (2012).

On the other hand, *de novo* assembly methods generate contigs without having known genomes as a reference. Informatically speaking, this method requires a large amount of processing power and memory to run the complex algorithms necessary to compare the reads to each other. The most common algorithms employed in the *de novo* assembly are the graphs algorithms, such as *de*-Bruijn and string graphs or greedy algorithms Quince et al. (2017); Bang-Jensen et al. (2004).

The *de*-Bruijn graph algorithm is the most popular choice for *de novo* assembly. A graph is a mathematical structure represented by vertices, or nodes, that are connected by edges or lines. The edges of a graph can be directed or undirected. *de*-Bruijn graph is a directed graph constructed by all reads in the raw data, represented as the nodes of the graphs, and the edge are the overlap between reads. In an assembly algorithm, the reads are broken down into smaller sequences of k size named k-mers, and a Hamiltonian cycle is performed, creating a path that travels through each node, reconstructing the genome. Assembly by this method is hampered by sequencing errors and repetitive sequences that can cause fragmentation of the assembly and/or misassembles Quince et al. (2017); Compeau et al. (2011).

MetaVelvet, Meta-IDBA or IDBA-UD, SOAPdenovo2, MetaSPAdes are some of the most used metagenomic assemblers. Overall, MetaSPAdes and IDBA-UD are the assemblers with the best performance. Metagenomic reads assembled with MetaSPAdes have the best quality, genome representation and number of complete genes, while for IDBA-UD, a narrow number of contigs with a larger size are obtained (Table 1).

2.6.3 *Binning*

The output of the assembly is a mixture of contigs from all organisms present in the environmental sample. Therefore, binning is applied to cluster the contigs into individual

Table 1.: Comparison of the most used metagenomic assemblers, adaptated from Gurevich et al. (2013); Vollmers et al. (2017)

Assembler	Method	Input Format	Read Pair Format	No. of contigs	NGA50 (bp)	Largest (bp)	Genome fraction (%)	No. of complete genes
IDBA-UD	de Bruijn multiple Kmer	.fasta	interleaved only	283	90 607	224 018	95.90	4030
SOAPdenovo	de Bruijn multiple Kmer	.fastq, fastq.gz, .fasta, fasta.gz, .bam	interleaved or seperate	817	16 606	87 533	81.36	3060
SPAdes	de Bruijn multiple Kmer	.fastq, .fastq.gz, .fasta, .fasta,gz, .bam	interleaved or seperate	532	99 913	211 020	96.99	4071
Velvet	de Bruijn multiple Kmer	.fastq, .fastq.gz, .fasta, .fasta.gz, .sam, .bam, .stdin	interleaved or seperate	310	22 648	132 865	75.53	3121
Velvet-SC	de Bruijn multiple Kmer	.fastq, .fastq.gz, .fasta, .fasta.gz, .sam, .bam, .stdin	interleaved or seperate	617	19 791	121 367	93.31	3662

genomes of each specie, or bins. The clustering of the genomes can be done based on, either contig composition or similarity. Composition-based binning employs genomic signatures, conserved oligonucleotide composition specific of species, that are detected in *k*-mers frequency. Similarity based binning applies alignment algorithms such as, BLAST or *profile Hidden Markov Models* (*pHMM*) to find similarities about specific genes from public databases. Afterwards, the contigs with similar classification are clustered together as an individual genome Quince et al. (2017); Oulas et al. (2015); Thomas et al. (2012).

VizBin

VizBin is a composition-based binning method that employs pentanucleotide signatures for the clustering of the reads. After binning, the visualization of the metagenomic contigs allows the user to manually select and retrieve the desirable *Composite Genome* (*CG*). The software is fast and easy to use, exploiting the capabilities of the human eye for pattern recognition, superior to automated system to provide better results Laczny et al. (2015).

MaxBin

MaxBin is a automated binning software using an expectation-maximization algorithm. This binning approach combines tetranucleotide signatures with differential coverage from

reads previously identified with single-copy marker genes. Since this software is completely automated, the errors originated from the user are excluded. After binning the CG's quality is evaluated using the 107 single-copy marker genes. Although the aumatation of the software allows less input from the user, unsupervised binning methods still loose efficiency when compared with human recognition and visualization Wu et al. (2014); Laczny et al. (2015).

Mmgenome

Mmgenome applies a two-step binning procedure, a first step using composition-independent information, followed by a refinement based on genome composition. First, the binning, by differential coverage maps the reads against other samples to produce the coverage by contig. Additional information, such as GC content, tetranucleotide frequency and essential single-copy genes, is also retrieved for the refinement. Finally, the plotting of the two samples using their coverage estimates clusters the contigs creating possible bins. The use of mmgenome needs at least two similar metagenomes to perform the analysis and prior knowledge of R, since the tool is a R package Albertsen et al. (2013).

2.6.4 *Quality Assessment*

Thousands of draft genomes have been produced from metagenomes, with the creation of new assembly and binning methods, increasing the need to ascertain the quality of the genomes. Besides the metrics employed to obtain the quality of an assembly, such as N50, the use of single-copy marker genes is fundamental to qualify the genome in terms of completeness and contamination Mikheenko et al. (2016). The percentage of universal single-copy marker genes found in a genome indicates comprehensiveness, while the presence of multiple copies reveals contamination, since each genome only contains one cop y Parks et al. (2015).

QUAST is a quality assessment tool that provides assembly and binning quality control with an easy to use interface and visualization Gurevich et al. (2013). Since QUAST uses universal single-copy marker genes, the accuracy of completeness can be decreased due to their uneven distribution across genomes. CheckM is another quality assessment tool that builds a library of single-copy genes specific to the genomes lineage. The use of lineage single-copy genes instead of the usual marker genes decreases the error in completeness and contamination estimates, making the process more robust. However, there are a few disadvantages with this tool which is the presence of bias in genome quality estimates and the lack of some reference genomes from lineages that are badly represented Parks et al.

(2015).

2.6.5 *Taxonomical annotation*

The 16S rRNA gene is a highly conserved between species, and present in all organism from the Bacteria and Archaea domain, therefore is the main gene used in taxonomic classification. But, in *MAGs*, the 16S rRNA gene is often missing as their conservation and repetitiveness impedes assembly or is lost is binning Parks et al. (2017). Some methods to attribute a taxonomic classification based on BLAST have been created, using the lowest common ancestor in a similarity search (MEGAN) Huson et al. (2007) or by comparing sequences (Kaiju) Menzel et al. (2016) or k-mers (Kraken) Wood et al. (2014) to proteins reference databases, to overcome these limitations Menzel et al. (2016); Luo et al. (2014); Jovel et al. (2016).

2.6.6 *Genome annotation*

The last step is a metagenomic analysis pipeline is the gene prediction followed by their annotation. Genome annotation characterizes the genes of a genome and attributes them a function, functional annotation. Gene prediction tools for metagenomics are more advantageous when considering di-codons frequency, bias in codon usage, patterns in start and stop codons and that incorporates the patterns of species-specific ribosome-binding sites Escobar-Zepeda et al. (2015). To obtain the functional annotation of a genomes, software and pipelines that use USEARCH Edgar (2010), BLAST or hidden Markov model-based algorithms (HMMER) Finn et al. (2011) to search homologies of the predicted coding sequences (CDS) against public databases (e.g. NCBI, Uniprot, KEGG; Ogata et al. (1999), SEED Overbeek et al. (2014) and Pfam Finn et al. (2014). Nowadays, several online rapid genome annotation software with integrated server exist, such as WebMGA Wu et al. (2011) and MG-RAST Aziz et al. (2008); Keegan et al. (2016). These tools are fast and perform the gene prediction and annotation automatically. Unfortunately, automatic annotation still requires improvement since only 20 to 50% of sequences can be annotated Thomas et al. (2012).

3

CASE STUDIES

Metagenomics analysis of a single-stage nitritation/anammox sequencing batch reactor.

Daniela Azevedo, Alejandro Palomo, Barth F. Smets

Manuscript in Preparation

Metagenomics analysis of a single-stage nitritation/anammox sequencing batch reactor.

Daniela Azevedo[1], Alejandro Palomo[1,2], Barth F. Smets[1]

[1]Department of Environmental Engineering, Technical University of Denmark, Kgs Lyngby, Denmark

[2]Department of Bio and Health Informatics, Technical University of Denmark, Kgs Lyngby, Denmark

1. Introduction

The increase of industrial, agricultural and domestic organic waste compromises the water quality, so the removal of excessive pollutants is necessary (Kneese, 2015). The increased input of reactive nitrogen into the environment leads to deregulation of the biogeochemical nitrogen cycle, causing eutrophication of water bodies and diminishing the ozone layer (Socolow, 1999; Galloway, 2005; Sutton *et al.*, 2011; Suddick *et al.*, 2013). In wastewater treatment plants, microbial communities carry nitrogen removal, commonly by nitrification-denitrification (Isaacs and Henze, 1995; Whitacre, 2013; Yin and Kong, 2014). This process requires continuous aeration for nitrification and the supply of large amounts of organic carbon for denitrification (Kartal, Kuenen and van Loosdrecht, 2010; Hu *et al.*, 2013). Complete denitrification efficiency greatly depends on the C:N ratio, a quantity of carbon not found in mainstream wastewater, needing external sources to be incorporated (Sobieszuk and Szewczyk, 2006; Volcke, Vanrolleghem and van Loosdrecht, 2006). The release of nitrous oxide (N_2O), an impactful greenhouse gas, during denitrification make the overall process costly and a risk for human lives (Tsuneda *et al.*, 2005; Kampschreur *et al.*, 2009; Zhao *et al.*, 2017).

Ammonia oxidation, a process thought to only occur in aerobic conditions, was discovered in anaerobic ammonium oxidizing (anammox) bacteria, converting ammonia (NH_4^+) to Nitrogen (N_2) using nitrite (NO_2^-) as an electron acceptor(Fux *et al.*, 2002; Hu *et al.*, 2013). These bacteria form a clade inside the Planctomycetes phylum, with five genera found in diverse aquatic environments (Kartal *et al.*, 2013; Oshiki *et al.*, 2015). They contain an unique organelle called anammoxosome, where anammox catabolism takes place, protecting the cell from toxic intermediates such as hydrazine (Pynaert *et al.*, 2003; Dietl *et al.*, 2015). The use of anammox bacteria is profitable but it is decreased by their slow growth rate of 10 to 12 days per generation time at 35ºC and need of NO_2^- supply. (Pynaert *et al.*, 2003).

In Partial Nitrification/Anammox (PNA) process, ammonium oxidizing bacteria (AOB) partially oxidize a portion of NH_4^+ to NO_2^- (nitritation) (Fux *et al.*, 2002; Schmidt *et al.*, 2003; Lackner *et al.*, 2014). Then, anammox bacteria use NH_4^+ and NO_2^- to catabolize N_2 (Kartal, Kuenen and van Loosdrecht, 2010; Hu *et al.*, 2013; Speth *et al.*, 2016). Since both bacteria are autotrophs, the need to externally provide carbon and other electron acceptors is removed. Besides, nitritation required less oxygen, decreasing the amount of aeration (Strous *et al.*, 1998; Kuenen, 2008; Hu *et al.*, 2013).

As a result, applying PNA in wastewater treatment requires less energy and produces a minimal quantity of greenhouse gases, coming from the heterotrophs present in the community, accomplishing an economic and ecological process (Siegrist *et al.*, 2008; Joss *et al.*, 2011).

PNA needs optimal ratios of NO_2^- and NH_4^+ and rely heavily on the synergy inside the community. Besides anammox bacteria and AOB, an enormous diversity of heterotrophic bacteria whose taxonomy and function in this systems is still unknown (Agrawal *et al.*, 2017; Bhattacharjee *et al.*, 2017).

Some studies have investigated PNA systems, but little is understandable of the organisms present and the role they partake (Costa *et al.*, 2014; Chu *et al.*, 2015). An understanding of the composition and metabolism inside the system may hint to an enhanced biological nitrogen removal. Most studies focus on the role of AOB and anammox bacteria, so only description of the denitrifiers' metabolism is found without them being the main actors (Langone *et al.*, 2014; Speth *et al.*, 2016; Bhattacharjee *et al.*, 2017; Lawson *et al.*, 2017). Further research is needed to mitigate the N_2O and nitrate (NO_3^-) production and the causes of the instabilities that arise (Fux and Siegrist, 2004; Lackner *et al.*, 2014).

In this work, over six months, seven samples were retrieved from a high-rate PNA sequencing batch reactor with regular periods of aeration time and minimal N_2O emissions. This reactor was fed ammonium-rich synthetic wastewater with no carbon supply and each sample was subject to shotgun sequencing. Metagenomics analysis and genome reconstruction was performed, recovering 45 almost complete genomes. With these genomes, an insight of heterotrophs' role can lead to an understanding of the community.

2. Materials and Methods

2.1. Library preparation, sequencing and *de novo* assembly

DNA-shearing and library preparation were conducted based on the NEXTflex Rapid DNA-Seq Kit, V13.08 (Bioo Scientific, Austin, TX, USA). Briefly, 250 ng genomic DNA was sheared with the Covaris E210 System using 10% duty cycle, intensity of 5, cycles per burst of 200 for 300 s to create 200-bp fragments. The samples were end-repaired and adenylated to produce an A-overhang. Adapters with unique barcodes were ligated to

the DNA. The samples were purified using bead-size selection for range ~ 300-400 bp with the Agencourt AMPure XP beads (Beckman Coulter, Beverly, MA, USA). Amplification was performed to the purified DNA libraries following the manufacturer's protocol: initial denaturation, (2 min, 98ºC), followed by 12 cycles of denaturation (98ºC, 30 s), annealing (65ºC, 30 s), extension (72ºC, 1 min) and final extension (72ºC, 4 min). DNA was quantified using NanoDrop ND-1000 UV-VIS Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and quality was checked on an Agilent 2100 Bioanalyzer using the High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, CA, USA). The DNA libraries were mixed in equimolar ratios. Sequencing was performed as a 100-bp pair-end run on HiSeq 2000 (Illumina Int., San Diego, CA, USA) at BGI (Copenhagen, Denmark). *Trimmomatic* v0.22 (Bolger, Lohse and Usadel, 2014) was used to remove adapters and trim the reads (threshold quality = 15; minimum length = 45). Quality control was carried out using *FastQC* (Andrews, 2010). Assembly of high-quality reads from each sample into contigs was performed using *IDBA-UD* (Peng *et al.*, 2012) with default parameters.

2.2. Binning composite genome assembly and annotation

For all samples, contigs larger than 1000 nucleotides were clustered into putative taxonomic groups based on pentanucleotide signatures using VizBin (Laczny *et al.*, 2015) or differential coverage using mmgenome (Albertsen *et al.*, 2013). Clustered reads were *de novo* assembled as described previously, to construct composite genomes (CGs). Resultant reconstructed CGs were manually evaluated through contig depth and GC content and accurate bin segregation was validated. Completeness and potential contamination of each CG was evaluated using 107 essential single-copy genes, using QUAST (Albertsen *et al.*, 2013; Gurevich *et al.*, 2013). When the same draft CG was obtained from several samples, comparison between them was implemented, retaining the one best assembled, most complete, and with the lowest contamination. CGs that contained at least 70% of the single-copy genes, or an inferior number without contamination were further analysed. Genomes were identified to appropriate taxonomic levels based on two approaches: (i) A set of 107 essential single-copy genes were blasted (BLASTP, E< $1e^{-5}$) against the NCBI-nr database with follow-up analysis in MEGAN (Huson *et al.*, 2007) to identify the lowest common ancestor of the blast output, (ii) using the standard settings of Kaiju (Menzel, Ng and Krogh, 2016). A taxon was assigned when at least 75% of the identity resulted in an agreeable taxonomy. For genomic comparison between CGs, or between CGs and reference genomes, the average amino-acid identity (AAI) was calculated between genomic data sets of proteins using the AAI calculator (Rodriguez-r and Konstantinidis, 2014). The relative abundance of the organisms represented by the CGs was calculated by mapping the original, quality-filtered reads from each sample against the CG sequence. Furthermore, to confirm protein functional assignment, Kyoto encyclopaedia of genes and genomes (KEGG)

annotations of the predicted proteins in each CG were obtained using the WebMGA server (Wu *et al.*, 2011). Presence of the complete operon was evaluated in MG-RAST (Keegan, Glass and Meyer, 2016).

2.3. Taxonomic and functional annotation

Predicted coding sequences retrieved using *Prodigal* 2.50 (Hyatt *et al.*, 2010) were annotated using USEARCH (Edgar, 2010) -*ublast* against the manually created databases of reference proteins encoded by genes of interest (best hit with $E < 1e^{-5}$, Bitscore > 60 and sequence similarity > 30%).

2.4. Phylogenetic analysis

Predicted amino-acid sequences from the metagenomic samples and CGs were aligned with reference sequences using MUSCLE (Edgar, 2004). Multiple alignments were manually revised before phylogenetic analysis. Bootstrapped maximum likelihood trees were constructed in MEGA7 using the Jones Taylor Thornton model with 500 replicates (Kumar, Stecher and Tamura, 2016).

3. Results and discussion

3.1. Community overview

Contigs retrieved from the co-assembly were binned using two distinct approaches of differential coverage, mmgenome (Albertsen *et al.*, 2013), per samples and coassembly, and pentanucleotide frequency, VizBin (Supplementary Information 1)(Laczny *et al.*, 2015). From the seven samples, a total of roughly 400 genomes were recovered, resulting in 57 unique metagenome-assembled genomes (MAGs), with 45 genomes being near complete genomes, with completeness above 75% (Supplementary Table 1).

These genomes account for 80% of the metagenome, represented by the most common phyla, *Proteobacteria* and *Chloroflexi*, and genomes from *Planctomycetes*, Bacteroidetes, *Armatimonadetes*, *Ignavibacteriae*, *Acidobacteria*, *Chlorobi*, *Verrucomicrobia*, *Actinobacteria* and *Gemmatimonadetes* phyla (Supplementary Table 1).

As expected, some of the most abundant genomes are anammox bacteria and AOB, accounting for 28% and 10% of the metagenomes, correspondingly (Figure 1). Other genomes considered abundant belong to the phyla, *Armatimonadetes* (one), *Bacteroidetes* (one), *Chloroflexi* (three), *Proteobacteria* (two), *Acidobacteria* (one) and

one genome from the Bacteroidetes/Chlorobi group (Supplementary Table 1). Each individual MAG was assigned a bin ID to ease their identification hereafter.

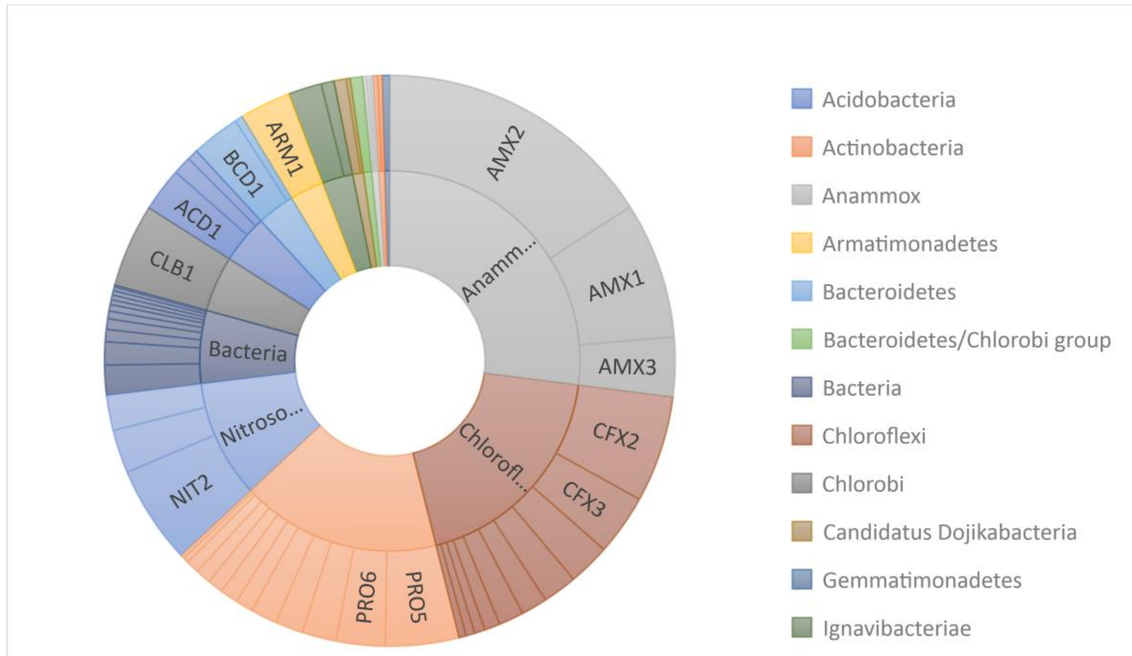


Figure 1 – Average abundance of the genomes retrieved in PNA system, divided by phylum (inner layer) and correspondent genomes (outer layer).

3.2. Autotrophs vs Heterotrophs

The PNA reactor was fed with carbon-free synthetic wastewater, making autotrophic organisms crucial in this community to provide organic carbon to heterotrophs. Of the four CO_2 fixation pathways usually found in bacteria, Reductive pentose phosphate, Reductive tricarboxylic acid, Reductive acetyl-CoA and 3-Hydroxylamine bicycle, only two pathways were present.(Hügler and Sievert, 2011). Eight organisms representing 40% of the community encode Reductive acetyl-CoA and Reductive Pentose Phosphate, sustaining the heterotrophic organisms in this reactor.

Genomes AMX1, AMX2, AMX3 and CFX7 had the Reductive acetyl-CoA pathway, a pathway found in Planctomycetes, Deltaproteobacteria, Spirochaetes and Firmicutes (Acetogens). Although species from the Chloroflexi phylum normally use either the 3-Hydroxylamine bicycle or the Reductive pentose phosphate cycle, in a recent study, it was discovered the presence of this fixation pathway in Chloroflexi (Ragsdale and Pierce, 2009; Hügler and Sievert, 2011).

NIT1, NIT2, NIT3 and PRO3, from the Betaproteobacteria class, use the Reductive pentose phosphate cycle for CO_2 fixation. Most autotrophs of the Alpha-, Beta- and

Gamma- proteobacteria utilize this cycle, therefore the results are congruent to previous studies (Hügler and Sievert, 2011).

3.3. Anammox and AOB

The 16S rRNA of anammox and AOB species was obtained exclusively when investigating the whole metagenome rather than the individual draft genomes. Comparison of these genomes to reference genomes based on the average amino acid identity revealed novel species of AOB and anammox bacteria, NIT1, NIT3 and AMX2 (Supplementary Figure 1) (Rodriguez-r and Konstantinidis, 2014). Taxonomy identification assigned NIT2 as a strain of *Nitrosomonas europeae*, similarity below 98%, AMX1 as *Candidatus* Brocadia fulgida, with similarity above 98% and AMX3 as a strain of *Candidatus* Brocadia sp. UTAMX1 (Luo, Rodriguez-R and Konstantinidis, 2014; Rodriguez-r and Konstantinidis, 2014; Lawson *et al.*, 2017).

All AOB genomes have the nitritation genes, ammonia monooxygenase (*amoCABDE*) operon and hydroxylamine oxidase (*hao*). Copper resistance (*copCD*) genes normally located downstream of the AMO operon were also present (Stein *et al.*, 2007)). Besides the nitrification process, where AOB retrieve most of their energy, these organisms can grow in oxygen limitation by using denitrification. The NIT genomes contain a copper-containing nitrite reductase (*nirK*) and cytochrome bc-type complex cNOR (respiratory nitric oxide reductase, *norBC*) (Hendriks *et al.*, 2000). Some studies suggest the use of denitrification in AOBs to confer tolerance of the bacteria to high quantities of NO_2^- (Beaumont *et al.*, 2004; Schmidt, van Spanning and Jetten, 2004). In PNA system, since the production of NO_3^- is supressed to reduce aeration costs, these genes confer protect to the cells against the NO_2^- produced.

Anammox bacteria harboured the hydrazine oxidoreductase (*hzo*) and nitrate oxidoreductase (*nxrAB*). Unexpectedly, only AMX2 encoded for the hydrazine synthase (*hzsABC*) operon. Analysis of the metagenome showed the presence of merely two copies of this operon, revealing the absence of these genes in one of the anammox genomes. As a biomarker, HZS operon indicates the presence of anammox pathway in a system, implying the existence of just two organisms capable of anaerobic ammonium oxidation (Harhangi *et al.*, 2012; Wang *et al.*, 2012). Analysis based on the amino acid identity reveal similarity below 85%, proving that each genomes belongs to different species but to the same genus (Luo, Rodriguez-R and Konstantinidis, 2014) (Supplementary Figure 2).

Nitrite reductase (*nirS* or *nirK*) genes is present in several organisms and is used by both denitrification and anammox pathway. Anammox bacteria require nitric oxide (NO), a substrate of hydrazine synthase, but studies have indicated the lack of these in some of these organisms. *Candidatus* Kuenenia encode the cytochrome cd_1 nitrite reductase

(*nirS*) gene where *Candidatus* Brocadia, do not have neither *nirS* or *nirK* (Oshiki *et al.*, 2015), or in some cases, only *nirK* (Ferousi *et al.*, 2013). Results show the presence of *nirK* in AMX1 and the non-existence of this gene in AMX2 and AMX3.

hao-like proteins were found in AMX organisms, ten in AMX1, ten in AMX2 and five in AMX3. Phylogenetic analysis corroborated the orthography of these proteins with previously published (Supplementary Figure 3) (Kartal *et al.*, 2011; Maalcke *et al.*, 2014; Oshiki *et al.*, 2016; Lawson *et al.*, 2017). Two novel reactions believed to replace the function of nitrite reductase in anammox pathway were observed, catalysed by *hao*-like proteins, which are amongst the most expressed proteins in these organisms (Lawson *et al.*, 2017). These new proposed reactions, the oxidation of hydroxylamine to nitric oxide (kustc1061 in *Candidatus* Kuenenia stuttgartiensis)(Maalcke *et al.*, 2014) and reduction of nitrite to hydroxylamine (brosi_A0131, brosi_A0501, brosi_A3534, brosi_A3864 in *Candidatus* Brocadia sinica JPN1) (Oshiki *et al.*, 2015, 2016) produce hydroxylamine or nitric oxide, two substrates accepted by hydrazine synthase (Dietl *et al.*, 2015). All AMX genomes have both proteins, but further information is needed to fully understand the complete anammox pathway.

3.4. Denitrification niches in a PNA reactor

Almost all genomes encoded at least one step of the denitrification pathway, but complete denitrification is present exclusively in PRO3. Denitrification is fragmented throughout the genomes, with absence of nitric oxide reductase genes compared to other steps of the process. Excluding the reduction of NO_3^-, the presence of nitrite reductase (*nirK* and *nirS*), nitric oxide reductase (*norBC* and *norZ*) and nitrous oxide reductase (typical and atypical *nosZ*) was studied. Similarities in the partitioning of the denitrification from NO_2^- to N_2 were explored, dividing the heterotrophic community in four niches (Figure 2).

Nitrite reducers, Internal deletion, nitrite to nitrous oxide reducers and nitrous oxide reducers were the name given to these niches. Nitrite reducers groups the organism that only reduce NO_2^- to NO, internal deletion, due to lack of NOR genes, contains all the organism that reduce NO_2^- to NO and N_2O to N_2, nitrite to nitrous oxide reducers do not encode the *nosZ* gene and nitrous oxide reducers only encode for this gene. The fragmentation in the denitrification pathway promotes the production of NO and the uptake of N_2O. Although a decrease of N_2O during the experiment was observed, the levels of NO detected in the reactor does not demonstrate an accumulation of this product (Domingo-Félez *et al.*, 2014).

Genomes encoding full denitrification, PRO3 and PRO9 contained cytochrome-cd1 nitrite reductase (*nirS*), while most of the nitrite reductase genes found in this metagenome are copper-containing nitrite reductase (*nirK*). Both nitrite reductases

were in PRO7. Of the heterotrophic community, of the 19 genomes with *nirK*, merely 31% also had *nosZ*, according to previous studies that depict a percentage between 10 to 30% (Hallin *et al.*, 2017). The reduction of NO was found only in NIT1, NIT2, NIT3 and PRO12.

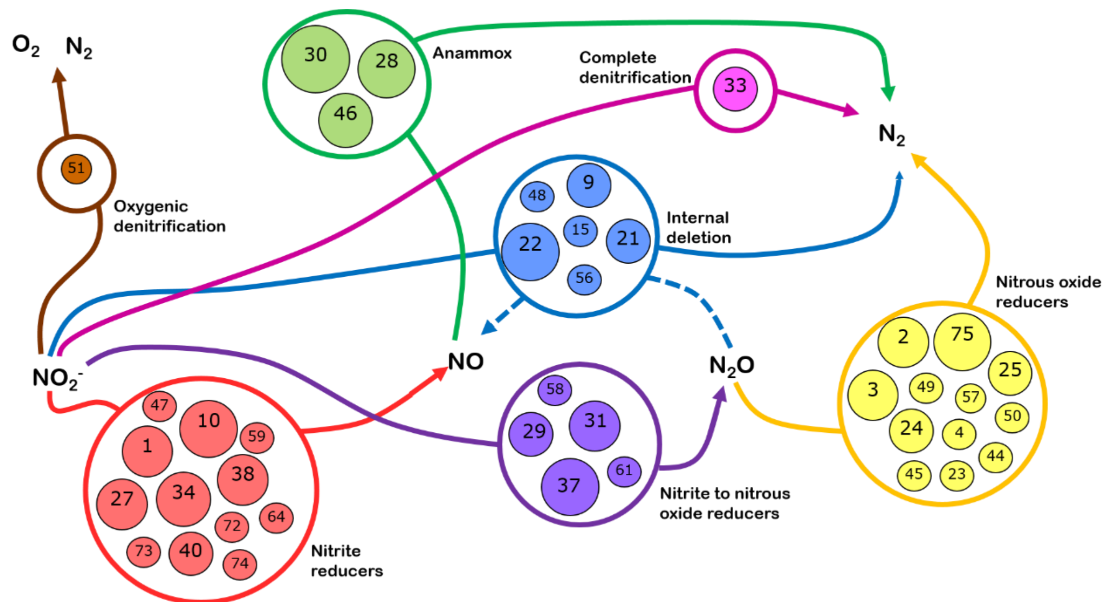


Figure 2 – Model of predicted niches based on the nitrogen oxides reduction from NO_2^- to N_2 gas, and anammox pathway (green). The circle represents each MAG by its numerical ID (Supplementary Table 1) and the size demonstrates the average abundance.

Organisms encoding *nosZ* are necessary to act as N_2O sinkers. Recently, *nosZ* has been divided into two clades, clade I or typical *nosZ*, discovered mainly in Proteobacteria with complete denitrification, and clade II or atypical *nosZ* (Sanford *et al.*, 2012). Genomes containing atypical *nosZ*, as partial denitrifiers, its main function is to reduce the amount of N_2O, reducing the quantity released to the atmosphere. This gene was previously found in organisms of the Bacteroidetes and Gemmatimonadetes phyla, as well as some of the Deltaproteobacteria and Gammaproteobacteria class (Hallin *et al.*, 2017). Interestingly, of the 17 genomes with atypical nosZ, merely six belong to the taxonomic groups mentioned previously. Around half are assigned as Chloroflexi, with the remaining consigned in Ignavibacteriae and Betaproteobacteria (Supplementary Figure 4).

Most studies of metagenomic analysis in PNA reactors describe this fragmentation but no information besides its annotation is provided (Speth *et al.*, 2016; Lawson *et al.*, 2017). Phenotypic and genotypic analysis comparing partial denitrifiers in soil prove the viability of the genotypes described in this work (Lycus *et al.*, 2017; Roco *et al.*, 2017). With the niche partition study performed in this work, we can describe a synergy

between all organisms to mutually share their synthesis products, promoting complete denitrification, to improve overall nitrogen removal.

3.5. Complexity of the NH_4^+ oxidation pathways

Ammonia oxidation in a PNA system is normally performed by AOBs (nitritation) and anammox bacteria (anaerobic ammonium oxidation). However, two new groups of heterotrophic organisms performing nitrification were found, putative heterotrophic hydroxylamine oxidizers and putative heterotrophic nitrite oxidizers.

Heterotrophic organisms, PRO5 and PRO11 have the genes responsible for hydroxylamine oxidation. Results from phylogenetic analysis of the *haoA* gene illustrate two distinctive clades, one exhibiting relatedness to Nitrosomonas and Nitrospira genus, whereas the second cluster with newly found genes (Kantor *et al.*, 2015) (Supplementary Figure 5). Previous studies have observed heterotrophic bacteria uptaking hydroxylamine and then reducing it to detoxify the cells (Padhi *et al.*, 2017). Although the mechanisms are different, the ability of these organisms to uptake hydroxylamine indicates the existence of transporters for hydroxylamine, enabling the described metabolism (Cabello *et al.*, 2004).

Nitrite oxidation is a process considered unique of NOB (Woznica *et al.*, 2013). Six heterotrophic bacteria, AMR1, PRO6, CFX9, BCT1 (with two distinct copies), PRO11 and PRO12, encode the *nxr* genes, contrary to normally believed. Hence, phylogenetic analysis was done to validate the acquired results (Figure 3). Results showed a novel cluster consisting of the sequences associated with heterotrophic nitrite oxidation. Interestingly, three metagenomic studies in aquifers, PNA reactor and subsurface groundwater found sequences from heterotrophs containing the *nxr* genes (Anantharaman *et al.*, 2016; Hug *et al.*, 2016; Speth *et al.*, 2016). The production of NO_3^- using NO_2^- has already been described in several studies, but by utilizing different pathways (Sakai *et al.*, 1996, 1997; Matsuzaka *et al.*, 2003; Abeliovich, 2006; Zhao *et al.*, 2012). This new group of heterotrophic bacteria display diversity with organisms from *Armatimonadetes* (Speth *et al.*, 2016), *Proteobacteria*, *Chloroflexi* and *Candidatus* Rokubacteria phyla (Hug *et al.*, 2016). Few information can be found, and the impact of this new process has in wastewater treatment has yet to be determined.

Almost half of the genomes encoded for respiratory nitrate reductase genes (*narGHIJ*), periplasmatic nitrate reductase (*napAB*) or the genes responsible for the dissimilatory nitrite reduction to ammonia (DNRA) via pentaheme nitrite reductase (*nrfHA*). Complete DNRA was merely found in IGN1, IGN2, ACD2, CFX8 and BCT11.

In wastewater systems where influents have low ration of COD/N, DNRA-anammox is favoured over denitrification under NO_3^- limiting conditions (Castro-Barros *et al.*, 2017).

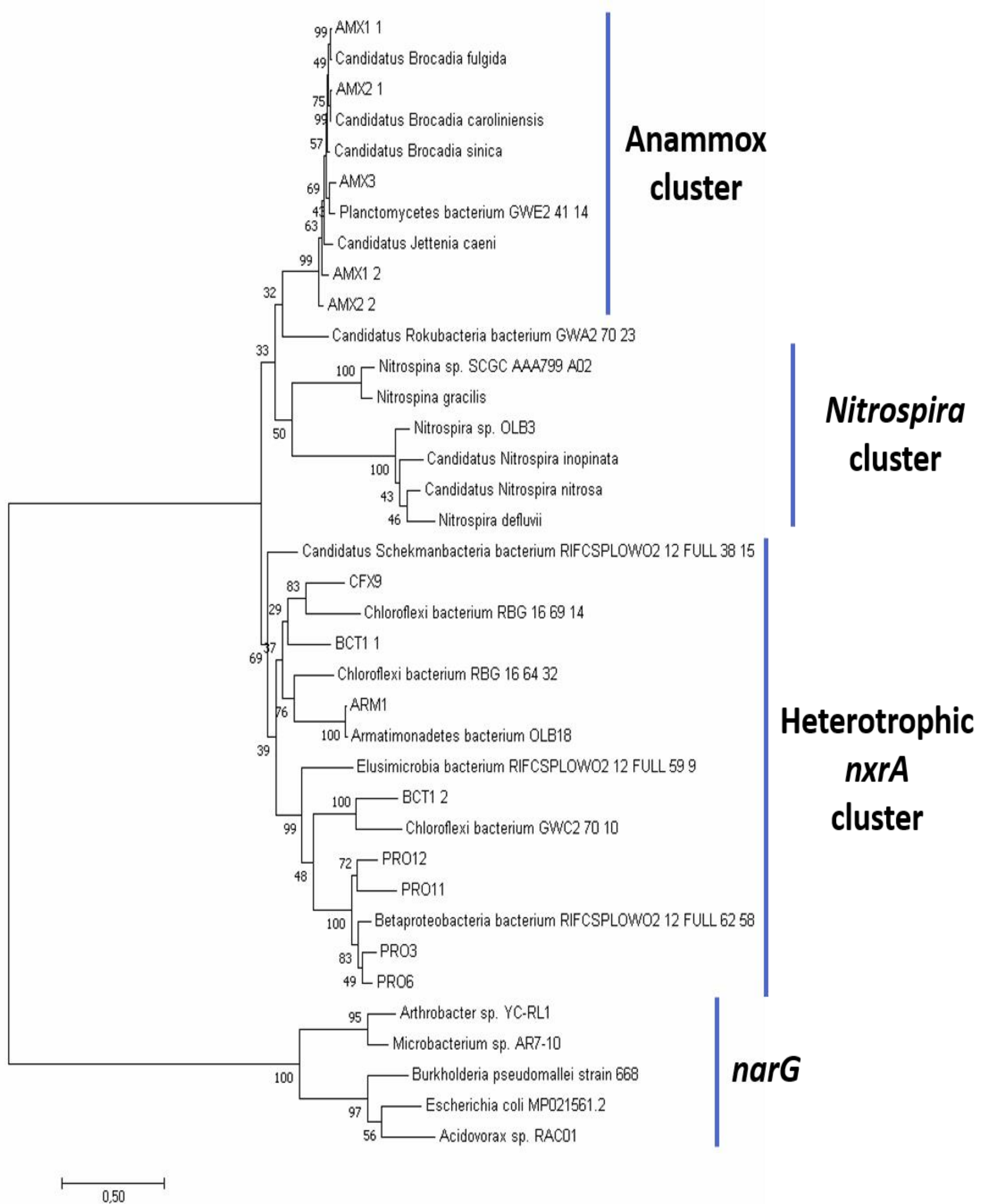


Figure 3 – Neighbour-joining phylogenetic tree of *nxrA* from Anammox and *Nitrospira*, putative heterotrophic *nxrA* and *narG*, using a bootstrap of 500 replicates. A total of three sub-clusters of *nxrA* are formed, with the putative heterotrophic nitrite oxidizers differing from the *narG* cluster. This separation confirms the functional annotation performed to these proteins.

With the discovery of a putative-NOB and the conversion of NO_2^- to NO_3^- performed by anammox bacteria, the partial denitrifiers that convert NO_3^- back to NO_2^- can create a nitrite loop. CFX2 and CLB1, the most abundant denitrifiers, encode for the *narGHIJ* genes, can be the main organism cooperating with anammox and the putative-NOB in the nitrite loop, as proposed by Steph et al. (2016) (Speth *et al.*, 2016). The recycling of nutrients for anammox bacteria, either by nitrite loop or DNRA still requires further studies but can lead to an improved nitrogen removal efficiency.

3.6. New NOB in Burkholderiales order

Nitrite oxidizing bacteria (NOB) are chemolithoautotrophic bacteria that use NO_2^- as energy source, during the second step of nitrification (Woznica *et al.*, 2013). PRO3, a putative nitrite oxidizing bacteria, encodes for the enzyme responsible for nitrite oxidation, nitrite oxidoreductase (*nxr*). Further analysis revealed the presence of the hydroxylamine oxidase (*hao*), the complete pathway of the CBB cycle for carbon fixation (Reductive Pentose Phosphate cycle) and several pathways of cytochrome oxidases. Furthermore, this organism also harbours the required genes for complete denitrification. Further investigation to the proteins present in the UniProt database shows an absence of a previously described NOB encoding the complete denitrification.

This autotrophic bacterium contains all the required genes to receive a classification as an NOB, but further studies need to be performed to fully confirm it.

3.7. NOD and unknown Nor-related genes

The presence of NOR-related genes was observed during genomic annotation, in PRO9 and BCD2. These genes belong to the novel oxygenic denitrification where N_2 and molecular oxygen are the final products, without emission of N_2O (Ettwig *et al.*, 2010; Luesken *et al.*, 2011; Zhu *et al.*, 2017). The new putative nitric-oxide dismutase (*nod*) and Nor-related gene from the FCB group (comprised of the Fibrobacteres, Bacteroidetes and Chlorobi phylum) convert NO to N_2.

A phylogenetic tree of all published *nod*, some unknown Nor-related, quinol-dependent nitric oxide reductase *qNor* and *cNor*, assigns PRO9 gene as a *nod* while BCD2 clustered with the unknown Nor-related genes (Figure 4).

PRO9 harbours the complete oxygenic denitrification pathway. The oxygen generated during the dismutase reaction can be used as an electron acceptor (Ettwig *et al.*, 2010; Zhu *et al.*, 2012).

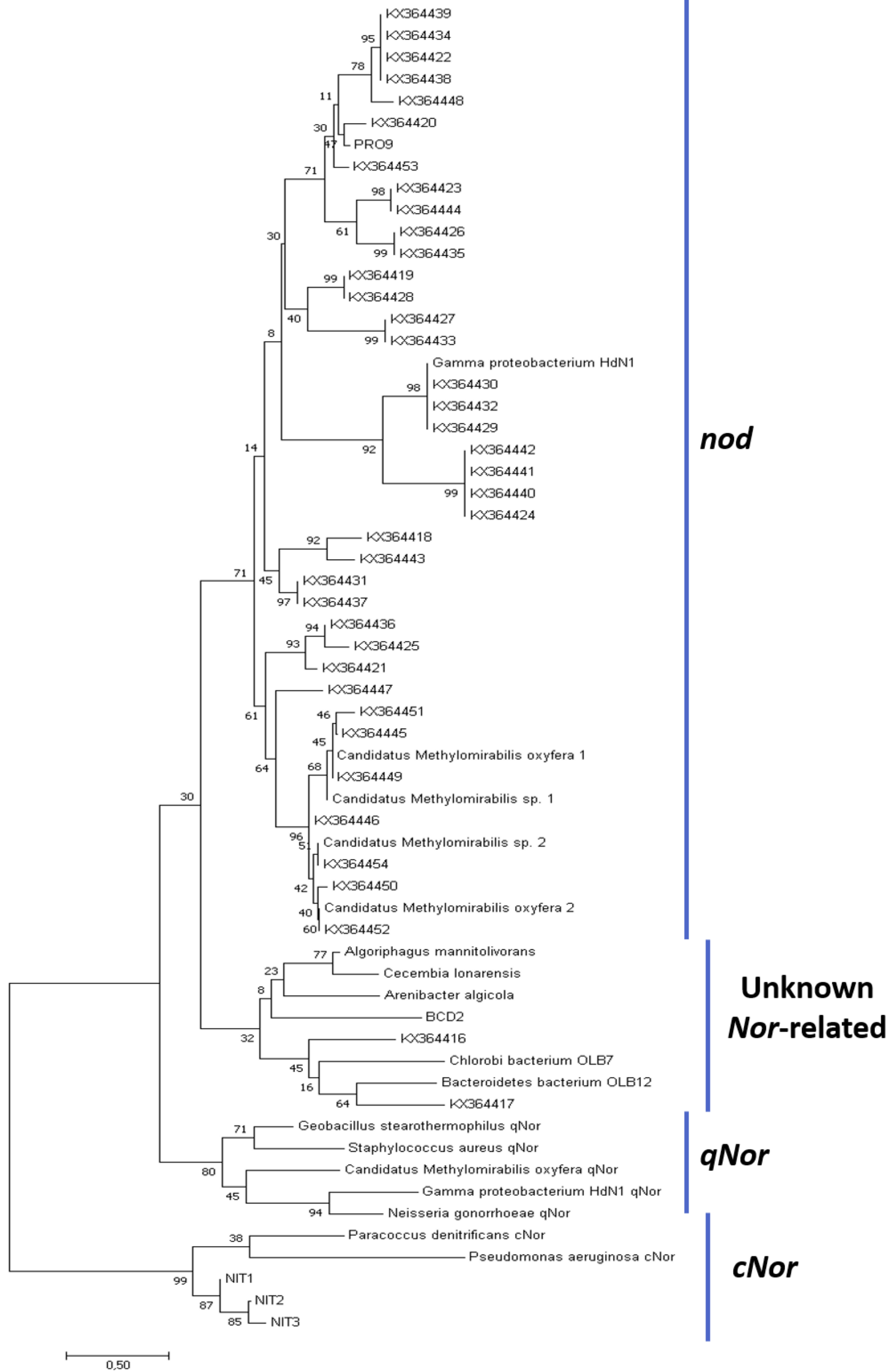

KX364439
KX364434
KX364422
KX364438
KX364448
KX364420
PRO9
KX364453
KX364423
KX364444
KX364426
KX364435
KX364419
KX364428
KX364427
KX364433
Gamma proteobacterium HdN1
KX364430
KX364432
KX364429
KX364442
KX364441
KX364440
KX364424
KX364418
KX364443
KX364431
KX364437
KX364436
KX364425
KX364421
KX364447
KX364451
KX364445
Candidatus Methylomirabilis oxyfera 1
KX364449
Candidatus Methylomirabilis sp. 1
KX364446
Candidatus Methylomirabilis sp. 2
KX364454
KX364450
Candidatus Methylomirabilis oxyfera 2
KX364452
Algoriphagus mannitolivorans
Cecembia lonarensis
Arenibacter algicola
BCD2
KX364416
Chlorobi bacterium OLB7
Bacteroidetes bacterium OLB12
KX364417
Geobacillus stearothermophilus qNor
Staphylococcus aureus qNor
Candidatus Methylomirabilis oxyfera qNor
Gamma proteobacterium HdN1 qNor
Neisseria gonorrhoeae qNor
Paracoccus denitrificans cNor
Pseudomonas aeruginosa cNor
NIT1
NIT2
NIT3
0,50
nod
Unknown
Nor-related
qNor
cNor

Figure 4 - Neighbour-joining phylogenetic tree of *nod,* unknown *Nor*-related*, qNor* and *cNor* using a bootstrap of 500 replicates. The reference genomes and their accession numbers were retrieved from the study performed by Zhu et al. (2017) (Zhu *et al.*, 2017). NIT proteins clustered with the cNor while the putative nitric-oxide dismutase retrieved in this study, belonged one to nod cluster and the other to the unknown Nor-related proteins.

Although the real function of the Nor-related genes is yet undefined, a recent study proposes a new cluster of *nod* or an atypical *qNor* role for this gene. Furthermore, having this gene, BCD2 also reduces N_2O (*nosZ*), potentially validating the study's theory of the role as atypical qNor or revealing the use of both to produce N_2 (Zhu *et al.*, 2017).

3.8. Organisms without nitrogen cycle genes

Out of the 57 MAGs, just nine genomes did not contain genes linked with the nitrogen cycle. Some of these genomes were not complete (less than 70% completeness) and one only had subunits of enzymatic complexes involved in this pathway. Three of the genomes that lacked respiratory nitrogen oxide genes were unclassified bacteria, DJK1, DJK2 and BCT8, of which DJK1 and DJK2 are similar to previously found genomes in recent studies of this type of wastewater treatment systems (Speth *et al.*, 2016).

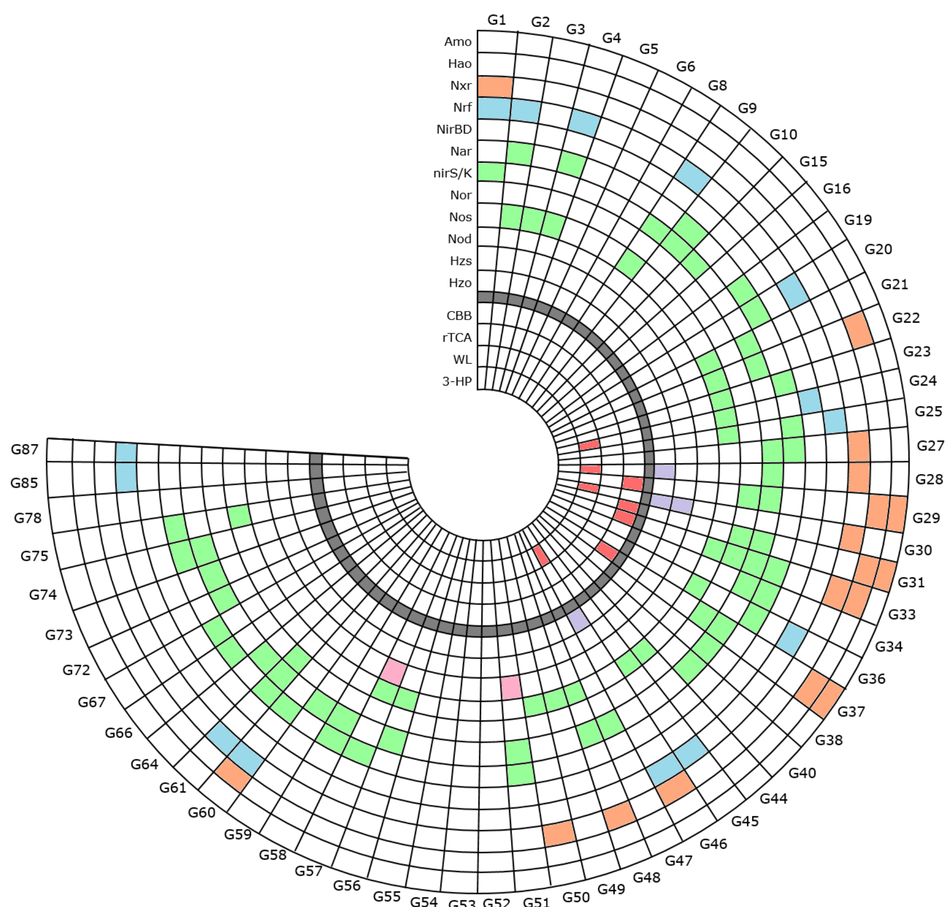


Figure 5 – Graphical overview of the genes encoded by each individual MAG, identified by their numerical ID (Supplementary Table 1). The pathways represented in the figure are the nitrogen cycle, Nitrification (orange), DNRA (blue), Denitrification (green), putative nitric-oxide dismutase (pink) and anammox (lilac), and carbon dioxide fixation, reductive pentose phosphate cycle, reductive tricarboxylic acid cycle, reductive acetyl-CoA pathway and 3-Hydroxylamine bicycle, respectively (red).

4. Conclusion

Applying two distinct binning approaches to the metagenomic analysis, we retrieved 57 MAGs, characterizing for most of the community. Also, we could assign a role in the nitrogen cycle to most of the organisms and describe de community's complexity with the discovery of the putative heterotrophic hydroxylamine oxidizers and putative heterotrophic nitrite oxidizers. This studied shows a diverse community and complex interactions surrounding the nitrogen cycle, specially due to the pathway partitioning (Figure 5). Further research is needed to specify the function of each individual organism in this community and fully characterize the metabolic interactions that binds them in this granular ecosystem.

5. References

Abeliovich, A. (2006) 'The Nitrite-Oxidizing Bacteria', *The Prokaryotes*, pp. 861–872. doi: 10.1007/0-387-30745-1.

Agrawal, S. *et al.* (2017) 'The role of inoculum and reactor configuration for microbial community composition and dynamics in mainstream partial nitritation anammox reactors', *MicrobiologyOpen*, (January), pp. 1–11. doi: 10.1002/mbo3.456.

Albertsen, M. *et al.* (2013) 'Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes.', *Nature biotechnology*, 31(6), pp. 533–8. doi: 10.1038/nbt.2579.

Anantharaman, K. *et al.* (2016) 'Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system', *Nature Communications*. Nature Publishing Group, 7, p. 13219. doi: 10.1038/ncomms13219.

Andrews, S. (2010) 'FastQC A Quality Control tool for High Throughput Sequence Data', *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Beaumont, H. J. E. *et al.* (2004) '*Nitrosomonas europaea* expresses a nitric oxide reductase during nitrification', *Journal of Bacteriology*, 186(13), p. 4417. doi: 10.1128/JB.186.13.4417.

Bhattacharjee, A. S. *et al.* (2017) 'Whole-Community Metagenomics in Two Different Anammox Configurations: Process Performance and Community Structure',

Environmental Science and Technology, 51(8), pp. 4317–4327. doi: 10.1021/acs.est.6b05855.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: A flexible trimmer for Illumina sequence data', *Bioinformatics*, 30(15), pp. 2114–2120. doi: 10.1093/bioinformatics/btu170.

Cabello, P. *et al.* (2004) 'Hydroxylamine assimilation by Rhodobacter capsulatus E1F1. Requirement of the hcp gene (hybrid cluster protein) located in the nitrate assimilation nas gene region for hydroxylamine reduction', *Journal of Biological Chemistry*, 279(44), pp. 45485–45494. doi: 10.1074/jbc.M404417200.

Castro-Barros, C. M. *et al.* (2017) 'Evaluating the potential for dissimilatory nitrate reduction by anammox bacteria for municipal wastewater treatment', *Bioresource Technology*, 233, pp. 363–372. doi: 10.1016/j.biortech.2017.02.063.

Chu, Z. rui *et al.* (2015) 'Microbial characterization of aggregates within a one-stage nitritation-anammox system using high-throughput amplicon sequencing', *Chemical Engineering Journal*. Elsevier B.V., 262, pp. 41–48. doi: 10.1016/j.cej.2014.09.067.

Costa, M. C. M. S. *et al.* (2014) 'Impact of inocula and operating conditions on the microbial community structure of two anammox reactors.', *Environmental technology*, 35(13–16), pp. 1811–22. doi: 10.1080/09593330.2014.883432.

Dietl, A. *et al.* (2015) 'The inner workings of the hydrazine synthase multiprotein complex', *Nature*, 527(7578), pp. 394–397. doi: 10.1038/nature15517.

Domingo-Félez, C. *et al.* (2014) 'Aeration strategies to mitigate nitrous oxide emissions from single-stage nitritation/anammox reactors', *Environmental Science and Technology*, 48(15), pp. 8679–8687. doi: 10.1021/es501819n.

Edgar, R. C. (2004) 'MUSCLE: Multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Research*, 32(5), pp. 1792–1797. doi: 10.1093/nar/gkh340.

Edgar, R. C. (2010) 'Search and clustering orders of magnitude faster than BLAST', *Bioinformatics*, 26(19), pp. 2460–2461. doi: 10.1093/bioinformatics/btq461.

Ettwig, K. F. *et al.* (2010) 'Nitrite-driven anaerobic methane oxidation by oxygenic bacteria.', *Nature*. Nature Publishing Group, 464(7288), pp. 543–548. doi: 10.1038/nature08883.

Ferousi, C. *et al.* (2013) 'Identification of the type II cytochrome c maturation pathway in anammox bacteria by comparative genomics', *BMC Microbiology*, 13(1), p. 265. doi: 10.1186/1471-2180-13-265.

Fux, C. *et al.* (2002) 'Biological treatment of ammonium-rich wastewater by partial nitritation and subsequent anaerobic ammonium oxidation (anammox) in a pilot plant', *Journal of Biotechnology*, 99(3), pp. 295–306. doi: 10.1016/S0168-1656(02)00220-1.

Fux, C. and Siegrist, H. (2004) ‘Nitrogen removal from sludge digester liquids by nitrification / denitrification or partial nitritation / anammox : environmental and economical considerations’, *Water Science & Technology*, 50(10), pp. 19–26. doi: 10.1111/j.1751-7915.2012.00341.x.

Galloway, J. (2005) ‘The global nitrogen cycle: past, present and future.’, *Science in China. Series C, Life sciences / Chinese Academy of Sciences*, 48 Suppl 2(326), pp. 669–678. doi: 10.1360/062005-261.

Gurevich, A. *et al.* (2013) ‘QUAST: Quality assessment tool for genome assemblies’, *Bioinformatics*, 29(8), pp. 1072–1075. doi: 10.1093/bioinformatics/btt086.

Hallin, S. *et al.* (2017) ‘Genomics and Ecology of Novel N2O-Reducing Microorganisms’, *Trends in Microbiology*, xx, pp. 1–13. doi: 10.1016/j.tim.2017.07.003.

Harhangi, H. R. *et al.* (2012) ‘Hydrazine synthase, a unique phylomarker with which to study the presence and biodiversity of anammox bacteria’, *Applied and Environmental Microbiology*, 78(3), pp. 752–758. doi: 10.1128/AEM.07113-11.

Hendriks, J. *et al.* (2000) ‘Nitric oxide reductases in bacteria’, *Biochimica et Biophysica Acta - Bioenergetics*, 1459(2–3), pp. 266–273. doi: 10.1016/S0005-2728(00)00161-4.

Hu, Z. *et al.* (2013) ‘Nitrogen removal by a nitritation-anammox bioreactor at low temperature’, *Applied and Environmental Microbiology*, 79(8), pp. 2807–2812. doi: 10.1128/AEM.03987-12.

Hug, L. A. *et al.* (2016) ‘Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages’, *Environmental Microbiology*, 18(1), pp. 159–173. doi: 10.1111/1462-2920.12930.

Hügler, M. and Sievert, S. M. (2011) ‘Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean’, *Annual Review of Marine Science*, 3(1), pp. 261–289. doi: 10.1146/annurev-marine-120709-142712.

Huson, D. *et al.* (2007) ‘MEGAN analysis of metagenome data’, *Gennome Res.*, 17, pp. 377–386. doi: 10.1101/gr.5969107.

Hyatt, D. *et al.* (2010) ‘Prodigal: prokaryotic gene recognition and translation initiation site identification’, *BMC Bioinformatics*, 11(1), p. 119. doi: 10.1186/1471-2105-11-119.

Isaacs, S. H. and Henze, M. (1995) ‘Controlled carbon source addition to an alternating nitrification-denitrification wastewater treatment process including biological P removal’, *Water Research*, 29(1), pp. 77–89. doi: 10.1016/0043-1354(94)E0119-Q.

Joss, A. *et al.* (2011) ‘Combined Nitritation - Anammox : Advances in Understanding Process Stability’, *Environmental Science & Technology*, 45, pp. 9735–9742. doi: 10.1021/es202181v.

Kampschreur, M. J. *et al.* (2009) ‘Nitrous oxide emission during wastewater treatment’, *Water Research*. Elsevier Ltd, 43(17), pp. 4093–4103. doi:

10.1016/j.watres.2009.03.001.

Kantor, R. S. *et al.* (2015) 'Bioreactor microbial ecosystems for thiocyanate and cyanide degradation unravelled with genome-resolved metagenomics', *Environmental Microbiology*, 17(12), pp. 4929–4941. doi: 10.1111/1462-2920.12936.

Kartal, B. *et al.* (2011) 'Molecular mechanism of anaerobic ammonium oxidation', *Nature*. Nature Publishing Group, 479(7371), pp. 127–130. doi: 10.1038/nature10453.

Kartal, B. *et al.* (2013) 'How to make a living from anaerobic ammonium oxidation', *FEMS Microbiology Reviews*, 37(3), pp. 428–461. doi: 10.1111/1574-6976.12014.

Kartal, B., Kuenen, J. G. and van Loosdrecht, M. C. M. (2010) 'Sewage Treatment with Anammox', *Science*, 328(5979), pp. 702–703. doi: 10.1126/science.1185941.

Keegan, K. P., Glass, E. M. and Meyer, F. (2016) 'MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function', in *Methods in molecular biology (Clifton, N.J.)*, pp. 207–233. doi: 10.1007/978-1-4939-3369-3_13.

Kneese, A. V. (2015) *Water pollution : economic aspects and research needs*. Available at: http://findit.dtu.dk/resolve?sid=google&auinit=AV&aulast=Kneese&title=Water+Pollution:+Economics+Aspects+and+Research+Needs&genre=book&isbn=1317387562&date=2015 (Accessed: 13 July 2017).

Kong, L. *et al.* (2013) 'Diversity and spatial distribution of hydrazine oxidoreductase (hzo) gene in the oxygen minimum zone off Costa Rica', *PLoS ONE*, 8(10), pp. 1–11. doi: 10.1371/journal.pone.0078275.

Kozlowski, J. A., Dimitri Kits, K. and Stein, L. Y. (2016) 'Comparison of nitrogen oxide metabolism among diverse ammonia-oxidizing bacteria', *Frontiers in Microbiology*, 7(JUL), pp. 1–9. doi: 10.3389/fmicb.2016.01090.

Kuenen, J. G. (2008) 'Anammox bacteria: from discovery to application', *Nature Reviews Microbiology*, 6(4), pp. 320–326. doi: 10.1038/nrmicro1857.

Kumar, S., Stecher, G. and Tamura, K. (2016) 'MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets', *Molecular biology and evolution*, 33(7), pp. 1870–1874. doi: 10.1093/molbev/msw054.

Lackner, S. *et al.* (2014) 'Full-scale partial nitritation/anammox experiences - An application survey', *Water Research*. Elsevier Ltd, 55(0), pp. 292–303. doi: 10.1016/j.watres.2014.02.032.

Laczny, C. C. *et al.* (2015) 'VizBin - an application for reference-independent visualization and human-augmented binning of metagenomic data', *Microbiome*, 3(1), p. 1. doi: 10.1186/s40168-014-0066-1.

Langone, M. *et al.* (2014) 'Coexistence of nitrifying, anammox and denitrifying bacteria in a sequencing batch reactor', *Frontiers in Microbiology*, 5(FEB), pp. 1–12. doi:

10.3389/fmicb.2014.00028.

Lawson, C. E. *et al.* (2017) 'Metabolic network analysis reveals microbial community interactions in anammox granules', *Nature Communications*. Nature Publishing Group, 8(May), p. 15416. doi: 10.1038/ncomms15416.

Luesken, F. A. *et al.* (2011) 'Diversity and enrichment of nitrite-dependent anaerobic methane oxidizing bacteria from wastewater sludge', *Applied Microbiology and Biotechnology*, 92(4), pp. 845–854. doi: 10.1007/s00253-011-3361-9.

Luo, C., Rodriguez-R, L. M. and Konstantinidis, K. T. (2014) 'MyTaxa: An advanced taxonomic classifier for genomic and metagenomic sequences', *Nucleic Acids Research*, 42(8), pp. 1–12. doi: 10.1093/nar/gku169.

Lycus, P. *et al.* (2017) 'Phenotypic and genotypic richness of denitrifiers revealed by a novel isolation strategy', *The ISME Journal*. Nature Publishing Group, 11(10), pp. 2219–2232. doi: 10.1038/ismej.2017.82.

Maalcke, W. J. *et al.* (2014) 'Structural basis of biological no generation by octaheme oxidoreductases', *Journal of Biological Chemistry*, 289(3), pp. 1228–1242. doi: 10.1074/jbc.M113.525147.

Matsuzaka, E. *et al.* (2003) 'Participation of Nitrite Reductase in Conversion of NO2- to NO3 - in a Heterotrophic Nitrifier, Burkholderia cepacia NH-17, with Denitrification Activity', *Microbes and Environments*, 18(4), pp. 203–209. doi: 10.1264/jsme2.18.203.

Menzel, P., Ng, K. L. and Krogh, A. (2016) 'Fast and sensitive taxonomic classification for metagenomics with Kaiju', *Nature Communications*. Nature Publishing Group, 7, p. 11257. doi: 10.1038/ncomms11257.

Oshiki, M. *et al.* (2015) 'Draft Genome Sequence of an Anaerobic Ammonium-Oxidizing Bacterium,"Candidatus Brocadia sinica"', *Genome Announcement*, 3(2), pp. 3–4. doi: 10.1128/genomeA.00267-15.Copyright.

Oshiki, M. *et al.* (2016) 'Hydroxylamine-dependent anaerobic ammonium oxidation (anammox) by ???Candidatus Brocadia sinica???', *Environmental Microbiology*, 18(9), pp. 3133–3143. doi: 10.1111/1462-2920.13355.

Padhi, S. K. *et al.* (2017) 'Aerobic and heterotrophic nitrogen removal by Enterobacter cloacae CF-S27 with efficient utilization of hydroxylamine', *Bioresource Technology*. Elsevier Ltd, 232, pp. 285–296. doi: 10.1016/j.biortech.2017.02.049.

Peng, Y. *et al.* (2012) 'IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth', *Bioinformatics*, 28(11), pp. 1420–1428. doi: 10.1093/bioinformatics/bts174.

Pynaert, K. *et al.* (2003) 'Characterization of an autotrophic nitrogen-removing biofilm from a highly loaded lab-scale rotating biological contactor', *Applied and Environmental Microbiology*, 69(6), pp. 3626–3635. doi: 10.1128/AEM.69.6.3626-3635.2003.

Ragsdale, S. W. and Pierce, E. (2009) 'Acetogenesis and the Wood-Ljungdahl Pathway of CO2 Fixation', *Biochim Biophys Acta*, 1784(12), pp. 1873–1898. doi: 10.1016/j.bbapap.2008.08.012.Acetogenesis.

Roco, C. A. *et al.* (2017) 'Modularity of nitrogen-oxide reducing soil bacteria: linking phenotype to genotype', *Environmental Microbiology*, 19(6), pp. 2507–2519. doi: 10.1111/1462-2920.13250.

Rodriguez-r, L. M. and Konstantinidis, K. T. (2014) 'Bypassing Cultivation to Identify Bacterial Species', *Microbe*, 9(3), pp. 111–118.

Sakai, K. *et al.* (1996) 'Nitrite oxidation by heterotrophic bacteria under various nutritional and aerobic conditions', *Journal of Fermentation and Bioengineering*, 82(6), pp. 613–617. doi: 10.1016/S0922-338X(97)81265-2.

Sakai, K. *et al.* (1997) 'Change in nitrite conversion direction from oxidation to reduction in heterotrophic bacteria depending on the aeration conditions', *Journal of Fermentation and Bioengineering*, 84(1), pp. 47–52. doi: 10.1016/S0922-338X(97)82785-7.

Sanford, R. A. *et al.* (2012) 'Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils', *Proceedings of the National Academy of Sciences*, 109(48), pp. 19709–19714. doi: 10.1073/pnas.1211238109.

Schmidt, I. *et al.* (2003) 'New concepts of microbial treatment processes for the nitrogen removal in wastewater', *FEMS Microbiology Reviews*, 27(4), pp. 481–492. doi: 10.1016/S0168-6445(03)00039-1.

Schmidt, I., van Spanning, R. J. M. and Jetten, M. S. M. (2004) 'Denitrification and ammonia oxidation by Nitrosomonas europaea wild-type, and NirK- and NorB-deficient mutants', *Microbiology*, 150(12), pp. 4107–4114. doi: 10.1099/mic.0.27382-0.

Shehzad, A. *et al.* (2016) 'Diversity, Community Composition and Abundance of Anammox Bacteria in Sediments of the North Marginal Seas of China', *Microbes and Environments*, 31(2), pp. 111–120. doi: 10.1264/jsme2.ME15140.

Shrestha, N. K. *et al.* (2002) 'Dinitrogen production from ammonia by Nitrosomonas europaea', *Applied Catalysis A: General*, 237(1–2), pp. 33–39. doi: 10.1016/S0926-860X(02)00279-X.

Siegrist, H. *et al.* (2008) 'Anammox brings WWTP closer to energy autarky due to increased biogas production and reduced aeration energy for N-removal', *Water Science & Technology*, 57(3), p. 383. doi: 10.2166/wst.2008.048.

Sobieszuk, P. and Szewczyk, K. W. (2006) 'Estimation of (C/N) ratio for microbial denitrification', *Environmental Technology*, 27(1), pp. 103–108. doi: 10.1080/09593332708618624.

Socolow, R. H. (1999) 'Nitrogen management and the future of food: lessons from the management of energy and carbon.', *Proceedings of the National Academy of Sciences*

of the United States of America, 96(11), pp. 6001–6008. doi: 10.1073/pnas.96.11.6001.

Speth, D. R. *et al.* (2016) 'Genome-based microbial ecology of anammox granules in a full-scale wastewater treatment system', *Nature Communications*, 7, p. 11172. doi: 10.1038/ncomms11172.

Stein, L. Y. *et al.* (2007) 'Whole-genome analysis of the ammonia-oxidizing bacterium, Nitrosomonas eutropha C91: Implications for niche adaptation', *Environmental Microbiology*, 9(12), pp. 2993–3007. doi: 10.1111/j.1462-2920.2007.01409.x.

Strous, M. *et al.* (1998) 'The sequencing batch reactor as a powerful tool for the study of slowly growing anaerobic ammonium-wxidizing microorganisms', *Appl Micorbial. biotechnol*, 50, pp. 589–596. doi: 10.1007/s002530051340.

Strous, M. *et al.* (2006) 'Deciphering the evolution and metabolism of an anammox bacterium from a community genome', *Nature*, 440(7085), pp. 790–794. doi: 10.1038/nature04647.

Suddick, E. C. *et al.* (2013) 'The role of nitrogen in climate change and the impacts of nitrogen-climate interactions in the United States: Foreword to thematic issue', *Biogeochemistry*, 114(1–3), pp. 1–10. doi: 10.1007/s10533-012-9795-z.

Sutton, M. A. *et al.* (2011) 'Too much of a good thing', *Nature*, 472(7342), pp. 159–161. doi: 10.1038/472159a.

Tsuneda, S. *et al.* (2005) 'Effect of salinity on nitrous oxide emission in the biological nitrogen removal process for industrial wastewater', *Journal of Hazardous Materials*, 119(1–3), pp. 93–98. doi: 10.1016/j.jhazmat.2004.10.025.

Volcke, E. I. P., Vanrolleghem, P. and van Loosdrecht, M. C. M. (2006) *Modelling, analysis and control of partial nitritation in a SHARON reactor*, *Faculty of Agricultural and Applied Biological Sciences*.

Wang, Y. *et al.* (2012) 'Co-occurrence and distribution of nitrite-dependent anaerobic ammonium and methane-oxidizing bacteria in a paddy soil', *FEMS Microbiology Letters*, 336(2), pp. 79–88. doi: 10.1111/j.1574-6968.2012.02654.x.

Whitacre, D. M. (2013) *Reviews of Environmental Contamination and Toxicology*. doi: 10.1007/978-1-4614-6470-9.

Woznica, A. *et al.* (2013) 'Stimulatory Effect of Xenobiotics on Oxidative Electron Transport of Chemolithotrophic Nitrifying Bacteria Used as Biosensing Element', *PLoS ONE*, 8(1). doi: 10.1371/journal.pone.0053484.

Wu, S. *et al.* (2011) 'WebMGA: a customizable web server for fast metagenomic sequence analysis', *BMC Genomics*, 12(1), p. 444. doi: 10.1186/1471-2164-12-444.

Yin, H. and Kong, M. (2014) 'Simultaneous removal of ammonium and phosphate from eutrophic waters using natural calcium-rich attapulgite-based versatile adsorbent', *Desalination*. Elsevier B.V., 351(August), pp. 128–137. doi:

10.1016/j.desal.2014.07.029.

Zhao, B. *et al.* (2012) 'N 2O and N 2 production during heterotrophic nitrification by Alcaligenes faecalis strain NR', *Bioresource Technology*. Elsevier Ltd, 116, pp. 379–385. doi: 10.1016/j.biortech.2012.03.113.

Zhao, B. *et al.* (2017) 'Characteristics of a heterotrophic nitrogen removal bacterium and its potential application on treatment of ammonium-rich wastewater', *Bioresource Technology*. Elsevier Ltd, 226, pp. 46–54. doi: 10.1016/j.biortech.2016.11.120.

Zhou, W. *et al.* (2014) 'Bias from removing read duplication in ultra-deep sequencing experiments', *Bioinformatics*, 30(8), pp. 1073–1080. doi: 10.1093/bioinformatics/btt771.

Zhu, B. *et al.* (2012) 'Anaerobic oxidization of methane in a minerotrophic peatland: Enrichment of nitrite-dependent methane-oxidizing bacteria', *Applied and Environmental Microbiology*, 78(24), pp. 8657–8665. doi: 10.1128/AEM.02102-12.

Zhu, B. *et al.* (2017) 'Unexpected diversity and high abundance of putative nitric oxide dismutase (Nod) genes in contaminated aquifers and wastewater treatment systems', *Applied and Environmental Microbiology*, 83(4). doi: 10.1128/AEM.02750-16.

Supplementary Information

Supplementary Method 1 – Binning Methods

In order to recover the majority of the community population, two binning approaches were used to recover draft genomes. The first approach, VizBin, clusters CGs using pentanucleotide frequency, yielding high quality CGs (Figure SM1-1). The second approach uses mmgenome, a R package that applies differential coverage analysis as a binning strategy. The samples were independently mapped against each other, separating the reads by coverage. Then, information of the GC content, tetranucleotide frequency and taxonomic classification based on the 107 single-copy maker genes was used to further separate the reads into CGs and estimate the quality. The CGs were manually recovered following the step-by-step guide (http://madsalbertsen.github.io/multi-metagenome/docs/overview.html) (Figure SM1-2).

To recover low abundance genomes, reads from sample 1-4 and reads from sample 5-7 were coassembled and the were binned using the same strategy described previously (Figure SM1-3). Approximately 40% of the total number of MAGs recovered in this work were obtained from the application of the coassembly strategy.

Concluding, the application of two binning methods allowed us to retrieve a high number of genomes representative of the community. VizBin obtained CGs are almost complete but inability of separating similar species forms composite genomes. These composite genomes can be further separated using mmgenome, with a slightly inferior completeness but a tremendous decrease in contamination (Figure SM1-4).

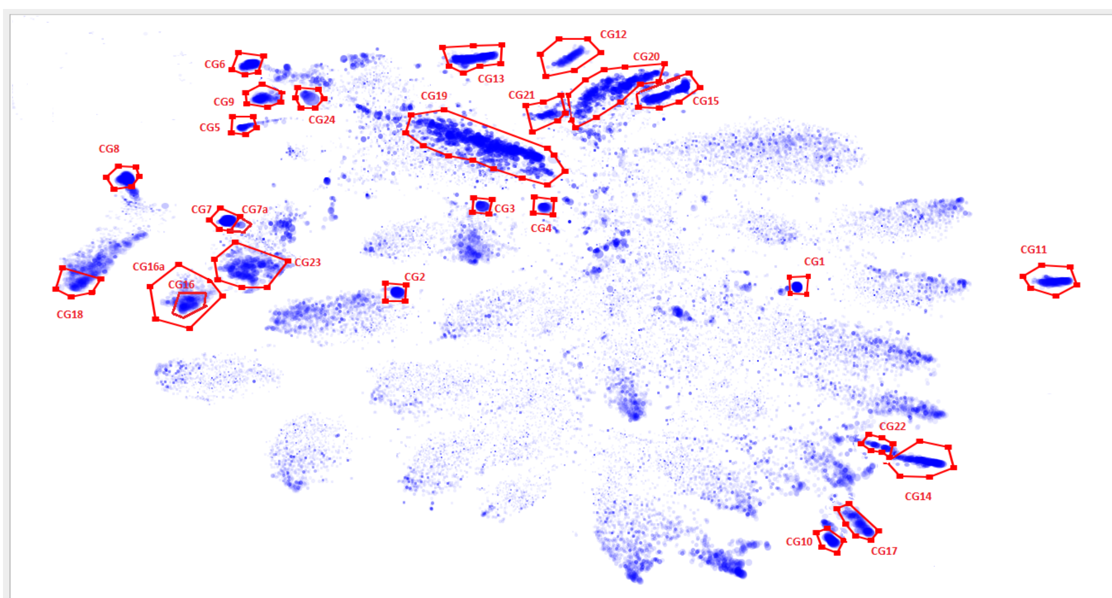


Figure SM1-1 – Scatter plot of contigs (≥ 1000-bp) assembled from metagenomic reads from the Sample 2. Contigs are represented by individual points, with point size representing contig length. Each geometric shape is a CG, with cluster similar to CG19 showing composite genomes.

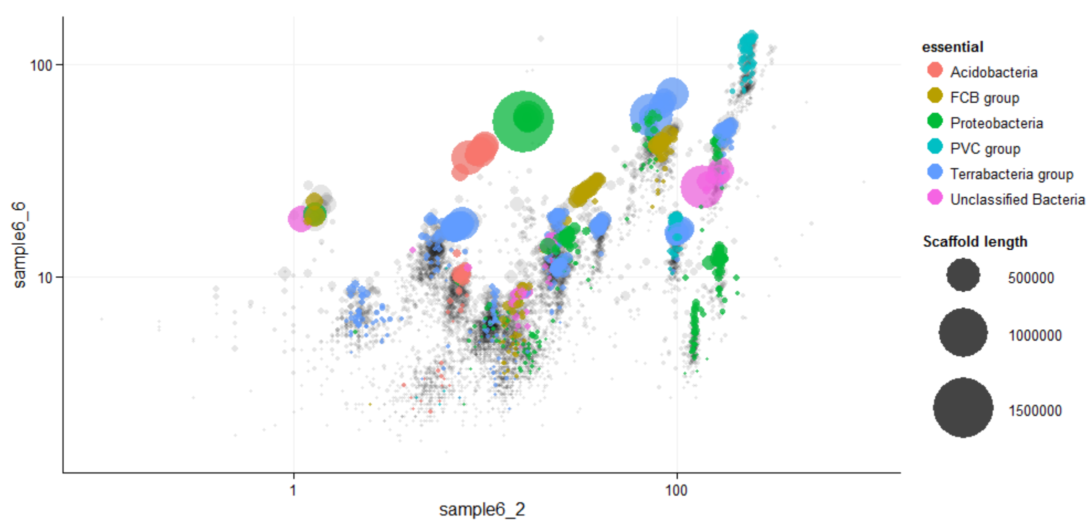


Figure SM1-2 – Differential coverage plot of Sample 6, comparing subsample 6_2 and subsample 6_6. Scaffold are represented as circles and coloured by the correspondent phylum or group taxonomic classification.

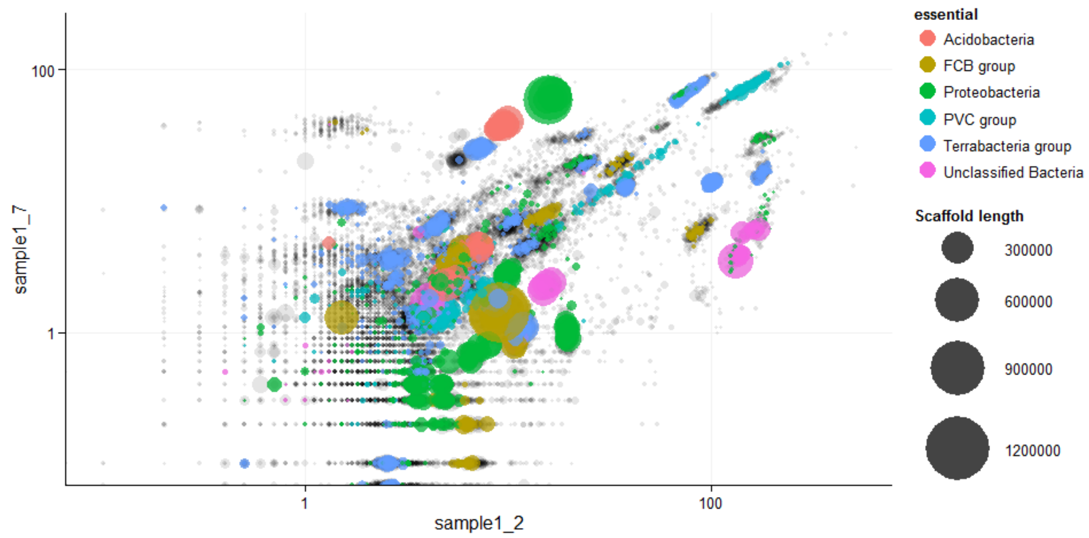


Figure SM1-2 – Differential coverage plot of coassembled Sample1, composed of sample 1 to 4, comparing subsample 1_2 and subsample 1_7. Scaffold are represented as circles and coloured by the correspondent phylum or group taxonomic classification.

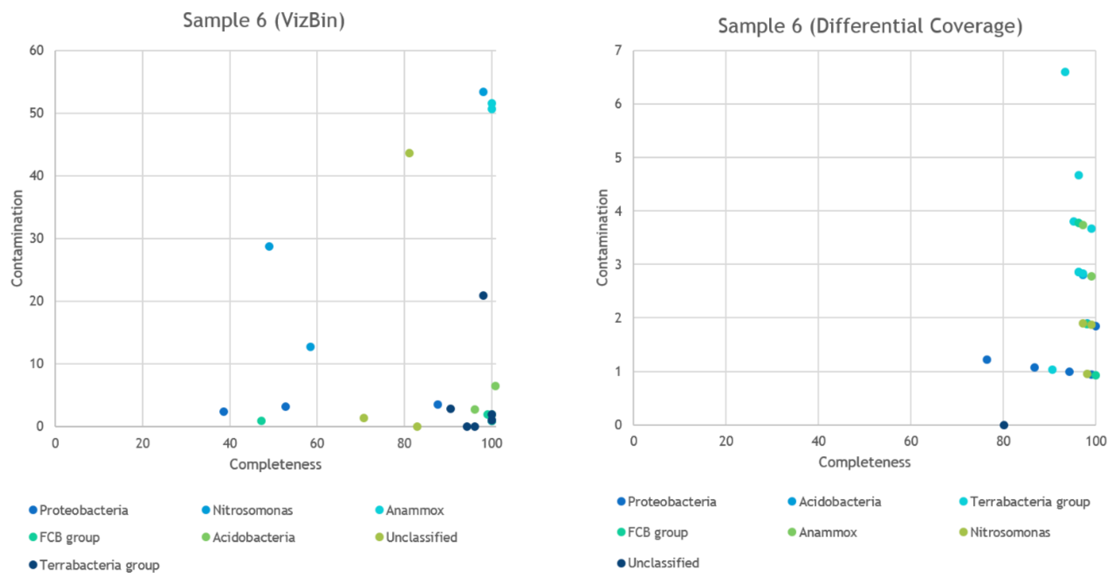


Figure SM1-2 – Comparison of the quality of the CGs recovered by the two binning methods. CGs retrieved using VizBin have high completeness and low contamination, with composite genomes holding elevated contamination. Through mmgenome, the percentage of contamination decreased significantly.

Supplementary Method 2 – Binning Problems

During genomic annotation of the genomes, some errors associated with the binning were detected. Therefore, some additional information related to the detection and problem-solving are described hereafter.

Unexpectedly, the AMO operon was solely detected in the NIT3 and CLB1 genomes, with CLB1 containing two copies. Whole metagenome examination showed the presence of only three operons, implication the absence of the genes in NIT1 and NIT2. Moreover, phylogenetic analysis suggest proximity between the three copies to the AOB cluster. So, the presence of these genes in the CLB1 organisms was a result of horizontal gene transfer or an error.

Initially, the three genes and CLB1, NIT1, NIT2 and NIT3 were compared based on their sequencing depth and GC content, producing two distinct boxplots (Figure SM2-1). The results imply that the AMO genes were not originally from CLB1, and a similarity between in boxplot of the genes and NIT genomes is detected. But, the genes values are several times higher than the genomes, so no possibility can be eliminated.

Thus, the binning of the metagenome by differential coverage was redone, but highlighting the interest genes instead. Results allowed to eliminate the possibility of transfer by HGT, since the interested genes were located in close proximity to the NIT genomes (Figure SM2-2).

NIT genomes also lacked the HAO genes. Phylogenetic analysis indicated the presence of three *haoA* genes, hence the same method previous described was applied. Once again, the genes were found near NIT genomes (Figure SM2-2).

The proximity proves that the NIT genomes were the original holders of this genes although its depth and coverage are a lot higher, when compared. Such difference can be explained by the number of copies of the AMO and HAO operon, around 2 or more, in AOBs from the Nitrosomonas genus (Kozlowski, Dimitri Kits and Stein, 2016). During de novo assemble the number of reads for a gene can increase when there is a duplication. As a result, the gene will present higher coverage and depth compared to the original genomes, generating a binning error (Zhou *et al.*, 2014).

For the anammox genomes, both HZS and *hzo* genes were missing, a comparable situation as described before. Boxplot results showed the same results, due to number of copies, leading to a high sequencing depth (Figure SM2-3)(Strous *et al.*, 2006; Kong *et al.*, 2013; Shehzad *et al.*, 2016). Three phylogenetic trees composed of AMX and representative anammox bacteria genomes were constructed for three different genes, *hzo*, *nxrB* and *amtB* (Figure SM2-4). Both AMX1 and AMX2 cluster with *Candidatus* Brocadia fulgida, with AMX1 closely related to this species. So, when extrapolating this results to *hzo* gene, we can appoint the gene CLB1_2 to AMX1 and CLB1_1 to AMX2.

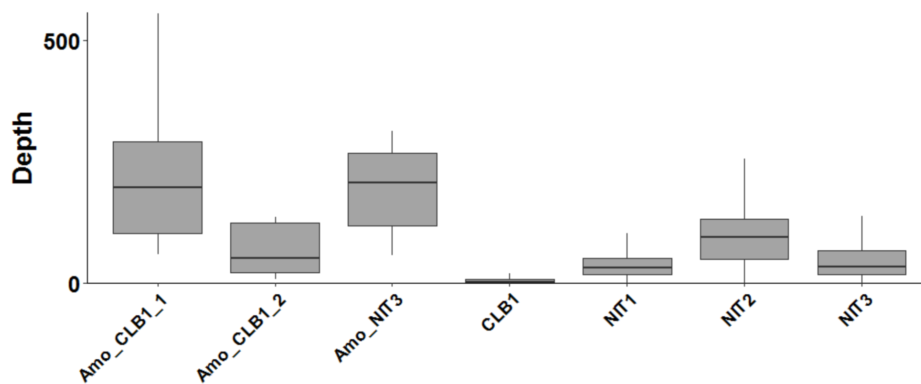


Figure SM2-1 – Comparison of depth between the *Nitrosomonas* genomes, the CLB1 genome and the *amo* genes.

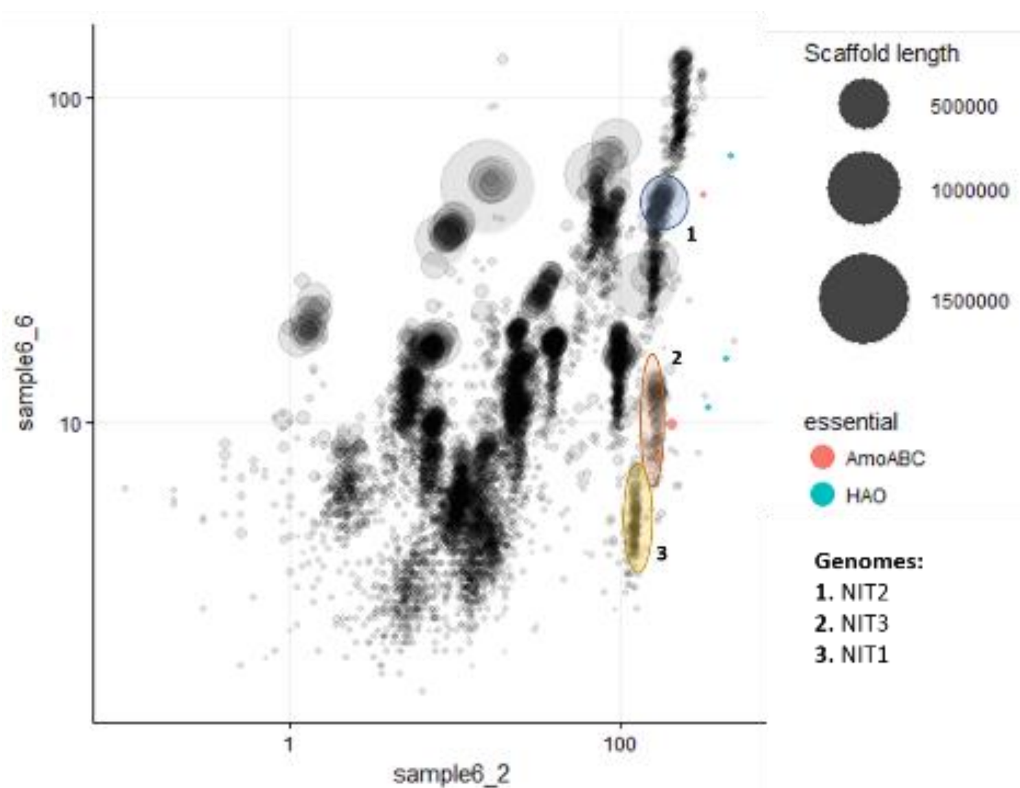


Figure SM2-2 – Differential coverage plot of Sample6, comparing subsample 6_2 and subsample 6_6. *amo* and *hao* genes are highlighted and the Nitrosomonas genomes position identified.

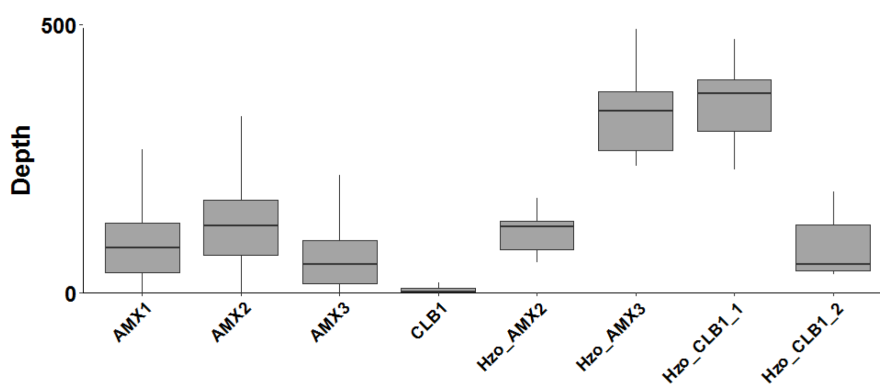


Figure SM2-3 – Comparison of depth between the Anammox genomes, the CLB1 genome and the *hzo* genes.

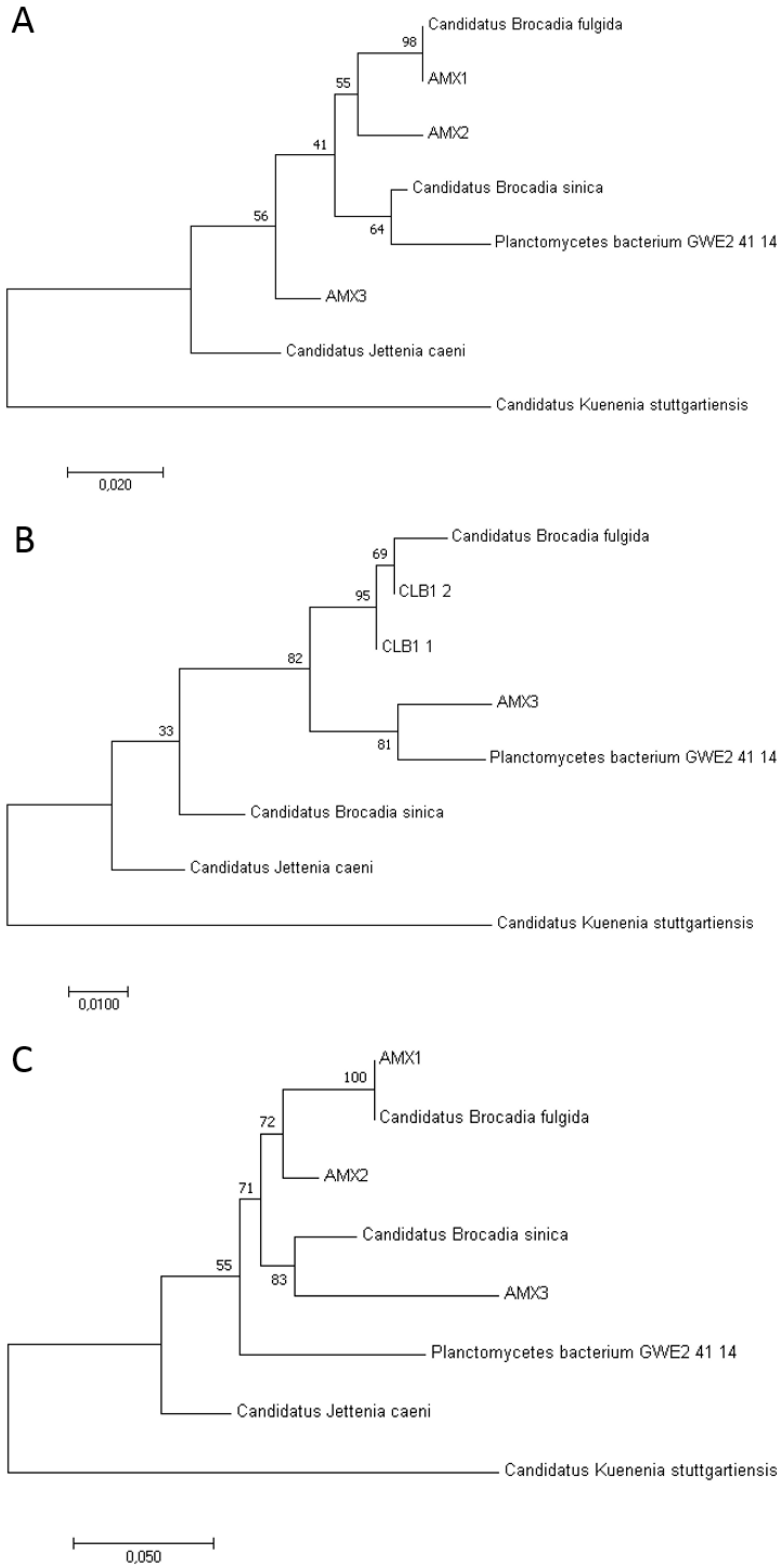


Figure SM2-4 – Neighbour-joining phylogenetic tree of Anammox genomes, retrieved in this work and reference genomes, of (A) *hzo* proteins, (B) nxrB proteins and (C) *amtB* proteins.

Supplementary Method 3 – Verification of the nitrogen cycle results

Functional annotation of each individual genome resulted in unexpected genotypes, like the internal deletion niche of the denitrification, the putative NOB, the putative heterotrophic hydroxylamine oxidizers and the putative heterotrophic nitrite oxidizers.

After the confirmation of each protein annotation using phylogeny with reference genomes, a pipeline for an extra confirmation was constructed to rule out the possibility of errors similar to the ones described in the supplementary method 2 (Figure SM3-1).

The following list characterizes the expected genotypes in this system:

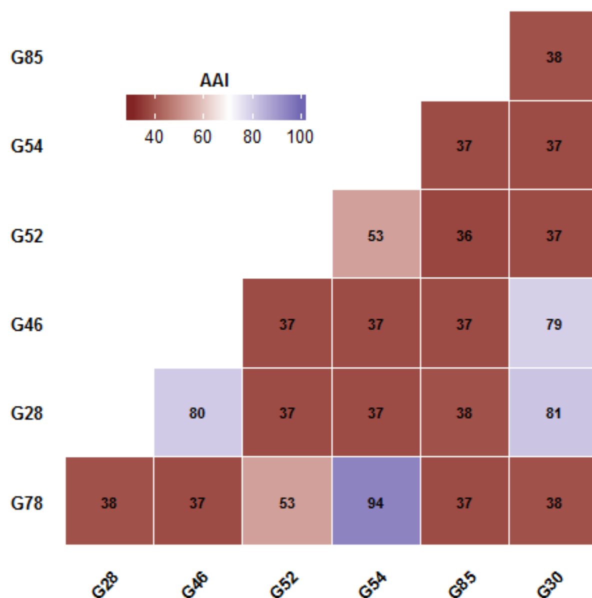
- AOBs – The presence of the *amo* and *hao* operon and CBB carbon fixation pathway are obligatory, while the presence of denitrification genes from the reduction of nitrite to nitrogen can be expected (Shrestha *et al.*, 2002).
- NOBs – The presence of the *nxr* operon and the rTCA or CBB carbon fixation pathway.
- Anammox Bacteria – These bacteria normally encode the *hzs*, *hzo*, *nxr* and *nirS/K* or a putative *hao*-like protein.
- Denitrifiers – Denitrifiers are expected to have *nar*, *nir*, *nor* and *nos* operon. Several organisms performing partial denitrification have been described in other works, so it is only considered an unexpected genotype if the absence occurs in the middle of the pathway. Such genotypes are for example, containing *nar* and *nos* and missing *nir* and nor and containing *nar* and *nor* and missing *nir*.
- DNRA – The *nar* or *nap* and *nrf* operon encode for the DNRA pathway.

To all genomes holding unexpected genotype the pipeline is applied and verification of the complete operon is performed in MG-RAST (Supplementary Table 2).

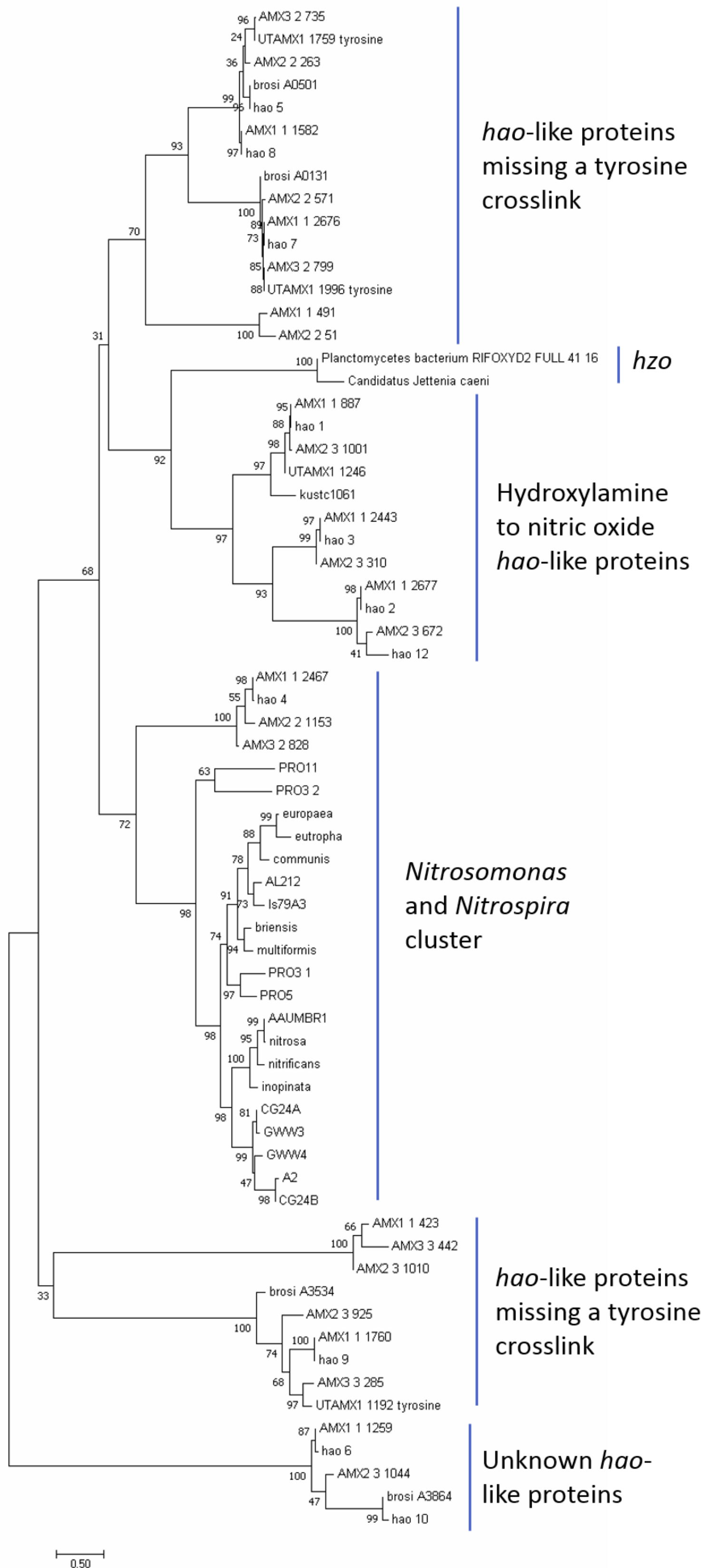
Figure SM3-1 – Schematic diagram of the pipeline created to confirm the unexpected genotypes obtained during functional annotation.

Supplementary Figures

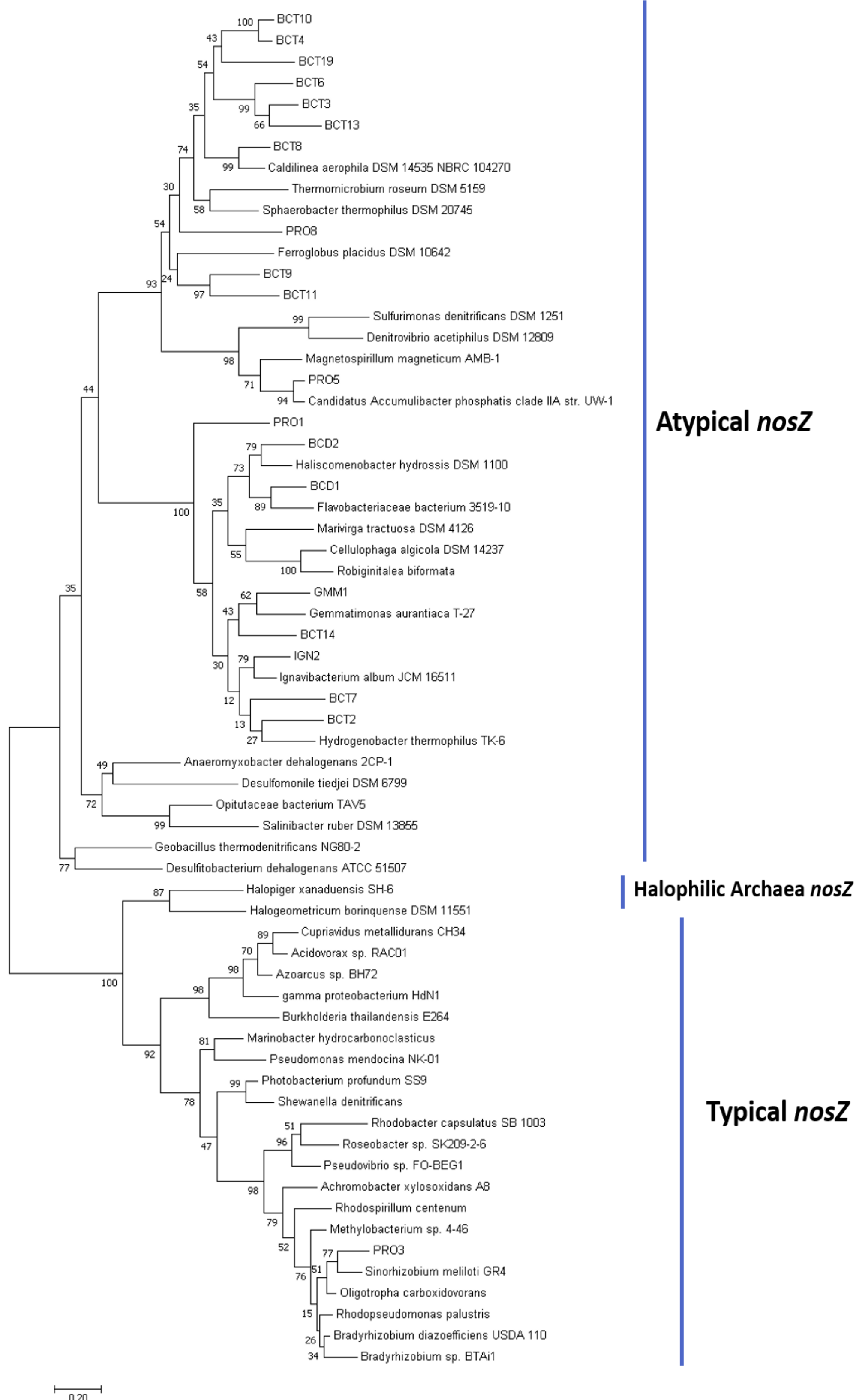
Supplementary Figure 1 – Comparison of anammox bacteria (right) and AOB (left) genomes to reference genomes previously published, based on the amino acid identity. Identification of the genomes retrieved in the work is as followed, NIT1 (Genome 29), NIT2 (Genome 31), NIT3 (Genome 37), AMX1 (Genome 28), AMX2 (Genome 30) and AMX3 (Genome 46).



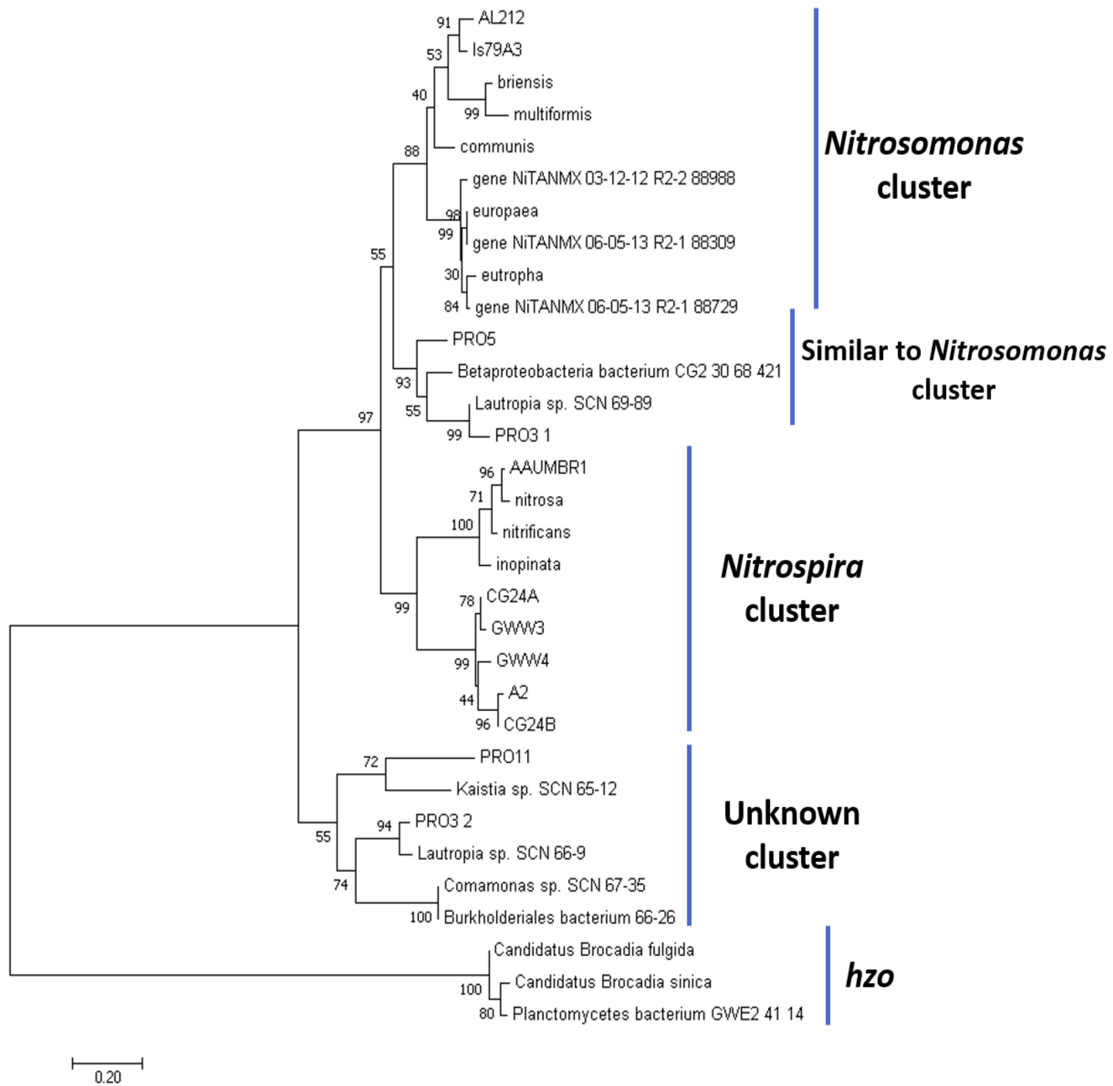
Supplementary Figure 2 – Comparison of genomes from the PVC group based on amino acid identity. Identification of the genomes retrieved in the work is as followed, AMX1 (G28), AMX2 (G30) and AMX3 (G46).



Supplementary Figure 3 – Phylogenetic tree of *hao*-like proteins.



Supplementary Figure 4 – Phylogenetic tree of the *nosZ* proteins.



Supplementary Figure 5 – Phylogenetic tree of the *haoA* proteins.

Supplementary Tables

Supplementary Table 1 – Characteristics of the MAGs reconstructed from the PNA system.

Bin ID	no ID	contigs	GC(%)	N50	Total length(bp)	Completeness(%)	Contamination(%)	Taxonomy
ACD1	16	80	54,2	164 998	3 661 027	100,0	2,8	Bacteria;Acidobacteria
ACD2	20	173	63,4	54 924	5 281 103	99,1	5,4	Bacteria;Acidobacteria
ACD3	58	185	54,3	173 466	3 162 093	95,3	4,7	Bacteria;Acidobacteria
ACT1	59	263	73,5	9 314	1 674 977	45,3	0,0	Bacteria;Actinobacteria
ACT2	66	113	72,7	17 450	1 206 875	38,7	0,0	Bacteria;Actinobacteria;Thermoleophilia; Solirubrobacterales;
AMX1	28	10	45	3 250 501	3 367 835	99,1	2,8	Bacteria;Planctomycetes;Planctomycetia; Candidatus Brocadiales; Candidatus Brocadiaceae;Candidatus Brocadia;
AMX2	30	121	43,6	2 284 717	3 754 724	97,2	3,7	Bacteria;Planctomycetes;Planctomycetia; Candidatus Brocadiales;Candidatus Brocadiaceae;
AMX3	46	118	42,2	832 282	2 474 235	90,6	5,0	Bacteria;Planctomycetes;Planctomycetia; Candidatus Brocadiales;Candidatus Brocadiaceae;
ARM1	1	42	60,9	142 842	2 870 527	94,3	1,0	Bacteria; Armatimonadetes;Fimbriimonadia
BCD1	3	107	41,4	55 265	3 530 112	100,0	1,9	Bacteria;Bacteroidetes
BCD2	57	53	66,9	98 468	3 204 765	87,7	1,1	Bacteria;Bacteroidetes
BCG1	23	33	37,7	178 301	2 387 745	100,0	0,9	Bacteria;Bacteroidetes/Chlorobi group
BCT1	50	93	66,3	76 467	2 858 101	88,7	2,1	Bacteria;
BCT2	21	1001	73	7 843	5 200 486	100,0	1,9	Bacteria;
BCT3	44	76	69,3	133 983	3 018 307	96,2	1,0	Bacteria;
BCT4	53	13	39,1	601 786	864 966	66,0	1,4	Bacteria
BCT5	56	191	57	28 215	3 684 708	87,7	0,0	Bacteria;
BCT6	64	204	72,1	14 670	1 821 896	39,6	0,0	Bacteria;
BCT7	67	418	64,5	4 777	1 596 158	38,7	0,0	Bacteria;
BCT8	8	35	47,5	185 700	722 001	81,1	0,0	Bacteria
BCT9	72	43	53,6	71 041	1 560 850	0,9	0,0	Bacteria
BCT10	85	126	63	88 537	2 314 385	74,5	1,3	Bacteria;
BCT11	36	1109	64,7	6 183	5 183 121	96,2	4,7	Bacteria;
CFX1	73	185	52,9	7 454	789 822	4,7	0,0	Bacteria;Chloroflexi;Anaerolineae; Anaerolineales;Anaerolineaceae;

CFX2	**10**	**105**	**52,7**	**58 196**	**2 722 981**	**96,2**	**2,9**	**Bacteria;Chloroflexi**
CFX3	34	362	53,1	20 077	4 167 777	96,2	2,9	Bacteria;Chloroflexi
CFX4	38	20	61,2	419 728	2 328 137	97,2	2,8	Bacteria;Chloroflexi
CFX5	47	651	57,1	7 715	3 348 042	92,5	4,9	Bacteria;Chloroflexi
CFX6	9	176	62,7	39 759	4 087 257	97,2	1,9	Bacteria;Chloroflexi;
CFX7	24	59	51,1	177 076	4 915 330	100,0	1,9	Bacteria;Chloroflexi;
CFX8	25	143	55	52 376	4 720 749	98,1	1,0	Bacteria;Chloroflexi;
CFX9	48	153	56,9	51 559	4 864 248	89,6	2,1	Bacteria;Chloroflexi
CFX10	87	250	67,4	10 884	1 908 586	60,4	1,5	Bacteria;Chloroflexi;
CLB1	75	209	49,7	20 990	2 464 280	82,1	2,2	Bacteria;Chlorobi;
DJK1	6	3	38,4	332 333	789 939	75,5	0,0	Bacteria;Candidatus Dojikabacteria;
DJK2	5	9	35,5	247 571	1 202 451	80,2	0,0	Bacteria;Candidatus Dojikabacteria;
GMM1	49	45	64,3	1 086 438	3 611 097	84,0	1,1	Bacteria;Gemmatimonadetes;
IGN1	2	74	33,4	71 666	3 211 749	100,0	0,9	Bacteria;Ignavibacteriae;Ignavibacteria;
IGN2	4	127	34,8	35 381	2 841 441	99,1	0,9	Bacteria;Ignavibacteriae
NIT1	29	100	48,5	1 689 662	1 908 318	99,1	0,9	Bacteria;Proteobacteria;Betaproteobacteria; Nitrosomonadales;Nitrosomonadaceae;Nitrosomonas
NIT2	31	37	50,7	2 429 488	2 677 054	98,1	1,9	Bacteria;Proteobacteria;Betaproteobacteria; Nitrosomonadales;Nitrosomonadaceae;Nitrosomonas
NIT3	37	32	50,3	2 081 098	2 164 667	99,1	1,9	Bacteria;Proteobacteria;Betaproteobacteria; Nitrosomonadales;Nitrosomonadaceae;Nitrosomonas
PRO1	15	63	62,9	76 361	3 124 961	91,5	1,0	Bacteria; Proteobacteria; Deltaproteobacteria; Bdellovibrionales
PRO2	19	15	55,6	1 529 228	2 722 707	100,0	0,9	Bacteria;Proteobacteria;Alphaproteobacteria
PRO3	33	120	68,9	58 552	3 578 718	100,0	0,9	Bacteria;Proteobacteria;Betaproteobacteria
PRO4	74	51	68,4	109 901	3 582 674	90,7	1,0	Bacteria;Proteobacteria;Gammaproteobacteria
PRO5	22	320	66,6	15 342	2 757 229	94,3	1,0	Bacteria;Proteobacteria;Betaproteobacteria
PRO6	27	326	70,1	25 192	4 238 876	100,0	1,9	Bacteria;Proteobacteria;Betaproteobacteria;
PRO7	40	401	69,3	7 737	2 618 829	87,7	0,0	Bacteria;Proteobacteria;Gammaproteobacteria

PRO8	**45**	**1521**	**73,3**	**5 657**	**6 512 352**	**95,3**	**3,8**	**Bacteria;Proteobacteria;Deltaproteobacteria; Myxococcales**
PRO9	51	245	64,2	53 017	3 648 696	100,0	2,8	Bacteria;Proteobacteria;Gammaproteobacteria
PRO10	55	101	68,3	8 340	489 986	18,9	0,0	Bacteria;Proteobacteria;Betaproteobacteria
PRO11	60	217	73,3	11 102	1 689 039	63,2	0,0	Bacteria;Proteobacteria;Alphaproteobacteria; Rhodobacterales;Rhodobacteraceae
PRO12	61	200	71	38 822	3 028 291	83,0	4,3	Bacteria;Proteobacteria;Betaproteobacteria
VER1	78	225	60	6 466	1 394 109	37,7	0,0	Bacteria;Verrucomicrobia
VER2	52	140	52,5	31 314	2 113 811	49,1	0,0	Bacteria;Verrucomicrobia
VER3	54	157	59,8	19 477	2 043 526	76,4	0,0	Bacteria;Verrucomicrobia

Supplementary Table 2 – Summary of the verification of the unexpected genotype procedure.

Bin ID	no ID	Completness(%)	Contamination(%)	Binning Method	Unexpected	Complete Operon	Incomplete Operon	Doubt	To do
ACD1	16	100,0	2,8	Differential Coverage	yes		nxrB		Search missing genes
ACD2	20	99,1	5,4	VizBin	No	narGHIJ, nrfAH			Nothing
ACD3	58	95,3	4,7	Differential Coverage	yes	nirK, narGHIJ			confirm qnor
ACT1	59	45,3	0,0	Differential Coverage	no	nirK			Nothing
ACT2	66	38,7	0,0	Differential Coverage	no	narGHIJ			Nothing
AMX1	28	99,1	2,8	Differential Coverage	No	nirK,NxrAB			Nothing
AMX2	30	97,2	3,7	Differential Coverage	No	NxrAB			Nothing
AMX3	46	90,6	5,0	Differential Coverage	no	nrfAH, nxrAB			Nothing
ARM1	1	94,3	1,0	VizBin	yes	nirK, NxrAB, nrfAH			DF
BCD1	3	100,0	1,9	VizBin	No	nosZDFY			Nothing
BCD2	57	87,7	1,1	Differential Coverage	yes	narGHIJK, nosZLDY, nod			search for nir
BCG1	23	100,0	0,9	VizBin	Yes	nosZDY, narGHJI			Search nir and nor
BCT1	50	88,7	2,1	Differential Coverage	yes	nosZDLY,nxrAB(2 copies)			Nothing
BCT2	21	100,0	1,9	VizBin	Yes	nirK, nosZDFYL			Search nor
BCT3	44	96,2	1,0	Differential Coverage	yes	nosZDY	norDQ, +1 copies of NarGH, nxrB		Search missing genes

BCT4	**53**	**66,0**	**1,4**	**Differential Coverage**	**no**				**Nothing**
BCT5	56	87,7	0,0	Differential Coverage	yes	nirK, nosZDY			Search nor
BCT6	64	39,6	0,0	Differential Coverage	yes	nosZLDY, nirK			search nor
BCT7	67	38,7	0,0	Differential Coverage	no	narGHIJ			Nothing
BCT8	8	81,1	0,0	VizBin	no				Nothing
BCT9	72	0,9	0,0	VizBin	no	nirK			Nothing
BCT10	85	74,5	1,3	Differential Coverage	no	nrfAH			Nothing
BCT11	36	96,2	4,7	Differential Coverage	yes	nosZDYL,nrfAH	narGH		Search nir and nor
CFX1	73	4,7	0,0	VizBin	no	nirK			Nothing
CFX2	10	96,2	2,9	Differential Coverage	No	nirK, narGHIJK(2 copies)			Nothing
CFX3	34	96,2	2,9	Differential Coverage	no	narGHIJ(2 copies), nirK			Nothing
CFX4	38	97,2	2,8	Differential Coverage	No	nirK			Nothing
CFX5	47	92,5	4,9	Differential Coverage	no			nirS	Nothing
CFX6	9	97,2	1,9	VizBin	yes	nirK, nosZDY,nrfAH			Search nir and nor
CFX7	24	100,0	1,9	VizBin	No	nosZDY,nirBD,nrfAH			Nothing
CFX8	25	98,1	1,0	VizBin	Yes	nosZFYL, narGHJIK,nrfAH			Search nir and nor

CFX9	**48**	**89,6**	**2,1**	**Differential Coverage**	**yes**	**nirK, nosZDYL,nxrAB**			**Search nor**
CFX10	87	60,4	1,5	Differential Coverage	no	nrfAH			Nothing
CLB1	75	82,1	2,2	VizBin	yes	narGHIJ, nosZDY			search nor
DJK1	6	75,5	0,0	Differential Coverage	no				Nothing
DJK2	5	80,2	0,0	Differential Coverage	no				Nothing
GMM 1	49	84,0	1,1	Differential Coverage	yes	nirK(new gene), nosZLDY, narGHIJK (new)			Search nor/ Check with Blast
IGN1	2	100,0	0,9	VizBin	yes	nosZLDFY, narGHIK(J), nrfAH			DF
IGN2	4	99,1	0,9	Differential Coverage	yes	nosZLDY, narGHIK(J),nrfAH			Search nir and nor
NIT1	29	99,1	0,9	Differential Coverage	No	nirK	norBC		Nothing
NIT2	31	98,1	1,9	Differential Coverage	No	nirK	norCBQ		Nothing
NIT3	37	99,1	1,9	Differential Coverage	no	nirK, norBCDQ			Nothing
PRO1	15	91,5	1,0	VizBin	Yes	nirK, nosZDFY,nrfAH			Search nor
PRO2	19	100,0	0,9	VizBin	no	narGHIJ			Nothing
PRO3	33	100,0	0,9	Differential Coverage	yes	haoABcycAB, narGHIJK, nosRZDFYL,NxrAB	HAO, nirSN, norBC		Nothing
PRO4	74	90,7	1,0	Differential Coverage	no	narGHJI	NirS		Nothing

PRO5	**22**	**94,3**	**1,0**	**Differential Coverage**	**Yes**	**nosZD(NapGH)FLY, narGHJIK, (nosR)nirSCFLGH, nirS(second copy) ,haoABcycAB**			**Search nor**
PRO6	27	100,0	1,9	Differential Coverage	Yes	narGHIJK, nirSCFLGHJN,NxrAB			Nothing
PRO7	40	87,7	0,0	Differential Coverage	yes	nirK,	norBC, nirS		Search missing genes
PRO8	45	95,3	3,8	Differential Coverage	yes	nosZDYL,nrfAH	nxrB		Search missing genes
PRO9	51	100,0	2,8	Differential Coverage	no	narGHJI, nirK, norZ and nod			Nothing
PRO10	55	18,9	0,0	Differential Coverage	no				Nothing
PRO11	60	63,2	0,0	Differential Coverage	yes	narGHJ, nxrAB,		haoAB(no cytochromes)	Nothing
PRO12	61	83,0	4,3	Differential Coverage	yes	narGHIJ, nirS	haoAB, norBC		Nothing
VER1	78	37,7	0,0	VizBin	no				Nothing
VER2	52	49,1	0,0	Differential Coverage	no				Nothing
VER3	54	76,4	0,0	Differential Coverage	no				Nothing

Supplementary Table 3 – Summary of the 16S rRNA retrieved from each MAG.

Bin ID	no ID	Blast uncultured 16S	%	Blast no uncultured 16S	%
ACD1	16				
ACD2	20				
ACD3	58	Opitutus terrae strain PB90-1 16S	95	Opitutus terrae strain PB90-1 16S	95
ACT1	59				
ACT2	66				
AMX1	28				
AMX2	30				
AMX3	46				
ARM1	1	Uncultured bacterium clone Ge64 16S	99	Fimbriimonas ginsengisoli Gsoil	90
BCD1	3				
BCD2	57	Uncultured bacterium clone F_SBR_37 16S	99	Owenweeksia hongkongensis strain DSM 17368 16S	86
BCG1	23	Uncultured bacterium clone Dok43 16S*	100	Unidentified bacterium clone K2-30-37 16S	85
BCT1	50				
BCT10	85				
BCT11	36	Uncultured bacteria partial 16S rRNA gene, clone b51	99	Caldilinea aerophila strain DSM 14535 16S	88
BCT2	21				
BCT3	44	Uncultured bacterium partial 16S rRNA gene, isolate BAG, T0, clone B131	98	Bacterium YC-LK-LKJ27 16S	90
BCT4	53	Uncultured bacterium clone SCS_HX36_99 16S	93	Candidatus Pacebacteria bacterium GW2011_GWF2_38_9 16S	87
BCT5	56				
BCT6	64				
BCT7	67				
BCT8	8	Bacterium enrichment culture clone EBa4 16S	99	Berkelbacteria bacterium GW2011_GWE1_39_12 16S	80
BCT9	72				
CFX1	73				
CFX10	87				
CFX2	10	Uncultured bacterium clone N-25 16S	100	Thermanaerothrix daxensis strain GNS-1 16S	91/90
CFX3	34				
CFX4	38	Uncultured bacterium clone B134 16S	100	Ornatilinea apprima strain P3M-1 16S	89
CFX5	47				
CFX6	9				
CFX7	24				
CFX8	25				
CFX9	48				
CLB1	75				

DJK1	6	Uncultured prokaryote clone Fr1-18 16S	95	Candidate division WS6 bacterium GW2011_GWC2_36_7 16S	88
DJK2	5	Uncultured bacterium clone B75 16S	95	Candidate division WS6 bacterium GW2011_GWE2_33_157 16S	80
GMM1	49				
IGN1	2				
IGN2	4				
NIT1	29				
NIT2	31				
NIT3	37				
PRO1	15	Uncultured bacterium clone GT30901f04 16S	98	Bdellovibrio bacteriovorus str. Tiberius	90
PRO10	55				
PRO11	60				
PRO12	61				
PRO2	19	Uncultured bacterium clone BT-57 16S	99	Micavibrio aeruginosavorus strain ARL-13 16S	88
PRO3	33	Uncultured bacterium clone B108 16S	99	Zhizhongheella caldifontis strain YIM 78140 16S	94
PRO4	74				
PRO5	22				
PRO6	27				
PRO7	40				
PRO8	45				
PRO9	51	Uncultured bacterium clone NIT_90 16S	99	Sorangiineae bacterium 706 16S	98
VER1	78				
VER2	52				
VER3	54				

4

CONCLUSION

4.1 CONCLUSIONS

This master thesis has contributed to the characterization of the interactions and dynamics of heterotrophic partial denitrifiers in a partial-nitrification/anammox reactor. Previous work offered some insight on the pathways representation in this type of wastewater treatment, but no other information besides the description was provided. Additionally, putative groups of organisms have been discovered, showing the complexity of these microbial communities and the lack of information to fully understand them.

Overall, the main aim of describing and analysis the interactions in this microbial community was accomplished. The results from this work can possibly enable future improvements to the *PNA* wastewater treatment, decreasing the instability, lowering the nitrous oxide emissions and improving the overall nitrogen removal.

The main finding of this work can be found bellow:

- Through metagenomic analysis, 57 *MAGs*, from which, 45 are near-complete genomes were reconstructed from a *PNA* reactor. The high number of obtained genomes was due to performing two different binning methods, and the co-assembly of similar samples to cluster low abundance genomes.

- Most of the recovered organisms have the capacity to perform at least one step of the nitrogen cycle.

- Although the system supresses the growth of *NOB*, a putative nitrite oxidizing bacteria was discovered, that has the HAO cluster and achieves complete denitrification.

- Two new groups of putative heterotrophic hydroxylamine oxidizers and nitrite oxidizers have been described for the first time.

- Finally, heterotrophic denitrifiers have been attributed a specific niche accordingly to their role in the denitrification. Four different niches have been created, nitrite reducers, nitrous oxide reducers, internal deletion and nitrite to nitrous oxide reducers.

4.2 PROSPECT FOR FUTURE WORK

In this Master Thesis, 57 metagenomic assembled genomes were successfully recovered from a *PNA* system. The functional annotation of these genomes allowed to describe the interactions involving the nitrogen cycle surrounding this community. But only the information of the metabolic analysis and average abundance of genomes were used in this work.

Throughout the seven samples, although the variation of abundance was relatively small, some organisms suffered a substantial increase or decrease so further analysis for the causes of this alterations are required (Appendix I). One of the possible explanations for the variation in abundance is the presence of predatory bacteria in the community. When two predators start increasing in abundance, some organisms, like the anammox bacteria start decrease (Appendix I). Therefore, functional annotation of genes associated with predation, like the ability to degrade polymeric compounds, genes that give mobility to the cells and gene associated with the isoprenoids, in which the mevalonic acid pathway has been associated with predation.

The complexity of this community was ascertained with the discovery of putative heterotrophic hydroxylamine oxidizers and putative heterotrophic nitrite oxidizers. *Hao*-like proteins divided in several clusters were also found, but further information on these proteins is required. A study based on the evolutionary analysis of the *hao*, *nxr* and *hao*-like proteins can provide the information of how these proteins differ from the typical found in their respective organism, how these organisms have acquired these new functions and how these organisms and their new role behave in the environment. Thus, metatranscriptomic, metaproteomic and evolutionary analysis to these proteins would insight on some of these questions that arise from these new discoveries.

Lastly, the analysis of the horizontal gene transfer and plasmid throughout the time to understand the community adaptation and possibility provide more information on how

the niches involving the partitioning of the denitrification pathway were created.

BIBLIOGRAPHY

N. Abdel-Raouf, A. A. Al-Homaidan, and I. B.M. Ibraheem. Microalgae and wastewater treatment. *Saudi Journal of Biological Sciences*, 19(3):257–275, 2012. ISSN 1319562X. doi: 10.1016/j.sjbs.2012.04.005. URL http://dx.doi.org/10.1016/j.sjbs.2012.04.005.

L Ackert. *Sergei Vinogradskii and the Cycle of Life: From the Thermodynamics of Life to Ecological Microbiology, 1850-1950*. Archimedes. Springer Netherlands, 2012. ISBN 9789400751989. URL https://books.google.pt/books?id=g-kKPYLzBR0C.

Mads Albertsen, Philip Hugenholtz, Adam Skarshewski, Kåre L Nielsen, Gene W Tyson, and Per H Nielsen. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature biotechnology*, 31(6):533–8, 2013. ISSN 1546-1696. doi: 10.1038/nbt.2579. URL http://www.ncbi.nlm.nih.gov/pubmed/23707974.

Muhammad Ali and Satoshi Okabe. Anammox-based technologies for nitrogen removal: Advances in process start-up and remaining issues. *Chemosphere*, 141:144–153, 2015. ISSN 00456535. doi: 10.1016/j.chemosphere.2015.06.094. URL http://dx.doi.org/10.1016/j.chemosphere.2015.06.094.

S Andrews. FastQC A Quality Control tool for High Throughput Sequence Data, 2010. URL http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

T Asano. *Wastewater Reclamation and Reuse: Water Quality Management Library*. Number v. 10. Taylor & Francis, 1998. ISBN 9781566766203. URL https://books.google.pt/books?id=9vAGkleP234C.

Ramy K Aziz, Daniela Bartels, Veronika Best, Andreas Wilke, and Olga Zagnitko. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, 9(1): 75, 2008. ISSN 1471-2164. doi: 10.1186/1471-2164-9-75. URL http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-9-75.

Jørgen Bang-Jensen, Gregory Gutin, and Anders Yeo. When the greedy algorithm fails. *Discrete Optimization*, 1(2):121–127, 2004. ISSN 15725286. doi: 10.1016/j.disopt.2004.03.007.

Ananda S. Bhattacharjee, Sha Wu, Christopher E. Lawson, Mike S.M. Jetten, Vikram Kapoor, Jorge W.Santo Domingo, Katherine D. McMahon, Daniel R. Noguera, and Ramesh Goel.

Whole-Community Metagenomics in Two Different Anammox Configurations: Process Performance and Community Structure. *Environmental Science and Technology*, 51(8):4317–4327, 2017. ISSN 15205851. doi: 10.1021/acs.est.6b05855.

Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120, 2014. ISSN 14602059. doi: 10.1093/bioinformatics/btu170.

Maria Tereza Borges, Antônio Galvão Nascimento, Ulisses Nunes Rocha, and Marcos Rogério Tótola. Nitrogen starvation affects bacterial adhesion to soil. *Brazilian Journal of Microbiology*, 39(3):457–463, 2008. ISSN 15178382. doi: 10.1590/S1517-83822008000300009.

H Bothe, W E Newton, and S Ferguson. *Biology of the Nitrogen Cycle*. Elsevier, 2007. ISBN 9780444531087. URL https://books.google.pt/books?id=FKe4ojTXw48C.

L. Bouwman, K. K. Goldewijk, K. W. Van Der Hoek, A. H. W. Beusen, D. P. Van Vuuren, J. Willems, M. C. Rufino, and E. Stehfest. Exploring global changes in nitrogen and phosphorus cycles in agriculture induced by livestock production over the 1900-2050 period. *Proceedings of the National Academy of Sciences*, 110(52):20882–20887, 2013. ISSN 0027-8424. doi: 10.1073/pnas.1012878108. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1012878108.

E Broda. Two kinds of lithotrophs missing in nature. *Zeitschrift f??r allgemeine Mikrobiologie*, 17(6):491–493, 1977. ISSN 15214028. doi: 10.1002/jobm.19770170611. URL http://www.ncbi.nlm.nih.gov/pubmed/930125.

Allen G. Burton and Robert E Pitt. Receiving Water Uses, Impairments, and Sources of Stormwater Pollutants. *Stormwater Effects Handbook*, pages 15–45, 2002.

Reviewed by Juan Ilerbaig. Enriching the Earth: Fritz Haber, Carl Bosch, and the Transformation of World Food Production. By Vaclav Smil. *The Quarterly Review of Biology*, 77(1): 86, 2002. doi: 10.1086/343654. URL https://doi.org/10.1086/343654.

Phillip E C Compeau, Pavel A Pevzner, and Glenn Tesler. How to apply de Bruijn graphs to genome assembly. *Nature Biotechnology*, 29(11):987–991, 2011. ISSN 1087-0156. doi: 10.1038/nbt.2023. URL http://www.nature.com/doifinder/10.1038/nbt.2023.

Holger Daims, Elena V. Lebedeva, Petra Pjevac, Ping Han, Craig Herbold, Mads Albertsen, Nico Jehmlich, Marton Palatinszky, Julia Vierheilig, Alexandr Bulaev, Rasmus H. Kirkegaard, Martin von Bergen, Thomas Rattei, Bernd Bendinger, Per H. Nielsen, and Michael Wagner. Complete nitrification by Nitrospira bacteria. *Nature*, 528(7583):504–509, 2015. ISSN 0028-0836. doi: 10.1038/nature16461. URL http://www.nature.com/doifinder/10.1038/nature16461.

Andreas Dietl, Christina Ferousi, Wouter J. Maalcke, Andreas Menzel, Simon de Vries, Jan T. Keltjens, Mike S. M. Jetten, Boran Kartal, and Thomas R. M. Barends. The inner workings of the hydrazine synthase multiprotein complex. *Nature*, 527(7578):394–397, 2015. ISSN 0028-0836. doi: 10.1038/nature15517. URL http://www.nature.com/doifinder/10.1038/nature15517.

Carlos Domingo-Félez, A. Gizem Mutlu, Marlene M. Jensen, and Barth F. Smets. Aeration strategies to mitigate nitrous oxide emissions from single-stage nitritation/anammox reactors. *Environmental Science and Technology*, 48(15):8679–8687, 2014. ISSN 15205851. doi: 10.1021/es501819n.

Robert C. Edgar. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19):2460–2461, 2010. ISSN 13674803. doi: 10.1093/bioinformatics/btq461.

A. Murat Eren, Özcan C. Esen, Christopher Quince, Joseph H. Vineis, Hilary G. Morrison, Mitchell L. Sogin, and Tom O. Delmont. Anvi'o: an advanced analysis and visualization platform for omics data. *PeerJ*, 3:e1319, 2015. ISSN 2167-8359. doi: 10.7717/peerj.1319. URL https://peerj.com/articles/1319.

Alejandra Escobar-Zepeda, Arturo Vera Ponce De León, and Alejandro Sanchez-Flores. The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in Genetics*, 6(DEC):1–15, 2015. ISSN 16648021. doi: 10.3389/fgene.2015.00348.

Katharina F Ettwig, Margaret K Butler, Kees-Jan Le Paslier, Henk Stunnenberg, Jean Weissenbach, Mike S M Jetten, and Marc Strous. Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature*, 464(7288):543–548, 2010. ISSN 0028-0836. doi: 10.1038/nature08883. URL http://dx.doi.org/10.1038/nature08883.

Scott Fields. Global Nitrogen: Cycling out of control. *Environmental Health Perspectives*, 112 (10):556–563, 2004. ISSN 00916765. doi: 10.1289/ehp.112-a556.

Robert D. Finn, Jody Clements, and Sean R. Eddy. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research*, 39(SUPPL. 2):29–37, 2011. ISSN 03051048. doi: 10.1093/nar/gkr367.

Robert D. Finn, Alex Bateman, Jody Clements, Penelope Coggill, Ruth Y. Eberhardt, Sean R. Eddy, Andreas Heger, Kirstie Hetherington, Liisa Holm, Jaina Mistry, Erik L.L. Sonnhammer, John Tate, and Marco Punta. Pfam: The protein families database. *Nucleic Acids Research*, 42(D1):1–9, 2014. ISSN 03051048. doi: 10.1093/nar/gkt1223.

J. A. Foley. Global Consequences of Land Use. *Science*, 309(5734):570–574, 2005. ISSN 0036-8075. doi: 10.1126/science.1111772. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1111772.

David Fowler, Mhairi Coyle, Ute Skiba, Mark A Sutton, J Neil Cape, Stefan Reis, Lucy J Sheppard, Alan Jenkins, Bruna Grizzetti, N Galloway, Peter Vitousek, Allison Leach, Alexander F Bouwman, Klaus Butterbach-bahl, Frank Dentener, David Stevenson, Marcus Amann, Maren Voss, and David Fowler. The global nitrogen cycle in the twenty- first century. *Philisophical transactions of the Royal Society B*, 2013. doi: http://dx.doi.org/10.1098/rstb.2013.0164. URL http://rstb.royalsocietypublishing.org/content/royptb/368/1621/20130164.full.pdf.

Eldon Franz. Ecology: Individuals, Populations and Communities. *Journal of Range Management*, 41:524, 1988. ISSN 0046-225X. doi: 10.1007/s13398-014-0173-7.2. URL https://www.cabdirect.org/cabdirect/abstract/19870540497.

Christian Fux, Marc Boehler, Philipp Huber, Irene Brunner, and Hansruedi Siegrist. Biological treatment of ammonium-rich wastewater by partial nitritation and subsequent anaerobic ammonium oxidation (anammox) in a pilot plant. *Journal of Biotechnology*, 99 (3):295–306, 2002. ISSN 01681656. doi: 10.1016/S0168-1656(02)00220-1.

J. Galloway. The global nitrogen cycle: past, present and future. *Science in China. Series C, Life sciences / Chinese Academy of Sciences*, 48 Suppl 2(326):669–678, 2005. ISSN 1006-9305. doi: 10.1360/062005-261.

J. N. Galloway, A. M. Leach, A. Bleeker, and J. W. Erisman. A chronology of human understanding of the nitrogen cycle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1621):20130120–20130120, 2013. ISSN 0962-8436. doi: 10.1098/rstb.2013.0120. URL http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2013.0120.

M H Gerardi. *Nitrification and Denitrification in the Activated Sludge Process*. Environmental Protection magazine series. Wiley, 2003. ISBN 9780471461319. URL https://books.google.pt/books?id=ccIKuXYNlBMC.

B. M. Griffin, J. Schott, and B. Schink. Nitrite, an Electron Donor for Anoxygenic Photosynthesis. *Science*, 316(5833):1870–1870, 2007. ISSN 0036-8075. doi: 10.1126/science.1139478. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1139478.

Alexey Gurevich, Vladislav Saveliev, Nikolay Vyahhi, and Glenn Tesler. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, 29(8):1072–1075, 2013. ISSN 13674803. doi: 10.1093/bioinformatics/btt086.

J Handelsman, J Tiedje, National Research Council (US) Committee on Metagenomics: Challenges and Functional, and Applications. *THE NEW SCIENCE OF METAGENOMICS: Revealing the Secrets of Our Microbial Planet*. 2007. ISBN 9780309106764. doi: NBK54006[bookaccession]. URL http://www.ncbi.nlm.nih.gov/pubmed/

21678629{%}5Cnhttp://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:The+new+science+of+metagenomics:+Revealing+the+secrets+of+our+microbial+planet{#}0.

Jo Handelsman, Michelle R. Rondon, Sean F. Brady, Jon Clardy, and Robert M. Goodman. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, 5(10):R245–R249, 1998. ISSN 10745521. doi: 10.1016/S1074-5521(98)90108-9. URL http://linkinghub.elsevier.com/retrieve/pii/S1074552198901089.

Kazuaki Hibiya, Akihiko Terada, Satoshi Tsuneda, and Akira Hirata. Simultaneous nitrification and denitrification by controlling v ertical and horizontal microen v ironment in a membrane- aerated biofilm reactor. 100:23–32, 2003. ISSN 0168-1656.

Brian M. Hoffman, Dmitriy Lukoyanov, Dennis R. Dean, and Lance C. Seefeldt. Nitrogenase: A draft mechanism. *Accounts of Chemical Research*, 46(2):587–595, 2013. ISSN 00014842. doi: 10.1021/ar300267m.

Ziye Hu, Tommaso Lotti, Merle de Kreuk, Robbert Kleerebezem, Mark van Loosdrecht, Jans Kruit, Mike S M Jetten, and Boran Kartal. Nitrogen removal by a nitritation-anammox bioreactor at low temperature. *Applied and Environmental Microbiology*, 79(8):2807–2812, 2013. ISSN 00992240. doi: 10.1128/AEM.03987-12.

D Huson, A Auch, J Qi, and S Schuster. MEGAN analysis of metagenome data. *Gennome Res.*, 17:377–386, 2007. ISSN 10889051. doi: 10.1101/gr.5969107.

S. H. Isaacs and M. Henze. Controlled carbon source addition to an alternating nitrification-denitrification wastewater treatment process including biological P removal. *Water Research*, 29(1):77–89, 1995. ISSN 00431354. doi: 10.1016/0043-1354(94)E0119-Q.

Juan Jovel, Jordan Patterson, Weiwei Wang, Naomi Hotte, Sandra O'Keefe, Troy Mitchel, Troy Perry, Dina Kao, Andrew L. Mason, Karen L. Madsen, and Gane K.S. Wong. Characterization of the gut microbiome using 16S or shotgun metagenomics. *Frontiers in Microbiology*, 7(APR):1–17, 2016. ISSN 1664302X. doi: 10.3389/fmicb.2016.00459.

Marlies J. Kampschreur, Hardy Temmink, Robbert Kleerebezem, Mike S M Jetten, and Mark C M van Loosdrecht. Nitrous oxide emission during wastewater treatment. *Water Research*, 43(17):4093–4103, 2009. ISSN 00431354. doi: 10.1016/j.watres.2009.03.001. URL http://dx.doi.org/10.1016/j.watres.2009.03.001.

B. Kartal, J. G. Kuenen, and M. C. M. van Loosdrecht. Sewage Treatment with Anammox. *Science*, 328(5979):702–703, 2010. ISSN 0036-8075. doi: 10.1126/science.1185941. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1185941.

Boran Kartal, Naomi M. De Almeida, Wouter J. Maalcke, Huub J M Op den Camp, Mike S M Jetten, and Jan T. Keltjens. How to make a living from anaerobic ammonium oxidation. *FEMS Microbiology Reviews*, 37(3):428–461, 2013. ISSN 01686445. doi: 10.1111/1574-6976.12014.

Kevin P. Keegan, Elizabeth M. Glass, and Folker Meyer. MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function. In *Methods in molecular biology (Clifton, N.J.)*, volume 1399, pages 207–233. 2016. ISBN 978-1-4939-3367-9. doi: 10.1007/978-1-4939-3369-3_13. URL http://www.ncbi.nlm.nih.gov/pubmed/26791506http://link.springer.com/10.1007/978-1-4939-3369-3{_}13.

Amelia K. Kivaisi. The potential for constructed wetlands for wastewater treatment and reuse in developing countries: A review. *Ecological Engineering*, 16(4):545–560, 2001. ISSN 09258574. doi: 10.1016/S0925-8574(00)00113-0.

Allen V Kneese. *Water Pollution: Economics Aspects and Research Needs*. Routledge, 2015.

Satish Kumar, Kishore Kumar Krishnani, Bharat Bhushan, and Manoj Pandit Brahmane. Metagenomics: Retrospect and Prospects in High Throughput Age. *Biotechnology Research International*, 2015:1–13, 2015. ISSN 2090-3138. doi: 10.1155/2015/121735. URL http://www.hindawi.com/journals/btri/2015/121735/.

Hannon Lab. FASTX Toolkit, 2009. URL http://hannonlab.cshl.edu/fastx{_}toolkit/.

Cedric C. Laczny, Nicolás Pinel, Nikos Vlassis, and Paul Wilmes. Alignment-free Visualization of Metagenomic Data by Nonlinear Dimension Reduction. *Scientific Reports*, 4 (1):4516, 2015. ISSN 2045-2322. doi: 10.1038/srep04516. URL http://www.nature.com/articles/srep04516.

Phyllis Lam and Marcel M.M. Kuypers. Microbial Nitrogen Cycling Processes in Oxygen Minimum Zones. *Annual Review of Marine Science*, 3(1):317–345, 2011. ISSN 1941-1405. doi: 10.1146/annurev-marine-120709-142814. URL http://www.annualreviews.org/doi/10.1146/annurev-marine-120709-142814.

Michela Langone, Jia Yan, Suzanne C M Haaijer, Huub J M Op den Camp, Mike S M Jetten, and Gianni Andreottola. Coexistence of nitrifying, anammox and denitrifying bacteria in a sequencing batch reactor. *Frontiers in Microbiology*, 5(FEB):1–12, 2014. ISSN 1664302X. doi: 10.3389/fmicb.2014.00028.

Michele Laureni, Per Falås, Orlane Robin, Arne Wick, David G. Weissbrodt, Jeppe Lund Nielsen, Thomas A. Ternes, Eberhard Morgenroth, and Adriano Joss. Mainstream partial nitritation and anammox: Long-term process stability and effluent quality at low temperatures. *Water Research*, 101:628–639, 2016. ISSN 18792448. doi: 10.1016/j.watres.2016.05.005. URL http://dx.doi.org/10.1016/j.watres.2016.05.005.

David J. Levy-Booth, Cindy E. Prescott, and Susan J. Grayston. Microbial functional genes involved in nitrogen fixation, nitrification and denitrification in forest ecosystems. *Soil Biology and Biochemistry*, 75:11–25, 2014. ISSN 00380717. doi: 10.1016/j.soilbio.2014.03.021.

Hanyan Li and Wendong Tao. Efficient ammonia removal in recirculating vertical flow constructed wetlands: Complementary roles of anammox and denitrification in simultaneous nitritation, anammox and denitrification process. *Chemical Engineering Journal*, 317:972–979, 2017. ISSN 13858947. doi: 10.1016/j.cej.2017.02.143. URL http://dx.doi.org/10.1016/j.cej.2017.02.143.

Chengwei Luo, Luis M. Rodriguez-R, and Konstantinos T. Konstantinidis. MyTaxa: An advanced taxonomic classifier for genomic and metagenomic sequences. *Nucleic Acids Research*, 42(8):1–12, 2014. ISSN 13624962. doi: 10.1093/nar/gku169.

Abid A. Ansari Sarvajeet Singh Gill (eds.) M. Nasir Khan, Firoz Mohammad (auth.). *Eutrophication: Causes, Consequences and Control: Volume 2*. Springer Netherlands, 1 edition, 2014. ISBN 978-94-007-7813-9,978-94-007-7814-6. URL http://gen.lib.rus.ec/book/index.php?md5=1BD129789E4E0FDA6F99A31209544A97.

Muhammad Arif Malik, Chunqi Jiang, Richard Heller, Jamie Lane, David Hughes, and Karl H. Schoenbach. Ozone-free nitric oxide production using an atmospheric pressure surface discharge - A way to minimize nitrogen dioxide co-production. *Chemical Engineering Journal*, 283:631–638, 2016. ISSN 13858947. doi: 10.1016/j.cej.2015.07.092. URL http://dx.doi.org/10.1016/j.cej.2015.07.092.

José L Martínez. Antibiotics and Antibiotic Resistance. *Science*, 321(July):365–367, 2008. ISSN 01460005. doi: 10.1002/9781118075609.

Céline Masclaux-Daubresse, Franoise Daniel-Vedele, Julie Dechorgnat, Fabien Chardon, Laure Gaufichon, and Akira Suzuki. Nitrogen uptake, assimilation and remobilization in plants: Challenges for sustainable and productive agriculture. *Annals of Botany*, 105(7): 1141–1157, 2010. ISSN 03057364. doi: 10.1093/aob/mcq028.

Peter Menzel, Kim Lee Ng, and Anders Krogh. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*, 7:11257, 2016. ISSN 2041-1723. doi: 10.1038/ncomms11257. URL http://www.nature.com/doifinder/10.1038/ncomms11257.

Alla Mikheenko, Gleb Valin, Andrey Prjibelski, Vladislav Saveliev, and Alexey Gurevich. Icarus: Visualizer for de novo assembly evaluation. *Bioinformatics*, 32(21):3321–3323, 2016. ISSN 14602059. doi: 10.1093/bioinformatics/btw379.

J W B Moir. *Nitrogen Cycling in Bacteria: Molecular Analysis*. Caister Academic Press, 2011. ISBN 9781904455868. URL https://books.google.pt/books?id=d3FMJgDPlY4C.

A Mulder, A A Graaf, and L A Robertson. Anaerobic ammonium oxidation discovered in a denitrifying fluidized bed reactor. *FEMS Microbiology Ecology*, 16:177–184, 1995. doi: 10.1111/j.1574-6941.1995.tb00281.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1574-6941.1995.tb00281.x/abstract.

Poonam Singh Nigam and Anoop Singh. Production of liquid biofuels from renewable resources. *Progress in Energy and Combustion Science*, 37(1):52–68, 2011. ISSN 03601285. doi: 10.1016/j.pecs.2010.01.003. URL http://dx.doi.org/10.1016/j.pecs.2010.01.003.

Hiroyuki Ogata, Susumu Goto, Kazushige Sato, Wataru Fujibuchi, Hidemasa Bono, and Minoru Kanehisa. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 27(1):29–34, 1999. ISSN 03051048. doi: 10.1093/nar/27.1.29.

Anastasis Oulas, Christina Pavloudi, Paraskevi Polymenakou, Georgios A. Pavlopoulos, Nikolas Papanikolaou, Georgios Kotoulas, Christos Arvanitidis, and Ioannis Iliopoulos. Metagenomics: Tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinformatics and Biology Insights*, 9:75–88, 2015. ISSN 11779322. doi: 10.4137/BBI.S12462.

Ross Overbeek, Robert Olson, Gordon D. Pusch, Gary J. Olsen, James J. Davis, Terry Disz, Robert A. Edwards, Svetlana Gerdes, Bruce Parrello, Maulik Shukla, Veronika Vonstein, Alice R. Wattam, Fangfang Xia, and Rick Stevens. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, 42 (D1):206–214, 2014. ISSN 03051048. doi: 10.1093/nar/gkt1226.

Alejandro Palomo, S Jane Fowler, Arda Gülay, Simon Rasmussen, Thomas Sicheritz-Ponten, and Barth F Smets. Metagenomic analysis of rapid gravity sand filter microbial communities suggests novel physiology of Nitrospira spp. *The ISME Journal*, 10(11):2569–2581, 2016. ISSN 1751-7362. doi: 10.1038/ismej.2016.63. URL http://www.nature.com/doifinder/10.1038/ismej.2016.63.

Donovan H Parks, Michael Imelfort, Connor T Skennerton, Philip Hugenholtz, and Gene W Tyson. CheckM: assessing the quality of microbial genomes recovered from. *Cold Spring Harbor Laboratory Press Method*, pages 1–31, 2015. ISSN 1549-5469. doi: 10.1101/gr.186072.114.

Donovan H. Parks, Christian Rinke, Maria Chuvochina, Pierre-Alain Chaumeil, Ben J. Woodcroft, Paul N. Evans, Philip Hugenholtz, and Gene W. Tyson. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature*

Microbiology, 903:1–10, 2017. ISSN 2058-5276. doi: 10.1038/s41564-017-0012-7. URL http://www.nature.com/articles/s41564-017-0012-7.

Jeremy Parr, Michael Smith, and Rod Shaw. Wastewater treatment options. *WELL - Water and Environmental Health at London and Loughborough*, pages 125–128, 1999. URL http://www.lboro.ac.uk/well/resources/technical-briefs/64-wastewater-treatment-options.pdf.

Ravi K. Patel and Mukesh Jain. NGS QC toolkit: A toolkit for quality control of next generation sequencing data. *PLoS ONE*, 7(2), 2012. ISSN 19326203. doi: 10.1371/journal.pone.0030619.

Yongzhen Peng and Guibing Zhu. Biological nitrogen removal with nitrification and denitrification via nitrite pathway. *Applied Microbiology and Biotechnology*, 73(1):15–26, 2006. ISSN 01757598. doi: 10.1007/s00253-006-0534-z.

James R. Pfafflin and Edward N. Ziegler. *Encyclopedia of Environmental Science and Engineering*. 2006. ISBN 9780849398438. doi: 10.1126/science.212.4499.1131-a.

S Pfister, A Koehler, and S Hellweg. Assessing the Environmental Impact of Freshwater Consumption in Life Cycle Assessment. *Environmental Science & Technology*, 43(11):4098–4104, 2009. ISSN 0013936X. doi: 10.1021/es802423e.

R. W. Portmann, J. S. Daniel, and A. R. Ravishankara. Stratospheric ozone depletion due to nitrous oxide: influences of other gases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1593):1256–1264, 2012. ISSN 0962-8436. doi: 10.1098/rstb.2011.0377. URL http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2011.0377.

P.R Postel, S.L.; Daily, G.C.; Ehrlich. Human Appropriation of Renewable Fresh Water Author (s): Sandra L . Postel , Gretchen C . Daily , Paul R . Ehrlich Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.org/stable/2889886. *Science*, 271(5250):785–788, 2008. URL http://www.jstor.org/stable/288988.

Christopher Quince, Thomas P Curtis, and William T Sloan. The rational exploration of microbial diversity. *The ISME Journal*, 2(10):997–1006, 2008. ISSN 1751-7362. doi: 10.1038/ismej.2008.69. URL http://www.nature.com/doifinder/10.1038/ismej.2008.69.

Christopher Quince, Alan W Walker, Jared T Simpson, Nicholas J Loman, and Nicola Segata. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 35(9):833–844, 2017. ISSN 1087-0156. doi: 10.1038/nbt.3935. URL http://www.nature.com/doifinder/10.1038/nbt.3935.

Nancy N Rabalais. Nitrogen in aquatic ecosystems. *AMBIO: A Journal of the Human Environment*, 31(2):102–112, 2002.

R Ramalho. *Introduction to Wastewater Treatment Processes*. Elsevier Science, 2012. ISBN 9780323151177. URL https://books.google.pt/books?id=r96wS3W0-sUC.

Ravi Ranjan, Asha Rani, Ahmed Metwally, Halvor S. McGee, and David L. Perkins. *Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing*, volume 469. Elsevier Ltd, 2016. ISBN 3124133382. doi: 10.1016/j.bbrc.2015.12.083. URL http://dx.doi.org/10.1016/j.bbrc.2015.12.083.

N P Revsbech and J Sørensen. *Denitrification in Soil and Sediment*. F.E.M.S. Symposium Series. Springer US, 2013. ISBN 9781475799699. URL https://books.google.pt/books?id=e4fdBwAAQBAJ.

Luana Marta de Almeida Rufino, Edenio Detmann, Daiany Íris Gomes, William Lima Santiago dos Reis, Erick Darlisson Batista, Sebastião de Campos Valadares Filho, and Mário Fonseca Paulino. Intake, digestibility and nitrogen utilization in cattle fed tropical forage and supplemented with protein in the rumen, abomasum, or both. *Journal of Animal Science and Biotechnology*, 7(1):11, 2016. ISSN 2049-1891. doi: 10.1186/s40104-016-0069-9. URL http://www.jasbsci.com/content/7/1/11.

R. A. Sanford, D. D. Wagner, Q. Wu, J. C. Chee-Sanford, S. H. Thomas, C. Cruz-Garcia, G. Rodriguez, A. Massol-Deya, K. K. Krishnani, K. M. Ritalahti, S. Nissen, K. T. Konstantinidis, and F. E. Loffler. Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. *Proceedings of the National Academy of Sciences*, 109(48):19709–19714, 2012. ISSN 0027-8424. doi: 10.1073/pnas.1211238109. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1211238109.

D.W. Schindler and J.R. Vallentyne. *The Algal Bowl: Overfertilization of the World's Freshwaters and Estuaries*. University of Alberta Press, 2008. ISBN 9780888644848. URL https://books.google.pt/books?id=xDwUAQAAIAAJ.

Ingo Schmidt, Olav Sliekers, Markus Schmid, Eberhard Bock, John Fuerst, J. Gijs Kuenen, Mike S M Jetten, and Marc Strous. New concepts of microbial treatment processes for the nitrogen removal in wastewater. *FEMS Microbiology Reviews*, 27(4):481–492, 2003. ISSN 01686445. doi: 10.1016/S0168-6445(03)00039-1.

Robert Schmieder and Robert Edwards. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6):863–864, 2011. ISSN 13674803. doi: 10.1093/bioinformatics/btr026.

Carsten J Schubert, Edith Durisch-kaiser, Bernhard Wehrli, Bo Thamdrup, Phyllis Lam, and Marcel M M Kuypers. Anaerobic ammonium oxidation in a tropical freshwater system (Lake Tanganyika). *Environ Microbiol*, 8(10):1857–1863, 2006. doi: 10.1111/j.1462-2920.2006.001074.x.

S Scott Shannon. Ammonia removal through simultaneous nitritation, anammox, and denitrification processes in recirculating vertical subsurface flow wetlands. 2016.

Vaclav Smil. Detonator of the population explosion. *Nature*, 400(July):415–415, 1999. ISSN 00280836. doi: 10.1038/22672.

Val H Smith. Eutrophication of freshwater and coastal marine ecosystems a global problem. *Environmental Science and Pollution Research*, 10(2):126–139, 2003.

P. Sobieszuk and K. W. Szewczyk. Estimation of (C/N) ratio for microbial denitrification. *Environmental Technology*, 27(1):103–108, 2006. ISSN 09593330. doi: 10.1080/09593332708618624.

R H Socolow. Nitrogen management and the future of food: lessons from the management of energy and carbon. *Proceedings of the National Academy of Sciences of the United States of America*, 96(11):6001–6008, 1999. ISSN 0027-8424. doi: 10.1073/pnas.96.11.6001.

Daan R. Speth, Michiel H. in 't Zandt, Simon Guerrero-Cruz, Bas E. Dutilh, and Mike S. M. Jetten. Genome-based microbial ecology of anammox granules in a full-scale wastewater treatment system. *Nature Communications*, 7:11172, 2016. ISSN 2041-1723. doi: 10.1038/ncomms11172. URL http://www.nature.com/doifinder/10.1038/ncomms11172.

J I Sprent. *The Ecology of the Nitrogen Cycle*. Cambridge Studies in Ecology. Cambridge University Press, 1987. ISBN 9780521310529. URL https://books.google.pt/books?id=GuVTFpNLGDwC.

David A. Stahl and José R. de la Torre. Physiology and Diversity of Ammonia-Oxidizing Archaea. *Annual Review of Microbiology*, 66(1):83–101, 2012. ISSN 0066-4227. doi: 10.1146/annurev-micro-092611-150128. URL http://www.annualreviews.org/doi/10.1146/annurev-micro-092611-150128.

M Strous, J J Heijnen, J G Kuenen, and M S M Jetten. The sequencing batch reactor as a powerful tool for the study of slowly growing anaerobic ammonium-wxidizing microorganisms. *Appl Micorbial. biotechnol*, 50:589–596, 1998. ISSN 0175-7598. doi: 10.1007/s002530051340.

Marc Strous, Jean Pelletier, Mike S. M. Jetten, Michael Wagner, and Denis Le Paslier. Deciphering the evolution and metabolism of an anammox bacterium from a community

genome. *Nature*, 440(7085):790–794, 2006. ISSN 0028-0836. doi: 10.1038/nature04647. URL http://www.nature.com/doifinder/10.1038/nature04647.

Emma C. Suddick, Penelope Whitney, Alan R. Townsend, and Eric A. Davidson. The role of nitrogen in climate change and the impacts of nitrogen-climate interactions in the United States: Foreword to thematic issue. *Biogeochemistry*, 114(1-3):1–10, 2013. ISSN 01682563. doi: 10.1007/s10533-012-9795-z.

Mark A. Sutton, Oene Oenema, Jan Willem Erisman, Adrian Leip, Hans van Grinsven, and Wilfried Winiwarter. Too much of a good thing. *Nature*, 472(7342):159–161, 2011. ISSN 0028-0836. doi: 10.1038/472159a. URL http://www.nature.com/doifinder/10.1038/472159a.

Torsten Thomas, Jack Gilbert, and Folker Meyer. Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2(1):3, 2012. ISSN 2042-5783. doi: 10.1186/2042-5783-2-3. URL http://microbialinformaticsj.biomedcentral.com/articles/10.1186/2042-5783-2-3.

Kenneth N. Timmis and Diemar H. Pieper. Bacteria designed for bioremediation. *Trends in Biotechnology*, 17(5):201–204, 1999. ISSN 01677799. doi: 10.1016/S0167-7799(98)01295-5. URL http://linkinghub.elsevier.com/retrieve/pii/S0167779998012955.

Satoshi Tsuneda, Makio Mikami, Yuzuru Kimochi, and Akira Hirata. Effect of salinity on nitrous oxide emission in the biological nitrogen removal process for industrial wastewater. *Journal of Hazardous Materials*, 119(1-3):93–98, 2005. ISSN 03043894. doi: 10.1016/j.jhazmat.2004.10.025.

UN WWAP. Water for People Water for Life. *Water*, page 36, 2003. ISSN 1098-6596. doi: 10.1017/CBO9781107415324.004.

U.S. EPA. Wastewater Technology Fact Sheet Package Plants. *United States Environmental Protection Agency*, pages 1–7, 2000. ISSN <null>. doi: EPA832-F-99-062. URL http://www3.epa.gov/npdes/pubs/final{_}sgrit{_}removal.pdf.

Maartje A H J van Kessel, Daan R Speth, Mads Albertsen, Per H Nielsen, Huub J M Op den Camp, Boran Kartal, Mike S M Jetten, and Sebastian Lücker. Complete nitrification by a single microorganism. *Nature*, 528(7583):555–559, 2015. ISSN 0028-0836. doi: 10.1038/nature16459. URL http://www.nature.com/doifinder/10.1038/nature16459.

Emilio Jimenez Vicente and Dennis R. Dean. Keeping the nitrogen-fixation dream alive. *Proceedings of the National Academy of Sciences*, 114(12):3009–3011, 2017. ISSN 0027-8424. doi: 10.1073/pnas.1701560114. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1701560114.

E I P Volcke, P Vanrolleghem, and M C M van Loosdrecht. *Modelling, analysis and control of partial nitritation in a SHARON reactor*. 2006. ISBN 9059891082.

John Vollmers, Sandra Wiegand, and Anne Kristin Kaster. *Comparing and evaluating metagenome assembly tools from a microbiologist's perspective - Not only size matters!*, volume 12. 2017. ISBN 1932-6203 (Electronic)\r1932-6203 (Linking). doi: 10.1371/journal.pone.0169662.

B B Ward, D J Arp, and M G Klotz. *Nitrification*. American Society Mic Series. ASM Press, 2011. ISBN 9781555814816. URL https://books.google.pt/books?id={_}QmsBAAAQBAJ.

Bess B. Ward and Marlene M. Jensen. The microbial nitrogen cycle. *Frontiers in Microbiology*, 5(October):2–3, 2014. ISSN 1664-302X. doi: 10.3389/fmicb.2014.00553. URL http://journal.frontiersin.org/article/10.3389/fmicb.2014.00553/abstract.

David M Whitacre. *Reviews of Environmental Contamination and Toxicology*, volume 225. 2013. ISBN 978-1-4614-6469-3. doi: 10.1007/978-1-4614-6470-9. URL http://link.springer.com/10.1007/978-1-4614-6470-9.

Kim Windey, Inge De Bo, and Willy Verstraete. Oxygen-limited autotrophic nitrification-denitrification (OLAND) in a rotating biological contactor treating high-salinity wastewater. *Water Research*, 39(18):4512–4520, 2005. ISSN 00431354. doi: 10.1016/j.watres.2005.09.002.

Derrick E Wood, Steven L SL Salzberg, C Venter, N Remington, B Ondov, N Bergman, and A Phillippy. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome biology*, 15(3):R46, 2014. ISSN 1474-760X. doi: 10.1186/gb-2014-15-3-r46. URL http://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-3-r46{%}5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/24580807{%}5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4053813.

Sitao Wu, Zhengwei Zhu, Liming Fu, Beifang Niu, and Weizhong Li. WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics*, 12(1):444, 2011. ISSN 1471-2164. doi: 10.1186/1471-2164-12-444. URL http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-12-444.

Yu-Wei Wu, Yung-Hsu Tang, Susannah G Tringe, Blake A Simmons, and Steven W Singer. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome*, 2(1):26, 2014. ISSN 2049-2618. doi: 10.1186/2049-2618-2-26. URL http://www.microbiomejournal.com/content/2/1/26.

Hongbin Yin and Ming Kong. Simultaneous removal of ammonium and phosphate from eutrophic waters using natural calcium-rich attapulgite-based versatile adsorbent. *Desalination*, 351(August):128–137, 2014a. ISSN 00119164. doi: 10.1016/j.desal.2014.07.029. URL http://dx.doi.org/10.1016/j.desal.2014.07.029.

Hongbin Yin and Ming Kong. Simultaneous removal of ammonium and phosphate from eutrophic waters using natural calcium-rich attapulgite-based versatile adsorbent. *Desalination*, 351:128–137, 2014b.

Hyungseok Yoo, Kyu Hong Ahn, Hyung Jib Lee, Kwang Hwan Lee, Youn Jung Kwak, and Kyung Guen Song. Nitrogen removal from synthetic wastewater by simultaneous nitrification and denitrification (SND) via nitrite in an intermittently-aerated reactor. *Water Research*, 33(1):145–154, 1999. ISSN 00431354. doi: 10.1016/S0043-1354(98)00159-6.

Bin Zhao, Meng Tian, Qiang An, Jun Ye, and Jin Song Guo. Characteristics of a heterotrophic nitrogen removal bacterium and its potential application on treatment of ammonium-rich wastewater. *Bioresource Technology*, 226:46–54, 2017. ISSN 18732976. doi: 10.1016/j.biortech.2016.11.120. URL http://dx.doi.org/10.1016/j.biortech.2016.11.120.

Yong zhen PENG, Yong MA, and Shu ying WANG. Denitrification potential enhancement by addition of external carbon sources in a pre-denitrification process. *Journal of Environmental Sciences*, 19(3):284–289, 2007. ISSN 10010742. doi: 10.1016/S1001-0742(07)60046-1.

Jizhong Zhou, Zhili He, Yunfeng Yang, and Ye Deng. High-Throughput Metagenomic Technologies for Complex Microbial Community Analysis: Open and Closed Formats. *CEUR Workshop Proceedings*, 1542(1):33–36, 2015a. ISSN 2150-7511. doi: 10.1128/mBio.02288-14.Copyright.

Qian Zhou, Xiaoquan Su, and Kang Ning. Assessment of quality control approaches for metagenomic data analysis. *Scientific Reports*, 4(1):6957, 2015b. ISSN 2045-2322. doi: 10.1038/srep06957. URL http://www.nature.com/articles/srep06957.

Wanding Zhou, Tenghui Chen, Hao Zhao, Agda Karina Eterovic, Funda Meric-Bernstam, Gordon B. Mills, and Ken Chen. Bias from removing read duplication in ultra-deep sequencing experiments. *Bioinformatics*, 30(8):1073–1080, 2014. ISSN 14602059. doi: 10.1093/bioinformatics/btt771.

A

APPENDIX I

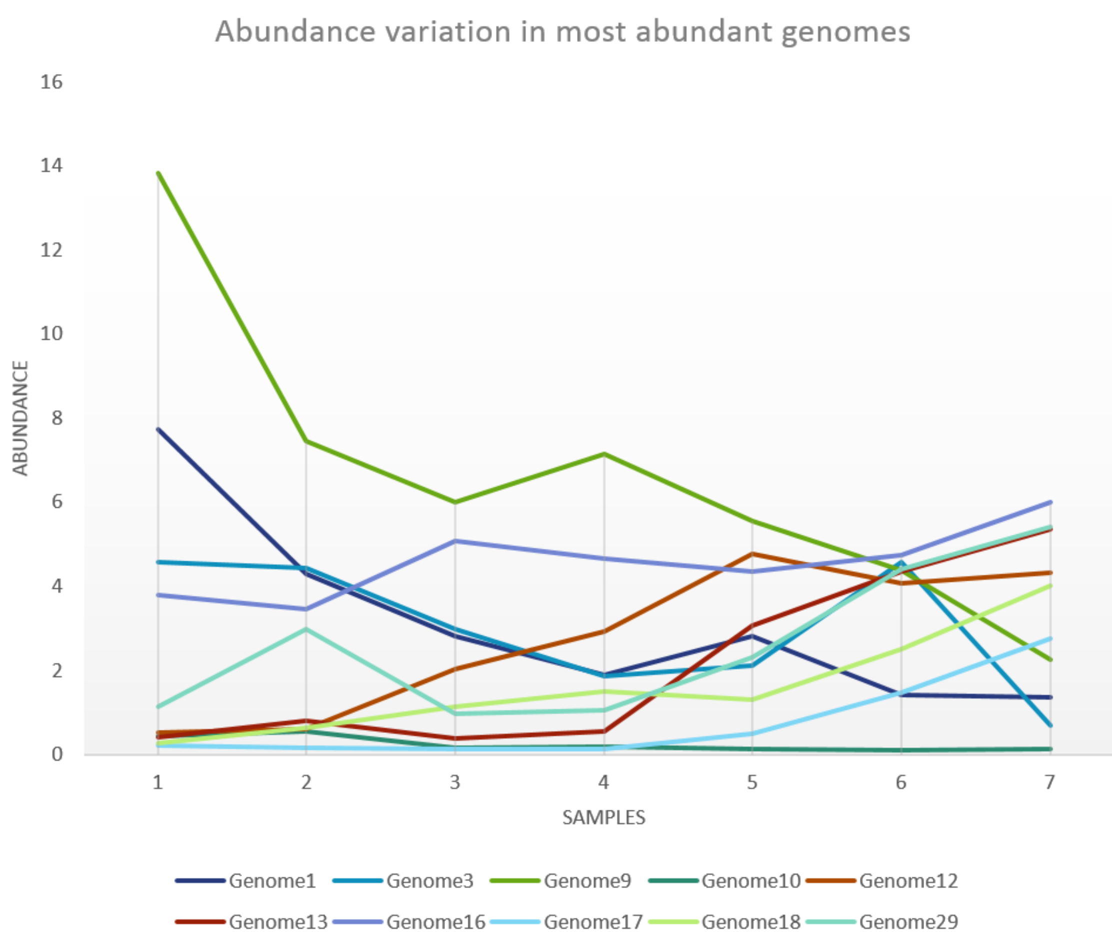


Figure 3.: A graphical representation of the variance of abundance throughout the time, in total six month, of the genomes displaying most fluctuation in abundance.

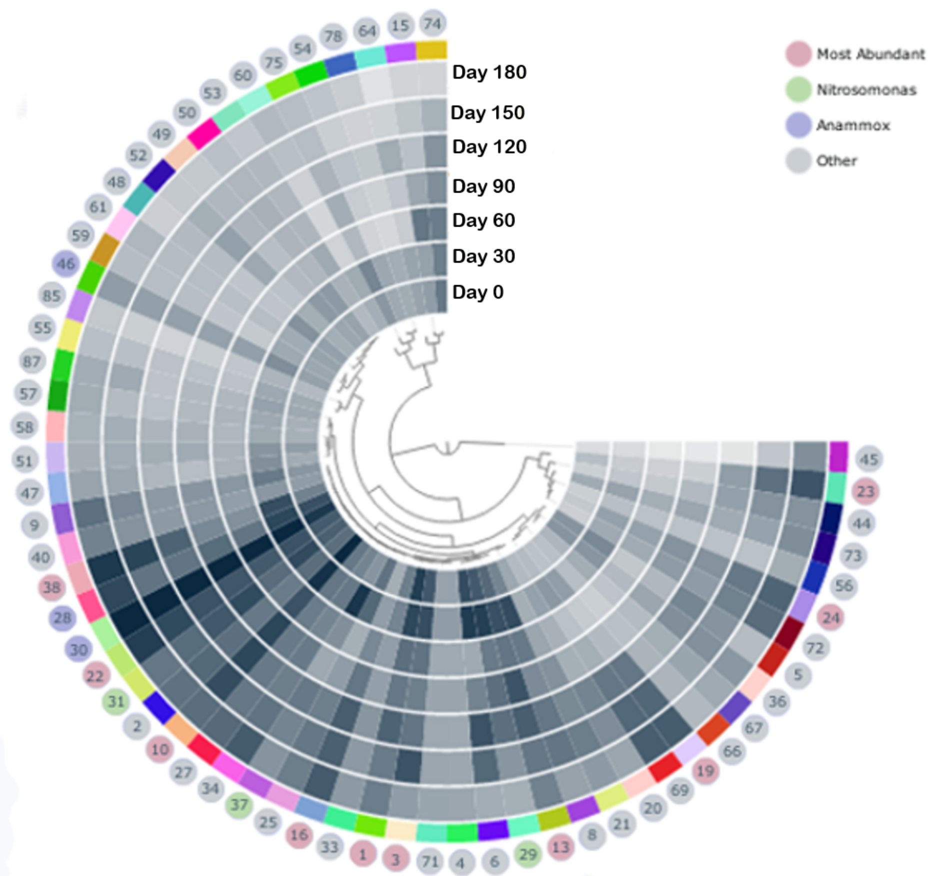


Figure 4.: A graphical representation of the genomes, organized by their variation of abundance throughout the samples. From genomes with decrease in abundance (left) to the most abundant(middle), ending with the MAGs with increase in abundance(right) Eren et al. (2015).

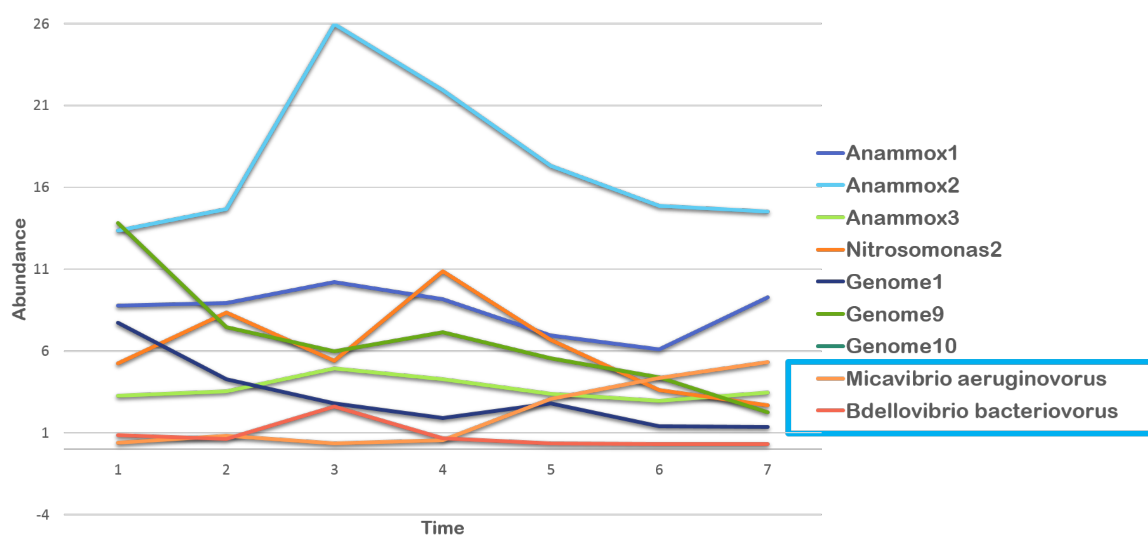


Figure 5.: A graphical representation of the decrease of certain organisms in abundance, while the two predators, highlighted by the blue rectangle, display growth.

NB: place here information about funding, FCT project, etc in which the work is framed. Leave empty otherwise.