Høgskolen i Innlandet

Faculty of Applied Ecology, Agricultural Sciences and Biotechnology

Sverre Branders

Masters Thesis

Analysis of the microbiota composition and gene expression in a phosphorus accumulating biofilm operating in a novel waste water treatment process

Masters in applied experimental biotechnology

2021 - 2022

Acknowledgement

I would like to thank Anders Øfsti and Torgeir Saltnes from Hias for their initiative in establishing the Hias research project in collaboration with HINN. Futhermore, I want to thank them for their accomodation and assistance in research activities at the Hias waste water plant. I would like to thank Knut Rudi and the team at NMBU including Inga Leena Angell for opening up their laboratory facillities and accomodate the use of their established sequencing library preparation pipeline and their guidance in doing so. In addition I want to thank them for the sequencing data processing preformed at NMBU and in general for their collaboration in the Hias research project. I want to thank HINN for the opportuniy to partake in this research and the facillitation of this master thesis project. I would like to thank Rob Willson for his assistance in operation and maintenance of laberatory equipment used in this thesis. Most notably I would like to thank Didrik Villard for the close collaboration and suport throughout this study. Finally and above all I would like to sincerely thank Wenche Johansen for her supervision, guidance and support throughout the study and the thesis writing process.

Abbreviations

Escherichia coli; *E.coli*, 16
Paracoccus pantotrophus; *P. pantotrophus*, 16

ammonium monooxygenase; AMO, 13
ammonium oxidising archae; AOA, 11
ammonium oxidising bacteria; AOB, 11, 31
amplicon sequence variants; ASVs, 20
analysis of variance; Anova, 31

biological nutrient removal; BNR, 10

complete ammonium oxidation; comammox, 11, 14
copper-dependent membrane monooxygenases; Cu-MMOs, 13
copper-dependent nitrite reductase; NirK, 17
cycle threshold; Ct, 37
cytochrome cd_1-containing nitrate reductase; NirS, 17

denitrifying phosphate accumulating organism; DPAO, 10
dissimilative nitrogen reduction to ammonia; DNRA, 15

enhanced biological phosphorus removal; EBPR; 9
exact sequence variants; ESVs, 20
extracellular polysaccharide matrix; EPS, 41

fluorescence in situ hybridization; FISH, 19

glycogen accumulating organisms; GAO, 31

heterotrophic nitrifying bacteria; HNB, 21
hydrazine synthase gene; *hzsB*, 22
hydroxylamine oxidoreductase; HAO, 13
hydroxylamine ubiquinone redox module; HURM, 13

membrane bound nitrate reductase; NAR, 16
molybdenum bis-molybdopterin guanine dinucleotide; $Mo[MGD]_2$, 16
moving bed biofilm reactor; MBBR, 10

nitric oxide oxidoreductase; NOO, 13
nitric oxide reductase; NOR, 17
nitrite oxidising bacteria; NOB, 11, 31
nitrite oxidoreductase complex; NXR, 14
no reverse transcriptase; no-RT, 37
non-metric multidimensional scaling; NMDS, 34

operational taxonomic unit; OTU, 20

particulate methane monooxygenase; pMmo, 13
periplasmic nitrate reductase; NAP, 16
phosphate accumulating organisms; PAO, 9, 31
polyhydroxybutyrate; PHB, 9
prokaryotic small ribosomal subunit RNA; 16S rRNA, 19

quantile-quantile; QQ, 114
quantitative polymerase chain reaction; qPCR, 19

reverse transcriptase; RT, 25

simple Z-type nitrous oxide reductase; NosZ, 18
simultaneous nitrification and denitrification; SND, 18
soluble chemical oxygen demand; SCOD, 10, 42

volatile fatty acids; VFAs, 9

zero-radius OTUs; zOTUs, 20

Table of Contents

Abstract

The Hias process is a novel biofilm based waste water treatment process designed for phosphorus removal and recovery. The process has also demonstrated significant nitrogen removal potential. The aim of this study was to investigate the microbial composition of the Hias biofilm throughout the reactor and how it relates to its functioning in the process. In addition, the expression of genes involved in nitrogen removal was evaluated as well as whether their expression is correlated with chemical components measured in the waste water. A sample handling method was developed for high quality nucleic acid extraction from the Hias biofilm. 16S rDNA sequencing revealed microbial composition of the biofilm to be stable while species evenness decreased throughout the Hias process. Relative abundance of ammonium oxidising bacteria increased in aerated parts of the Hias reactor while abundance of phosphate accumulating organisms decreased. Real time quantitative PCR showed no correlation between *amoA* transcript levels and decreasing NH_4 concentration.

1. Introduction

1.1 Biological Nutrient Removal

Human activity such as use of chemical fertilisers has caused high phosphorus and nitrogen levels in many bodies of surface water. Phosphorus stimulates fast growth of algae such as toxic *cyanobacteria*, causing eutrophication. Nitrogen containing compounds can be harmful to the aquatic biotope because of both inherent toxicity as well as by causing eutrophication. Therefore it is important to remove the excess of these compounds from waste water before releasing it into the environment. Chemical methods for nitrogen removal such as ion-exchange, chemical precipitation, etc. can be highly effective at removing nitrogen and are often used for water with high nitrogen loads but can be expensive and can require great amounts of chemicals and processing of waste products [78, 109]. Chemical phosphorus removal often uses chemical precipitation with metal salts. This process achieves low residual phosphate levels at the cost of the salts added. In addition, as ecological concers raise the demand for sustainability [38, 94], there can be additional cost associated with recovering phosphorus. Biological nutrient removal methods offer a sustainable, often cost effective alternative; biological nitrogen removal is an effective and inexpensive process that is widely adopted in waste water treatment with lower nitrogen loads where the lack of chemicals used in biological nitrogen removal can add to its sustainability and economic potential, as long as nitrogen discharge limits can be reached. Using nitrifying organisms (such as ammonium oxidising bacteria, AOB and archaea, AOA, complete ammonium oxidising bacteria, also known as comammomx bacteria and nitrite oxidising bacteria, NOB) to aerobically convert ammonium to nitrite or nitrate and subsequently denitrifying organisms to converte these products anoxically ultimately to gaseous nitrogen. These processes however are slow as responsible organisms exhibit slow growth, with ammonia oxidation being widely recognised as the rate-limiting step [24, 48]. Many different methods have been developed to increase the efficiency of nitrogen removal by, for example, using plastic discs as carriers for biofilm containing nitrogen metabolising organisms. However, many of these methods preform poorly when facing influent with higher nitrogen levels, sparking further innovation to overcome these limitations [48, 111].

Enhanced biological phosphorus removal (EBPR) emplois phosphate accumulating organisms (PAO) in a sludge based process to remove phosphate from waste water. Typically this process starts anaerobically, facillitating uptake of carbon substrate compounds such as volatile fatty acids (VFAs) by PAO. VFAs are stored intracellulary (as an energy reserve) as polyhydroxybutyrate (PHB) at the expence of energy generated through polyphosphate hydrolysis. This energy metabolism causes the release of PO_4 into the waste water [75, 81]. Anaerobic uptake of VFAs gives PAO a growth advantage over other organisms looking to take advantage of carbon under subsequent aerobic conditions. Under aerobic conditions PAOs then use their internally stored carbon to fuel cell growth while taking up phosphate to replenish their internal polyphosphate levels [69]. As a result of this growth, more phosphate

is taken up in aerobic growth than was initially released during anaerobic VFA uptake.

Biological nitrogen and phosphorus removal can be combined into a single process (biological nutrient removal (BNR)). Typically this requires three types of reactor zones; an initial anaerobic zone to provide an advantage to PAOs, typically followed by an anoxic zone for denitrification and an aerated zone for phosphate removal as well as nitrification where nitrogen removal can be facillitated through gas stripping [37]. However simultaneous BNR has also been observed in fully aerobic reactors where conditions are right to create anaerobic micro environments in biofilm or flocs [20].

1.2 The Hias process

In search of a more economic and sustainable method for phosphorus recovery and removal from waste water, the Hamar municipality waste water treatment plant (Hias, Hamar, Norway) developed a process combining traditional sludge based enhanced biological phosphorus removal (EBPR) with the cost effective, low process volume moving bed biofilm reactor (MBBR) technology. As shown in Figure 1.1, the process start with an inlet for pretreated waste water (1.1a) into an anaerobic reactor containing floating biofilm carriers. This reactor facilitates uptake of volatile carbon by PAO resulting in released PO_4. Then the waste water and biofilm carriers move to an aerobic reactor zone (1.1b) where PAO take up high amounts of PO_4 from the water to recover polyphosphate levels, subsequently using up more PO_4 than initially released in the aerobic zones, resulting in a net removal of phosphate to less than 0.20 mg/l $PO_4 - P$. The byproduct of this high phosphate uptake is high PAO cell growth. Using filters, the excess biomass material which does not remain attached to the biofilm carrier material can be recovered from the waste water. After processing, this biomass can be used as P-fertiliser. The reactor is divided into ten holding zones to achieve optimal rate of water and biofilm carrier flow throughout the reactor. Finally, the carriers are moved over a conveyor belt back to the first anaerobic zone (1.1c), whereas the treated water is discharged to a sludge recovery process (1.1d), meaning the biofilm carriers are recycled, whereas the waste water goes through the reactor only once [41]. The phosphate recovery process used has a higher potential for use of recovered phosphorus as fertiliser than widely used precipitation with metal salts. This way the Hias process achieves more than 95% phosphate removal and 80% soluble chemical oxygen demand (SCOD) removal with high phosphorus recovery [85].

While the process was developed with phosphorus removal in mind, nitrification and denitrification have been observed with up to 40% nitrogen removal without increasing cost. Though nitrogen discharge for this treatment plant is not regulated, this combined capability of nitrogen and phosphorus removal in a single reactor may be attractive in terms of sustainability as well as cost of waste water treatment. Furthermore, occurrence of nitrification and denitrification may have additional benefits. Lab scale experimentation indicates denitrification rates are constant under aerobic and anaerobic conditions and occur by consuming carbon stored under anaerobic conditions [85]. This suggests phosphate accumulating organisms could be present that use nitrogen instead of oxygen as final electron acceptor (denitrifying PAOs, DPAOs) that can aid in phosphorus as well as nitrogen removal and may have a growth advantage over other organisms in the initial anaerobic reactor zones.

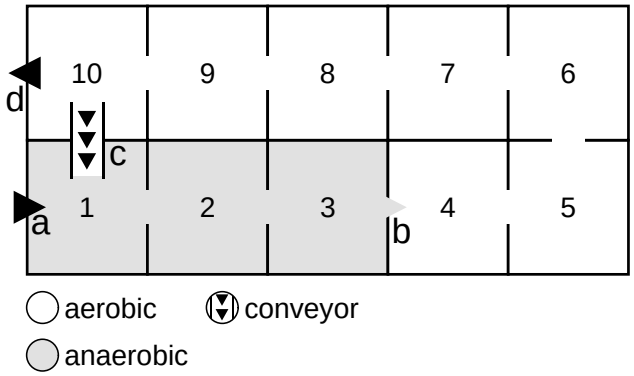


Figure 1.1: Schematic overview of the Hias process, including inlet stream for pre-treated water (a), transition for carrying water and biofilm carriers from an anaerobic (grey) to an aerobic (white) zone (b), a method for moving the carriers back to the anaerobic zone without the water (c) and an outlet stream for the processed water (d). (Adapted from [41, 82])

Better insight into the biology of the biofilm may help in reaching higher nitrogen removal potential and as a result ensure organisms can more efficiently consume waste water carbon as they accumulate phosphorus using nitrogen as final electron acceptor. [82, 85].

1.3 Nitrification

Nitrification is a three step process (as shown in Figure 1.2) that aerobically converts ammonia to nitrate preformed by two separate groups of organisms (with the exception of complete ammonium oxidation, comammox). Ammonia oxidising bacteria (AOB) or archae (AOA) oxidise ammonia to nitrite in two steps; ammonia oxidation and hydroxylamine oxidation. Under anaerobic conditions, AOB may reduce this nitrite to NO_2 or N_2 in denitrification. Alternatively, nitrite may be oxidised to nitrate by nitrite oxidising bacteria (NOB). All bacteria involved in nitrification (AOB and NOB) belong to the family of *Nitrobacteraceae*. [11, 12, 28, 40, 89, 93].

Figure 1.3 shows a schematic overview of how oxidations involved in nitrification may generally occur *in vivo* and of the key enzymes involved.

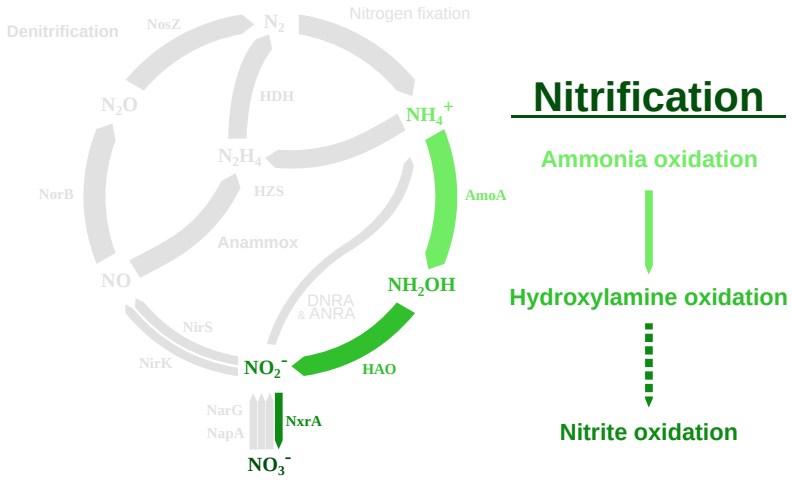


Figure 1.2: Overview of the major steps and catalysing enzymes involved in nitrification. In ammonia oxidising bacteria (AOB), ammonium monooxygenase subunit A (AmoA) is the catalysing enzyme in the first and hydroxylamine oxidoreductase (HAO) in the second stage of nitrification. Nitrite oxidising bacteria (NOB) preform the final oxidation in nitrification with the catalytic nitrite oxidoreductase complex subunit A (NxrA). Comammox bacteria are able to preform all three steps of nitrification. (Adapted from [21, 62, 91])

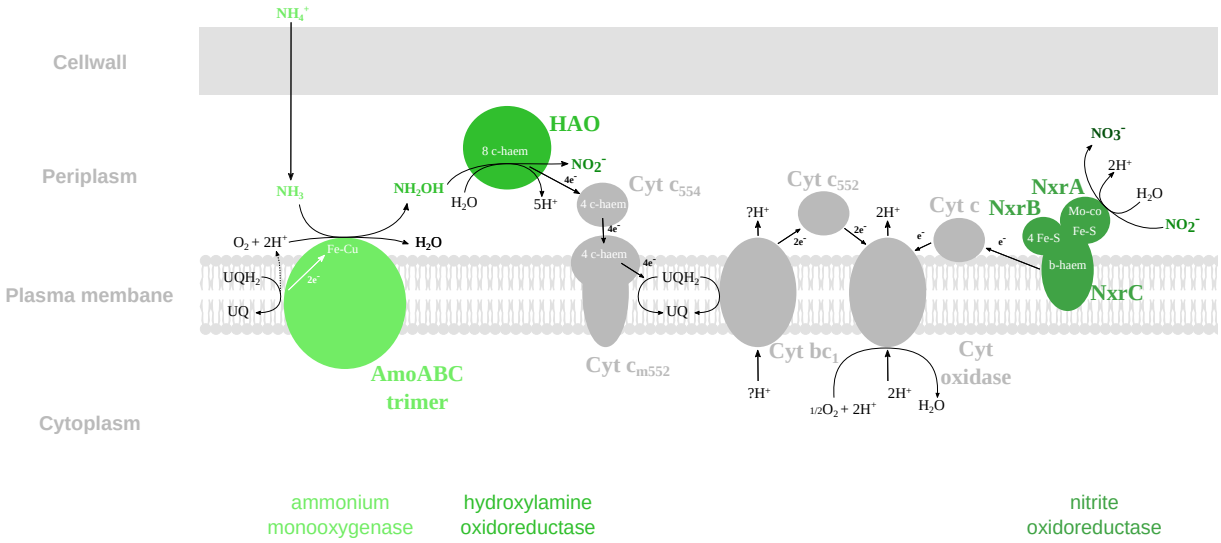


Figure 1.3: Schematic overview of key enzymes involved in the NH_4 oxidation methabolism and their cellular localisation. The catalysing enzyme complexes involved in conversion of ammonium to nitrate are represented in shades of green: the trimeric ammonium monooxygenase (AMO) complex catalysing ammonia oxidation, hydroxylamine oxidoreductase (HAO) oxidising hydroxylamine to nitrite and the nitrite oxidoreductase (NXR) complex oxidising nitrite to nitrate. Important structures such as haem groups involved in electron transfer and known catalytic sites are labeled and electron flow is indicated. In grey, accompanying enzymes involved in the respiratory chain are represented as well as the electron flow and chemical conversions involved in the nitrification respiratory chain according to current understanding. (Adapted from [4, 11, 15, 22, 30, 35, 40, 51, 61, 66, 83, 89, 105])

1.3.1 Ammonia oxidation

Though ammonia oxidation and hydroxylamine oxidation are two separate enzymatic steps, they are mutually dependent for energy generation; in the oxidation of NH_3 catalysed by ammonium monooxygenase (AMO), four electrons are required to oxidise O_2, two of which are supplied by NH_3 forming NH_2OH, the remaining two come from the subsequent hydroxylamine oxidation, leaving two electrons to power the cells metabolism. According to current understanding, the electrons generated by hydroxylamine oxidation are transported to the ubiquinol pool, from which AMO can draw its two required electrons to sustain the reaction [4, 40].

AMO is a membrane-bound hetero-trimeric Cu-dependent protein, consisting of three complexes each made up of three subunits; AmoA, AmoB and AmoC. In addition to *amoA*,*B* and *C*, two orthologous genes, *amoD* and *amoE*, have been found, encoding a membrane protein complex with unknown function. Difficulties in isolating these protein complexes means their structure has not been crystallographically confirmed. However, AMO is highly homologous to particulate methane monooxygenase (pMmo) and both encoded proteins belong to the same copper-dependent membrane monooxygenases (Cu-MMOs) superfamily, implying they serve similar biological functions. Homology modelling of AMO against pMMO shows a Cu_B site at the interface of AmoB and AmoC and a solvent accessible Cu_C site near the AmoC trimer interface. Though much is unknown about the reaction mechanisms of AMO, its structure implicates that one of these sites may be the active catalytic site [66, 89].

Genomic studies of AOA have shown they do not encode hydroxylamine detoxifying enzymes nor produce cytochrome c, prompting different models for potential archaeal AMO mechanisms compared to AOB, non however have been experimentally confirmed [89].

1.3.2 Hydroxylamine oxidation

Hydroxylamine oxidoreductase (HAO) is a complex periplasmic haem containing homotrimer. Each subunit contains eight haem groups, one of which, having unique properties, is the active site of hydroxylamine oxidation to nitrite. Some of the additional haem groups are presumed to transport the four generated electrons away from the active site. As noted before, the generated electrons likely move to the quinone pool were two of the electrons are used to sustain ammonia oxidation. To arrive in the quinone pool, the electrons likely pass through cytochrome *c*554 to cytochrome c_m552. Because genes encoding these proteins cluster in AOB, this group of proteins is known as the Hydroxylamine Ubiquinone Redox Module (HURM). Once the electrons are in the quinone pool, it has been shown that the two remaining electrons flow trough the cytochrome bc_1 complex. Finally these electrons are used to oxidise O_2 by cytochrome oxidase, adding to the proton gradient. The transfer of electrons from cyt bc_1 to cyt oxidase may be facilitated by the periplasmic cytochrome c_{552} [4, 30, 51, 89].

While it is known that AOB can produce NO under anaerobic conditions through denitrification, Caranto and Lancaster [11] showed production of NO under aerobic conditions by HAO. This finding implies another enzyme may be involved in ammonium oxidation to nitrite. Non have been identified definitively however, though Stein [93] proposes potential pathways based on candidate nitric oxide oxidoreductases (NOO). Interestingly, AOA don't

harbour HAO homologues while they have been shown to be abundant in some waste water treatment reactors and can contribute significantly to nitrite removal. The higher substrate affinity for ammonia of AOA compared to AOB means AOA show greater nitrogen removal potential under extreme or highly fluctuating environmental conditions [4, 30, 51, 57, 89].

Cytochrome bc_1

In eukaryotes, the cytochrome bc_1 complex serves a critical role in the cellular respiration pathway. By pumping protons from ubiquinol hydrogen atoms across the lipid bilayer membrane, the resulting proton gradient can be used to generate ATP. In this process, one electron is released to cytochrome c outside the membrane and one electron is released inside the membrane, perpetuating the cycle (known as the Q-cycle) [35]. This complex is also found in the plasma membrane of many bacteria and is involved in reduction pathways of different electron acceptors, nitrite being one of them [97]. Prokaryotic cytochrome bc_1 complexes typically contain the membrane protein cytochrome b and two proteins containing cofactors: cytochrome c_1 and the Riekse Fe/S protein [90].

1.3.3 Nitrite oxidation

The nitrite oxidoreductase complex (NXR) catalyses the oxidation of nitrite to nitrate in a reversible reaction releasing two electrons. These complexes have been shown to be genetically diverse, however in most aerobic nitrite oxidisers, they are membrane bound. NXR is composed of three subunits, with the NxrA subunit containing the catalytic molybdopterin-containing active site where oxidation of nitrite occurs. The NxrB subunit contains a chain of iron-sulphur clusters transporting electrons from (and to in the case of nitrate reduction) the active site to the b-haem in NxrC. Whereas the resulting electrons are used for nitrite reduction to nitric oxide by anammox bacteria, NOB use these electrons for oxygen reduction in energy generation. Generated electrons are likely transported from the NxrC b-haem to cytochrome c, and finally to a terminal cytochrome oxidase (such as cyt oxidase aa3). The shorter electron transport chain in NOB as compared to AOB is in accordance with the comparatively lower oxido-reduction potential of NO_2^-/NO_3^- [15, 61, 79, 105].

1.3.4 Complete ammonium oxidation

The separation of nitrification between two groups of organisms means these organisms often occur in co-aggregation. Kinetic studies have long theorised that complete nitrification in one organism would provide growth advantages, particularly in environments with low growth and nutrient concentrations, such as in biofilms. In 2015, Daims et al. [22] identified a *Nitrospira* bacteria oxidising ammonia to nitrate and harbouring *amo*, *hao* and *nxr* genes, confirming complete ammonium oxidation (comammox) can occur in one organism. Comammox *Nitrospira* cannot be differentiated by 16S rRNA-based identification and not all *Nitrospira* are capable of comammox. So far however, all comammox *Nitrospira* have been found to have a unique AmoA subunit which can be used to differentiate between comammox and canonical NOB. Koch, Kessel, and Lücker [51] found comammox bacteria often occur together with canonical ammonia oxidisers. Taken together with metabolic differences

observed through genomic studies, this implies comammox bacteria likely fill a specific niche in microbial communities. Higher ammonia affinity seems to confirm a growth advantage in low nutrient environments, even more than in (terrestrial) AOA. Sakoula et al. [83] isolated a new comammox *Nitrospira* species enrichment culture. in addition to high ammonia affinity, this culture showed higher affinity for nitrite. Furthermore, inhibition of ammonium oxidation occured at different ammonia concentration, implying differences in substrate tolerance between species of comammox bacteria. These findings suggest different comammox bacteria may occupy different niches based on their substrate tollerance and ammonia affinity.

1.4 Denitrification

Plants and many fungi as well as some bacteria use nitrate to grow in a process known as assimilative nitrate reduction. This process requires energy and results in assimilation of organic nitrogen. Some bacteria can use nitrate as an energy source by dissimilative nitrate reduction, using nitrogen as an alternative final electron acceptor under anaerobic conditions. The end product can be ammonia, in dissimilative nitrogen reduction to ammonia (DNRA), or N_2 in denitrification [6].

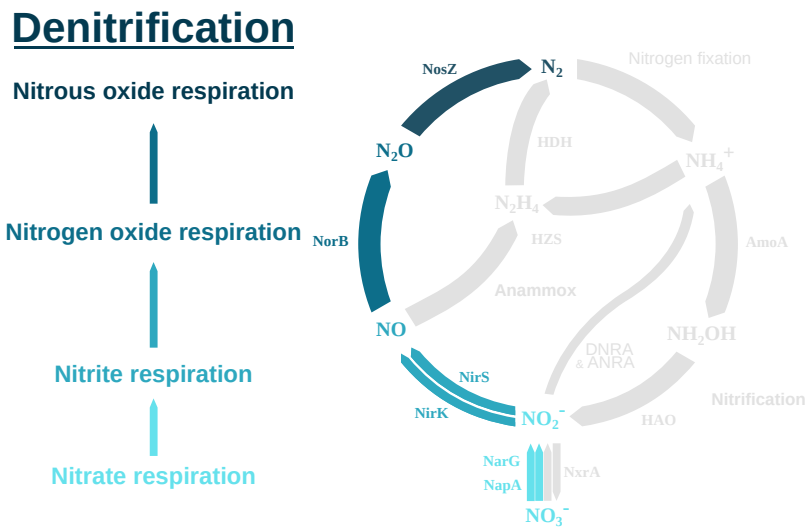


Figure 1.4: Overview of the major steps and catalysing enzymes involved in denitrification. Nitrate reduction to nitrite can be catalysed by membrane bound nitrate reducatase subunit G (NarG) or periplasmic nitrate reductase subunit A (NapA). The catalysis of nitrite to NO can be facilitated by copper containing (NirK) or cytochrome cd_1 containing (NirS) nitrite reductase. Nitric oxide reductase subunit B (NorB) catalyses NO respitration and Z-type nitous oxide reductase (NosZ) catalyses reduction of nitrous oxide to gaseous nitrogen. (Adapted from [21, 62, 91])

Denitrifying bacteria are a very diverse group, both in terms of biochemistry and in taxonomy, however the most commonly found denitrifiers belong to the genus *Pseudomonas* [50]. Denitrification typically occurs in four enzymatic reductions as shown in Figure 1.4: NO_3^- respiration, NO_2^- respiration, NO respiration and N_2O respiration. Figure 1.5 shows a schematic overview of how reductions involved in denitrification may generally occur *in vivo* and of the key enzymes involved.

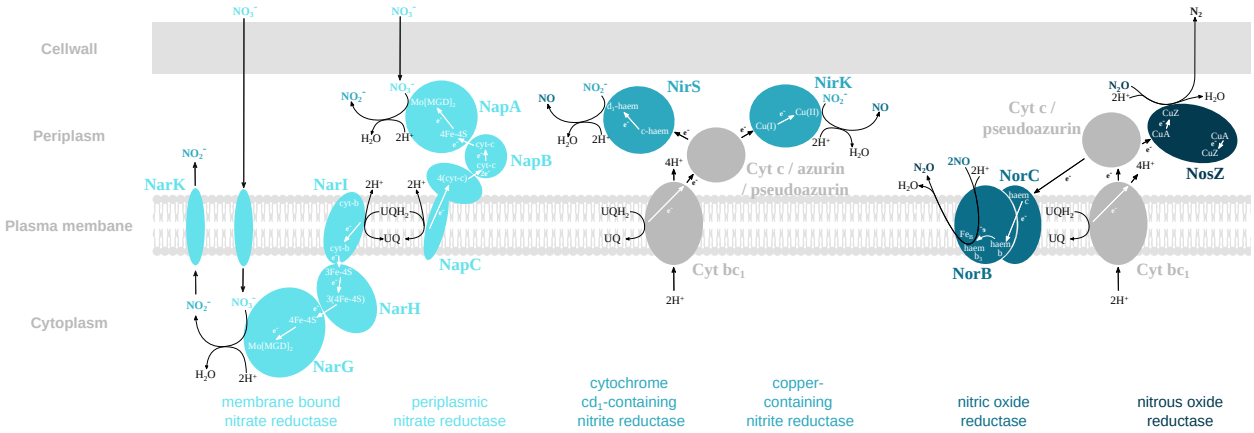


Figure 1.5: Schematic overview of key enzymes and reaction compounds involved in denitrification and their cellular localisation. The catalysing enzyme complexes involved in the reduction reactions of denitrification are represented in shades of blue: the G, H and I subunits making up the membrane bound nitrate reductase complex (NAR) reduce nitrate to nitrite, whereas NarK transports the reagent and reaction product of this step to and from the cytoplasm, in addition to NAR, the periplasmic nitrate reductase complex (NAP) can also reduce nitrate but takes on primarily a redox balancing role, cytochrome cd_1 (NirS) or copper containing (NirK) nitrite reductase catalyse the reduction of nitrite to NO, the nitric oxide reductase (NOR) complex catalyses reduction of NO to nitrous oxide and finally nitrous oxide reductase (NOS) catalyses the reduction of nitrous oxide to gaseous nitrogen. Important structures in electron transport and known active sites are labeled and electron flow is indicated. Other key proteins involved in the denitrification energy metabolism are represented in grey. (Adapted from [1, 6, 14, 19, 21, 27, 34, 42, 43, 44, 47, 52, 63, 65, 73, 89, 90, 104, 112])

1.4.1 Nitrate respiration

In *Paracoccus pantotrophus* (*P. pantotrophus*), two distinct dissimilative nitrate reductases have been characterised; membrane bound nitrate reductase (NAR) and periplasmic nitrate reductase (NAP). Both use molybdenum bis-molybdopterin guanine dinucleotide ($Mo[MGD]_2$) cofactors for nitrate reduction and obtain electrons by oxidising ubiquinol in the plasma membrane. NAR serves as the primary reductase in nitrate conversion and NAP takes on a redox balancing role, optimising growth rate [6, 34]. *Escherichia coli* (*E. coli*) expresses two different NAR proteins: NRA and NRZ, where NRZ is expressed constitutively and NRA is expressed under denitrifying conditions. NRZ is highly homologous to NRA and has very similar biological function. Furthermore it has been shown subunits of NRZ and NRA can form hybrid complexes [6]. In addition to *P. pantotrophus* and *E. coli*, NAR enzymes have been found in several other denitrifying and nitrate respirating bacteria [63].

Structurally, NAR enzymes consist of three subunits. NarI (subunit γ) is membrane bound and anchors the β and consequently α subunits to the membrane. In addition it oxidises ubiquinol from the plasma membrane quinone pool as it transfers electrons across the plasma membrane with its two haem-b groups on either side the membrane, releasing two protons into the periplasm. NarH (subunit β) is soluble and has internal iron-sulphur clusters (3Fe-4S and three 4Fe-4S), transporting electrons from NarI to NarG (subunit α). Another 4Fe-4S cluster in NarG accepts the electron from NarG and relays it to the $Mo[MGD]_2$ cofactor where nitrate is reduced using two protons from the cytoplasm. Other proteins involved in NAR include NarJ which helps in assembling or stabilizing the $\alpha\beta$ complex

before it is attached to NarI and activates the reductase after MGD synthesis, as well as NarK, transporting NO_3^- into and NO_2^- out of the cytoplasm [6, 34, 63].

Similarly to NAR, NAP consists of three subunits. The membrane bound NapC catalyses ubiquinol oxidation and mediates electron transfer to the tightly associated NapAB complex, serving a similar function to NarI in NAR. NapB contains two haem cyt-c groups relaying electrons through the iron-sulphur cluster in NapA to the $Mo[MGD]_2$ cofactor where catalysis of nitrate reduction occurs [34]. Ji et al. [44] found NAP almost exclusively in the phylum *Proteobateria* and its activity has been correlated with aerobic respiratory nitrate reduction. As stated, NAP is presumed to serve as an electron sink, maintaining periplasmic redox balance [6, 34].

1.4.2 Nitrite respiration

Typically, one of two nitrite reductases is found in denitrifying bacteria; cytochrome cd_1-containing nitrate reductase (NirS) or Cu-dependent nitrite reductase (NirK), with NirS being more common [108]. NirS is a dimeric protein containing a haem c covalently bound to its N-terminal domain and a haem d_1 non-covalently bound to the active site where nitrite reduction occurs [67]. In NirS proteins, the C-terminal domain interacting with d_1 haem is much more conserved than the N-terminal domain. In addition, the d_1 haem is found to be unique to denitrifiers [19]. In *Pseudomonas aeruginosa*, the electron required for nitrite reduction is carried from cytochrome bc_1 by cytochrome c_{551} *in vivo* [19, 46], whereas in *Paracoccus denitrificans*, this role is filled by cytochrome c_{550} or alternatively pseudoazurin [74].

In contrast to NirS, the NirK protein is trimeric. It contains three type II copper sites where nitrite reduction occurs, as well as four type I copper sites where electrons are received from pseudoazurin [65]. In addition to pseudoazurin and cytochrome c, azurin has been shown to play a role in transporting electrons from cytochrome bc_1 to nitrite reductase. The importance of azurin as an electron transporter in this process *in vivo* has been questioned however. Vijgenboom, Busch, and Canters [99] proposed it may play a role under stress conditions. Yuan, Liu, and Lu [108] studied the differential response of *nirS* and *nirK* genes in anoxic soil when adding nitrate and found *nirK* denitrifiers to be active at the start of anaerobic incubation and more *nirS* activity when anaerobic conditions were fully established.

1.4.3 NO respiration

NO is a highly reactive compound. Its toxicity against Fe and Cu containing proteins and reactivity with bacterial DNA makes it cytotoxic in high concentrations. Despite this, NO is an obligatory free intermediate in denitrification. Nitric oxide reductase (NOR) is essential in maintaining a low concentration of NO by reducing it to N_2O [52]. The NOR protein consists of a small (NorC) and a large (NorB) subunit. The NorC subunit contains a haem c which receives electrons from cytochrome c_{551} and relays them to the haem b in NorB. From there they are moved to the binuclear centre, consisting of haem b_3 and non-haem Fe_B, where reduction of nitric oxide is catalysed. The active site of NOR is buried inside the enzyme. Both nitric oxide and protons used in the reduction reaction originate from the

periplasmic side of the membrane and never cross the membrane. It has been proposed the reaction components are transported from the periplasmic region to the active site through a water channel and hydrogen-bonding network and a hydrophobic channel [42].

1.4.4 N_2O respiration

In the final step of denitrification, nitrous oxide is reduced to non reactive gaseous nitrogen by nitrous oxide reductase. Two families of N_2O reductases have been identified, however the three-domain Z-type reductase has only been found in a few organisms whereas simple Z-type nitrous oxide reductase (NosZ) is more common and better characterised. NosZ is a homodimer with each monomer consisting of two domains. The N-terminal domain binds the catalytic CuZ centre and the C-terminal domain binds the CuA centre which is responsible for electron transfer. The dimer is arranged such that the CuA and CuZ centres from opposing monomers are in close proximity to accommodate electron transfer. Similarly to NorC and nitrite reductases, CuA in NosZ receives electrons from small electron carrier proteins (such as pseudoazurin and cytochrome c_{550} in *Pa. denitrificans*, cytochrome c_{552} in *Ma. hydrocarbonoclasticus* and pseudoazurin in *Ac. cycloclastes*) and relays them to the CuZ centre where reduction to N_2 occurs [73]. N_2O is a potent greenhouse gas and can account for a large portion of the environmental impact of waste water treatment plants, particularly when employing biological nitrogen removal. While accumulation of N_2O by denitrifiers can occur as a result of NosZ activation being slower than other denitrifying enzymes at the beginning of anaerobic conditions or even because some denitrifiers simply do not reduce nitrogen all the way to N_2, this accumulation has not been found to result in significant emission of nitrous oxide. Instead most emission is a result of N_2O produced under aerated conditions where it is immediately released into the atmosphere as a result of the aeration [56].

1.5 Simultaneous Nitrification and Denitrification

Though nitrification requires oxygen and denitrification typically happens under anaerobic conditions, simultaneous nitrification and denitrification (SND) is a well known phenomenon which can occur for two reasons. Firstly, SND can occur as a result of an oxygen gradient in a biofilm or in microbial flocs caused by oxygen diffusion limitations in those structures. Second, SND can be a biological phenomenon; heterotrophic nitrifying bacteria have been shown to convert ammonium to nitrogen gas under aerobic conditions and organisms capable of aerobic denitrification have been found [44]. Notably, the NAP protein can be expressed under both aerobic and anaerobic conditions. Typically in biofilm based aerobic SND reactors, fast growing heterotrophic bacteria are most abundant in the outer layers of the biofilm, with nitrifying bacteria being found deeper inside the biofilm and denitrifying bacteria even further internally where conditions may be more anoxic. Because of this, competition for oxygen between nitrifying and heterotrophic bacteria can decrease nitrogen removal efficiency, whereas overly oxygenating can affect denitrification rates. Decreasing C/N ratio has been shown to increase relative abundance of nitrifying bacteria while lowering nitrification efficiency. [13, 32, 64]

1.6 Molecular methods for detection of micro organisms in complex communities

One method for studying micro organisms that contribute to nitrogen removal is to grow enrichment cultures of individual or groups of organisms, potentially giving much insight into pathways and kinetics in those organisms. Though many have already been successfully grown in enrichment cultures, some other nitrogen removing organisms are notoriously difficult to isolate [44, 59, 71, 83, 103].

When the goal is to evaluate a large community of organisms responsible for nitrogen removal, such as in a waste water treatment reactor, using this approach would be prohibitively time consuming. Therefore other approaches have been developed. One common approach is to detect the abundance of genes or transcripts of specific genes involved in nitrogen removal using methods such as quantitative polymerase chain reaction (qPCR). A prerequisite for the use of this approach is having access to highly specific primer pairs able to detect these targets, as the accuracy of the assay is strongly affected by the specificity of the primers used [24].

Alternatively, or in addition, sequencing approaches can be used. Meta sequencing methods can be used to detect genes or transcripts in the studied community directly. A cheaper and commonly used method to detect groups of prokaryotic micro organisms is 16S rDNA sequencing [33, 95]. When interested in the localisation of genes of interest or particular organisms in communities, methods such as fluorescence in situ hybridization (FISH) can be used [32].

1.6.1 16S rDNA and sequencig

Prokaryotic small ribosomal subunit RNA (16S rRNA) is functionally conserved and comprised of conserved, variable and hyper variable regions that evolve at different rates due to their secondary structure. As a result, their genes can be used as a universal marker for taxonomy. Designing primers targeting regions of varying conservation, allows for the detection of a set range of organisms; for example targeting specific genera or all organisms in a domain [106]. By sequencing PCR products obtained from these primer sets (16S rDNA sequencing), researchers can get a taxonomic overview of microbial communities of interest. In order to get an accurate picture of a community however, it is important to use primers sets able to amplify 16S rDNA of all micro organisms to be studied, without introducing bias. Additionally, other molecular biology methods can make use of these primer sets, such as qPCR for quantification of microbial DNA or cDNA, or FISH for localisation of micro organisms. Yu et al. [107] designed a universal primer pair (PRK341F/PRK806R) for qPCR amplification targeting the V3- and V4-hyper variable regions (see Figure 1.6) of bacterial and archaeal 16S rDNA. Results showed high specificity to the target and insignificant over- or underestimation of the target groups. Independent assessment of the PRK341F/PRK806R primer pair for the use of 16S rDNA sequencing showed the primer pair targets approximately 97.4% of bacterial and 93.8% of archaeal rRNA gene sequences in the Ribosomal Database Project database [95].

A limitation of this approach is the variable copy number of the 16S rRNA gene in

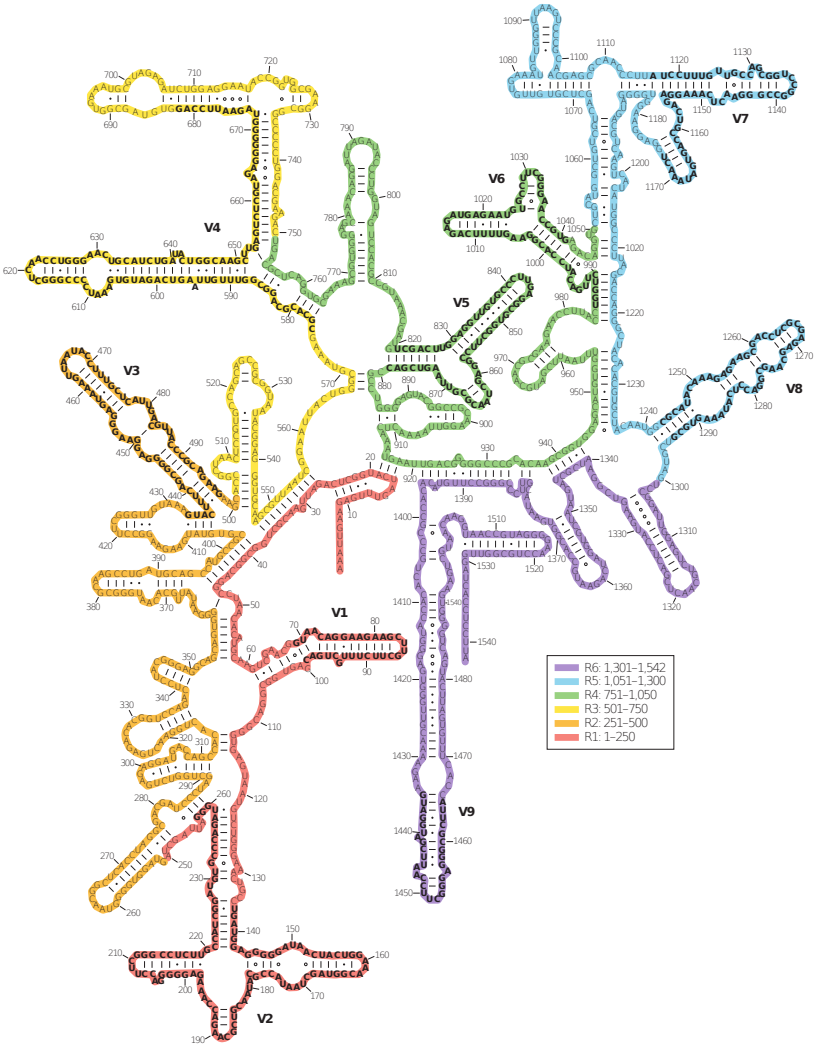


Figure 1.6: Secondary structure of *E. coli* 16S rRNA divided into gene segments (R) in different colours, according to the 9 known variable regions (V) [106].

prokaryotes [98]. While software approaches exist to attempt to correct for this variation, the accuracy of these corrections have been shown to be low [60]. Another limitation is the inherent error rate of any sequencing platform. In order to minimise the effect of these errors, similar sequences (typically 97% similarity) are clustered and their consensus sequence is represented as an operational taxonomic unit (OTU). This approach however may underestimate actual species richness by grouping (low abundance) sequences with similar but phylogenetically different sequences into a single OTU [86]. To avoid this, new methods have been developed to retain all unique sequences, while using algorithmic approaches to correct for sequencing errors through denoising and error correction. These include zero-radius OTUs (zOTUs), exact sequence variants (ESVs) and amplicon sequence variants (ASVs) [45]. While culturing of micro organisms is a slow process, the advances in sequencing technology have meant many previously unknown and uncultured organisms are being identified through sequencing. A lack of clear thresholds for taxonomic boundaries and biases introduced in clustering methods may however cause challenges in identifying and classifying

previously unknown organisms [86, 106]. Therefore, comprehensive databases of specific microbial communities are valuable. When studying communities of a specific type, such as those responsible for waste water treatment, using universal reference databases may lack important species. Thus the MiDas 4 database was constructed to provide better coverage of waste water microbial communities [25].

1.6.2 Other targets for detection of relevant micro organisms

In addition to 16S rDNA, other targets can be used for identification of organisms, when sequences are unique and ubiquitous for the studied group. Duan, Zhang, and Zheng [24] identified a group of primer pairs that has been widely used to detect AOB, AOA, comammox bacteria and anaerobic AOB (A.1). As Duan et al. note, autotrophic ammonium oxidising organisms are a group of low phylogenetic diversity with AOB belonging to the class of β- or γ-*proteobacteria*, AOA belonging to the phylum *Thaumarchaeota*, all anaerobic AOB being affiliated with the *Brocadiaceae* family of the phylum *Planctomycetes* and comammox bacteria all belonging to the *Nitrospira* sub lineage. Whereas heterotrophic nitrifying bacteria (HNB) are much more diverse.

1. The amoA-1F/amoA-2R primer pair was designed to target the *amoA* gene of *Nitrosomonas europaea* and was shown to be specific to the *amoA* gene of AOB of the β-*proteobacteria* subclass, as the assay resulted in amplification of targets form all tested AOB except for *Nitrosococcus oceanus*, which belongs to the γ subclass. Furthermore the primers have been shown to discriminate between *amoA* and the highly homologous *pmoA* [80]. However, as noted by Duan, Zhang, and Zheng [24], Lang et al. [55] among others reported amplification of some HNB *amoA* genes using this primer pair, whereas others HNB isolates result in no amplification with these primers. This suggests a bias of this primer pair towards certain groups of AOB and HNB.

2. Like Rotthauwe, Witzel, and Liesack [80], Kowalchuk et al. [53] aimed to study the β subgroup of AOB, but targeted 16S ribosomal RNA gene fragments instead of *amoA*. The resulting CTO189fA/B, CTO189fC / CTO654r primer pairs showed good specificity for the β subdivision of AOB.

3. Both of these primer pairs targeting AOB may overlook certain groups of AOB because of their specificity for the β subgroup. Therefore, Calvó and Garcia-Gil [8] used the *Nitrosomonas europaea amoB* gene sequence to design the amoBMf/amoBMr primer pair, claiming to target all AOB. While no fully conserved domains were identified, some amino acid motifs were found in AmoB of both subgroups of AOB. These motifs were used as targets for primer design. Comparing the primer pair to amoA-1F/amoA-2R however, showed 10 times more template is required to reach detectable levels of amplification with the primer pair targeting *amoB*.

4. The Arch-amoAF/Arch-amoAR primer pair was designed based on alignments of archaeal *amoA* sequences from environmental samples. While sequencing showed many archeal sequences were picked up by this primer pair, the authors mention more degenerate primers may be able to pick up a more diverse range of archaeal *amoA*. Archaea

have been shown to harbour unique *amoA* genes which cannot be picked up by primers targeting AOB [31].

5. The 771F/957R primer pair was designed based on known conserved 16S ribosomal subunit regions to target the *crenarchaeota* group of AOA while discriminating against *Euryarchaeota* and were shown both *in silico* and experimentally to have good specificity [68].

6. The HSBeta396F/HSBeta742R primer pair was designed to determine the abundance of an anammox bacteria in a river bed based on the hydrazine synthase gene (*hzsB*) [102]. *hzs* represents a unique phylogenetic marker for anammox bacteria [39].

7. Scnir374F/Scnir845R target the *nirS* gene of the anammox clade *Candidatus Scalindua* [54]. Kirkpatrick et al. [49] found environmental distribution of anammox *nirS* obtained using sequencing techniques with these primers was consistent with distribution of 16S anammox rDNA.

8. Amx368f/Amx820r is a primer pair targeting the anammox 16S rRNA gene. Amano et al. [2] found 39 of 40 clones obtained using this primer pair had more than 94% similarity to anammox 16S rDNA sequences, with the last clone being separate from the anammox lineage.

9. As mentioned before, comammox bacteria harbour a unique *amoA* gene. Both Nino_amoA_19F / Nino_amoA_252R and comamoA AF/comamoA SR target this gene as comammox bacteria cannot be distinguished from non-comammox *Nitrospira* using 16S rRNA genes [22].

10. The heteroamo378f/634r primer pair was designed based on the highly conserved regions of bacterial *amoA* sequences from different genera belonging to HNB, however, their specificity has not yet been evaluated and many dominant HNB genera have not been included in their design [24].

1.7 Aim of the study

The aim of this study was to evaluate the microbial composition of phosphorus accumulating bioflim in the Hias process. Specifically, to determine the microbial composition of the biofilm in different reactor zones and how potential changes in composition are relevant to phosphorus accumulation and nitrogen removal. In order to obtain relevant data for these investigations, a method was developed for sample handling with the goal of high-quality nucleic acid extraction. DNA isolated from biofilm samples obtained from each zone of interest was used for 16S rDNA sequencing. In addition, the aim was to analyse, using RT-qPCR, the expression of genes involved in nitrogen removal, how the expression of those genes changes throughout the process and how the expression relates to the observed levels of chemical components in the waste water.

2. Materials & Methods

2.1 Biofilm sample handling

The Hias waste water treatment process is divided into three anaerobic (1-3) and 7 aerobic (4-10) zones (See Figure 1.1). Three biofilm carriers, representing biological replicates, were taken at random from the middle of each zone of interest (1, 2, 3, 4, 6, 8 and 10) using sterile tweezers, treated with RNAse Away (Thermo Scientific, USA). Excess water was removed by briefly touching a clean paper towel before carriers were sealed in plastic bags (never combining carriers from different zones in the same bag) and submerged in liquid nitrogen.

Working surfaces were cleaned and treated with RNAse Away (Thermo Scientific, USA). Mortar and pestle and scoops, tweezers and other material to come in contact with the samples were thoroughly washed with dish soap and thoroughly rinsed with deionised water before flaming with ethanol. All material was cooled down before coming in contact with the sample to prevent any thawing. Biofilm carriers were pulverised manually while remaining submerged in liquid nitrogen. When pulverised, excess loose plastic carrier material was separated from the powder using clean sterilised tweezers. The powder was then transferred to pre-cooled falcon tubes and stored at -80 °C. All materials used were cleaned, rinsed and flamed before moving on to the next biofilm carrier.

2.2 RNA isolation

Roughly 500 mg of pulverised biofilm carrier material was transferred to Powerbead tubes (DNeasy Powersoil Pro kit, Qiagen, Germany) containing silica beads, using a sterile cooled scoop and dissolved in 1 ml of cold (4 °C) TRIzol reagent (Invitrogen, USA). The tubes containing biofilm powder, TRIzol and beads were processed in a Mini Beadbeater 96 (Biospec, USA) three times for 40 s, resting one minute on ice to avoid heating of the sample. After 5 min of incubation on ice, 300 μl of chloroform and 30 μl of RNase free water were added and the samples were centrifuged for 15 min at 12000 g at 4 °C after mixing well and incubating for 2 minutes on ice. The resulting upper aqueous phase was mixed with a 1:1 volume of ethanol (70%) and loaded onto a GeneJET RNA Purification Column (GeneJET RNA Purification Kit, Thermo Scientific, USA) by centrifuging at 12000 g for 15 s. The RNA bound to columns was washed by adding 700 μl of GeneJET Wash Buffer I and two times 500 μl of GeneJET Wash Buffer II and centrifuging the spin columns at 12000 g for 15 s. The RNA Purification Column, with bound RNA, was dried by centrifuging at 12000 g for 1 min and RNA was eluted twice in 50 μl RNase free water by centrifuging at 12000 g for 2 min. The resulting RNA was analysed by 1% agarose gel electrophoresis and concentration and purity were measured using NanoDrop (Thermo Scientific, USA). The samples were stored at -80 °C.

2.3 DNA isolation

Roughly 500 mg of pulverised biofilm carrier material was transferred to Powerbead tubes (DNeasy Powersoil Pro kit, Qiagen, Germany) before adding 800 μl Solution CD1 and continuing with the DNeasy Powersoil Pro protocol according to manufacturer instructions, eluting in 50 μl of Solution CD6. The resulting DNA was analysed by 1% agarose gel electrophoresis and concentration and purity were measured using NanoDrop (Thermo Scientific, USA).

2.4 Optimisation and validation of qPCR

Gradient PCR was used to determine the optimal annealing temperature for each primerpair and the resulting products were purified by molecular cloning and sequenced to verify specific amplification.

2.4.1 Determining optimal annealing temperature

Extracted RNA was converted to cDNA using the SuperScript IV First-Strand Synthesis System (Invitrogen, USA). First, contaminating gDNA from around 250 to 450 ng (8 μl) of RNA was removed by gDNA digestion using ezDNase enzyme as instructed by the manufacturers protocol. The ezDNase enzyme was inactivated in the presence of 10 mM of DTT before continuing with 10 μl of ezDNase treated RNA for the SuperScript IV reverse transcriptase reaction using random hexamer primers according to the manufacturers instructions.

The resulting cDNA (1 μl) was diluted 1:5 and PCR amplified using the HOT FIREPol EvaGreen qPCR Supermix (Solis Biodyne, Estonia) and 200 nM of forward and reverse primers specific for different genes of prokaryotes relevant to nitrogen removal in waste water (A.1 [24]) using the following thermal profile: initial denaturation for 12 min at 95 °C followed by 40 cycles of 15 s at 95 °C, 30 s at 56 °C and 30 s at 72 °C followed by a dissociation analysis. qPCR was preformed using the Applied Biosystems 7500 Fast Real-Time PCR System (USA) and the dissociation curve along with wether or not amplification was observed served as an initial indication of whether or not the primer pairs amplified a specific product in samples from the Hias reactor.

The optimal annealing temperature for each primer pair was determined by preforming gradient PCR with varying annealing temperature (52 °C, 55 °C, 58 °C and 61 °C). The same template cDNA and similar reaction conditions were used as for qPCR. The 1:5 diluted cDNA was used for gradient PCR under the following conditions: 1X HOT FIREPol Buffer B1 and 0.04 U/μl HOT FIREPol DNA Polymerase (Solis Biodyne, Estonia), 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 150 nM forward and reverse primers. The thermal profile used was similar to the profile used for qPCR: initial denaturation for 12 min at 95 °C followed by 40 cycles of 15 s at 95 °C, 30 s annealing at variable temperature and 30 s at 72 °C.

2.4.2 Validation of extension product

After gradient PCR, PCR products of expected molecular weight were identified using 1% agarose gel electrophoresis and products of expected molecular weight were purified for Sanger sequencing to verify their complementarity to expected sequences. In order to purify the identified putative expected PCR products, 3 μl of the product was cloned into the pCR4-TOPO vector and transformed into OneShot TOP10 chemically competent cells according to the TOPO-TA Cloning Kit manual (Invitrogen, USA). Resulting colonies were picked into 3 ml LB medium (10 g/l tryptone, 10 g/l $NaCl$, 5 g/l yeast extract) containing 100 mg/l ampicillin and grown for 16-18 h at 37 °C while shaking at 230 rpm. Plasmid DNA was extracted using the PureYield Plasmid Miniprep System (Promega, USA) using the manufacturer provided alternative protocol for larger volumes. Positive clones were identified by restriction digestion by incubating >300 ng of isolated plasmid DNA for one hour at 37 °C in presence of 10 U EcoRI and 1X EcoRI Buffer (NEB, USA). Sequencing was preformed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) according to Platt, Woodhall, and George [76].

The sequencing products were purified by incubating for 15 minutes at room temperature in presence of 3 mM EDTA, 70 mM NaOAc pH 5.2 and 60 % ethanol before centrifuging at 21100 g for 30 minutes at 4 °C. The resulting pellet was rinsed with 70 % ethanol, dried and resuspended in deionised formamide. The products were then analysed by capillary electrophoresis using the 3130xl Genetic Analyzer (Applied Biosystems, USA).

Sequences were manually trimmed for quality and to only include the insert and aligned using CAP3 (see Appendix A.3). The resulting consensus contigs were blasted against the NCBI nucleotide database.

2.5 RT-qPCR

Reverse transcriptase (RT)-qPCR was used to assess the expression of genes involved in the nitrogen cycle throughout the Hias process. RNA samples were diluted to 100 $ng/\mu l$ in RNAse free water and concentration was confirmed using NanoDrop (Thermo Scientific, USA). Genomic DNA was digested using DNase I amplification grade (Invitrogen, USA) by adding 1X DNase I Reaction Buffer and 1 U DNase I Amp Grade enzyme to 800 ng of RNA and incubating 15 min at room temperature before inactivating the reaction by adding 2.5 mM EDTA and incubating at 65 °C for 10 min. Random hexamer primers (SuperScript IV First-Strand cDNA Synthesis System, Invitrogen, USA) were annealed to the gDNA digested RNA samples by adding 40 ng/µl of random hexamer primers and 8mM of dNTPs to 600 ng of RNA sample and incubating at 65 °C for 5 min. Two technical replicates of the RT reaction were set up by adding 1X SSIV Buffer, 100 mM of DTT, 1 μl of Ribonuclease inhibitor and 200 U of SuperScript IV Reverse Transcriptase, and one no RT control reaction was set up by adding 1X SSIV Buffer and 100 mM of DTT. The reactions were incubated at 23 °C for 10 minutes and 55 °C for 10 minutes and inactivated by incubating at 80 °C for 10 minutes. The resulting cDNA samples were diluted 1:10 in RNAse free water.

The cDNA was PCR amplified using the HOT FIREPol EvaGreen qPCR Supermix (Solis Biodyne, Estonia) and 200 nM of forward and reverse primer of one of six primer

pairs: PRK (PRK341F/PRK806R [107]), amoA (amoA-1F/amoA-2R [80]), 771 (771F/957R [68]), CTOAB (CTO189fA/B /CTO654r [53]), CTOC (CTO189fC/CTO654r [53]) or Nino (Nino_amoA_19F/Nino_amoA_252R [22]) (see Appendix A.1), using the following thermal profile: initial denaturation for 12 min at 95 °C followed by 40 cycles of 15 s at 95 °C, 30 s annealing (PRK and amoA: 62 °C, CTOAB and CTOC: 58 °C and 771 and Nino: 52 °C) and 30 s at 72 °C. Template used for amplification with PRK primers was diluted 1:10 000 as opposed to 1:10. Samples were amplified on a Microamp Fast Optical 96 Well Reaction plate (0.1 ml), covered with a qPCR compatible Microamp Optical Adhesive Film (Applied Biosystems, USA) using the Applied Biosystems 7500 Fast Real-Time PCR System (USA).

2.6 16S rDNA sequencing library preparation

1 µl (∼100 - 500 ng) of genomic DNA isolated from each of the Hias waste water biofilm samples was PCR amplified using the HOT FIREPol Blend Master Mix Ready to Load (Solis Biodyne, Estonia) and 200 nM of the PRK341F and PRK806R primers in a 2720 Thermal Cycler (Applied Biosystem, USA) using the following thermal profile: 15 min at 95 °C, 30 cycles of 30 s at 95 °C, 30 s at 55 °C and 45 s at 72 °C and final extension for 7 min at 72 °C. The PCR products were purified using AMPure XP (Beckman Coulter, USA) according to the manufacturer provided protocol. Each sample was assigned a unique combination of two indexing primers (200nM). Indexing PCR was preformed using the same HOT FirePol Blend Mix (Solis Biodyne, Estonia) in a 2720 Thermal Cycler using the following thermal program: 5 min at 95 °C, 10 cycles of 30 s at 95 °C, 1 min at 55 °C, 45 s at 72 °C followed by a final extension for 7 min at 72 °C. The indexed PCR products were quantified using the Qubit dsDNA BR Assay Kit (Invitrogen , USA), according to the manufacturer provided protocol and normalised to equal concentration before pooling. The resulting sequencing library was again purified using AMPure XP as before.

The sequencing library was sequenced using the MiSeq platform (Illumina, San Diego, CA, USA) with 300 bp paired-end reads. The sequences were processed at NMBU (Ås, Norway) according to an established pipeline using QIIME according to [10], Usearch v8.0 according to [26] and OTU clustering according to [77] using 97% sequence similarity for clustering.

2.7 16S rDNA sequencing data analysis and statistics

Clustered OTUs, taxonomy data, raw OTU table and rarefied (25000 reads) OTU table were provided. In addition to taxonomy data provided by NMBU, NCBI BLAST+ 2.13.0 blastn [9, 110] was used to assign taxonomic data from the MiDas 4 database [25] to the obtained OTUs. BLAST was also used to determine the presence of AOA by blasting known ref_seq AOA 16S rRNA gene sequences (accession NR_159208.1, NR_159207.1, NR_159206.1, NR_102913.1 and NR_134097.1) against the obtained OTUs (Appendix B.1). QIIME core diversity analysis [10] was used to calculate alpha and beta diversity metrics. R was used to generate figures, scan data for functional groups as well as comammox bacteria and AOA, and statistical analysis of diversity metrics. Anova and linear regression were used for sta-

tistical analysis of alpha diversity metrics and adonis from the vegan package [70] was used for statistical analysis of beta diversity metrics (Appendix B.2).

3. Results

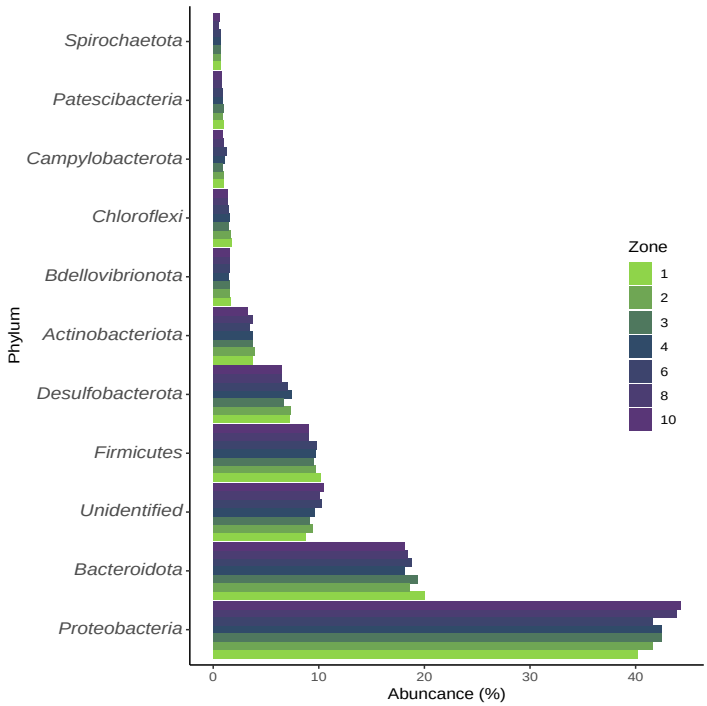
3.1 Microbial composition of biofilm in the Hias process based on 16S rDNA sequencing

Triplicate (at least) biofilm samples were taken from each reactor zone of interest (1, 2, 3, 4, 6, 8 and 10). DNA was isolated and 16S rDNA sequencing was preformed and used to retrieve sequences which were processed at NMBU (Ås, Norway); OTUs were clustered using 97% sequence similarity, rarefied to 25000 reads and taxonomy was assigned using the RDP v18 database [16]. In additions, the MiDas 4 database was used. This database is specifically designed for identifying organisms in waste water treatment plants and provides additional information about their function in waste water treatment. The RDP database on the other hand contains general universal quality controlled sequences. The resulting taxonomy assignment data obtained using the MiDas 4 database was in agreement with the taxonomic data obtained using the RDP v18 database, however fewer OTUs were marked unidentified when using the MiDas 4 database for taxonomic assignment compared to RDP.

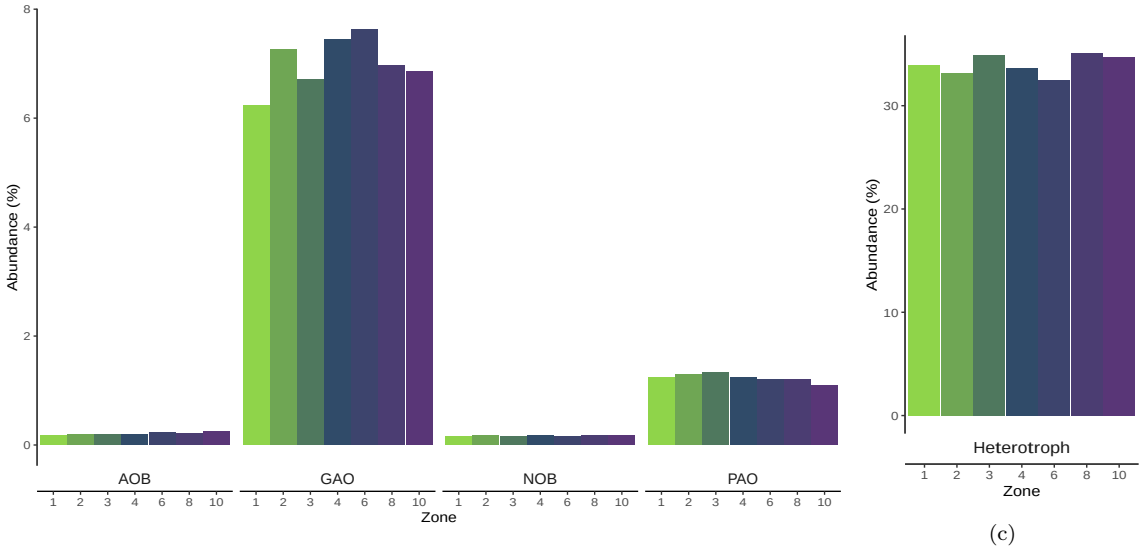
3.1.1 Important functional groups for nitrogen and phosphorus removal in the Hias process

Taxonomic data shows *Proteobacteria* are the most abundant phyla ($>$ 40 % abundance) present in the Hias reactor, this is in agreement with previous studies of the Hias process [82]. Figure 3.1a shows the ten most abundant phyla in the Hias process grouped by reactor zone. In addition to standard taxonomic identification, the MiDas database provides information about which group of organisms the identified sequence belongs to. This data was used in combination with relative abundance calculated from rarefied (25000 reads) abundance data provided by NMBU. Using this method, the relative abundance of groups relevant to nitrogen and phosphorus removal was determined and shown in Figure 3.1b and 3.1c. Ammonium oxidising bacteria (AOB) identified in the Hias reactor all belong to the *Nitrosomonas* genus. Both AOB and nitrogen oxidising bacteria (NOB) are low abundance with AOB $<$ 0.25 % abundance and NOB $<$ 0.20 % abundance. Interestingly glycogen accumulating organisms (GAO) are more abundant than phosphate accumulating organisms (PAO) with GAO around 6.2 - 7.6 % abundance and PAO around 1.0 - 1.3 % abundance. All identified PAO belong to the genus of *Tetrasphaera*. Heterotrophic bacteria are roughly 32 - 35 % abundant. Both ammonium oxidising archaea (AOA) and comammox bacteria were of interest in this study. Neither taxonomic assignment using the RDP nor MiDas database identified any obtained OTUs as belonging to AOA (belonging to the *Nitrosopumilus*, also known as *Thaumarchaeota* phylum). In addition to the taxonomy data, NCBI BLAST+ 2.13.0 blastn [9, 110] was used to blast known ref-seq AOA 16S rRNA gene sequences (accession NR_159208.1, NR_159207.1, NR_159206.1, NR_102913.1 and NR_134097.1) against

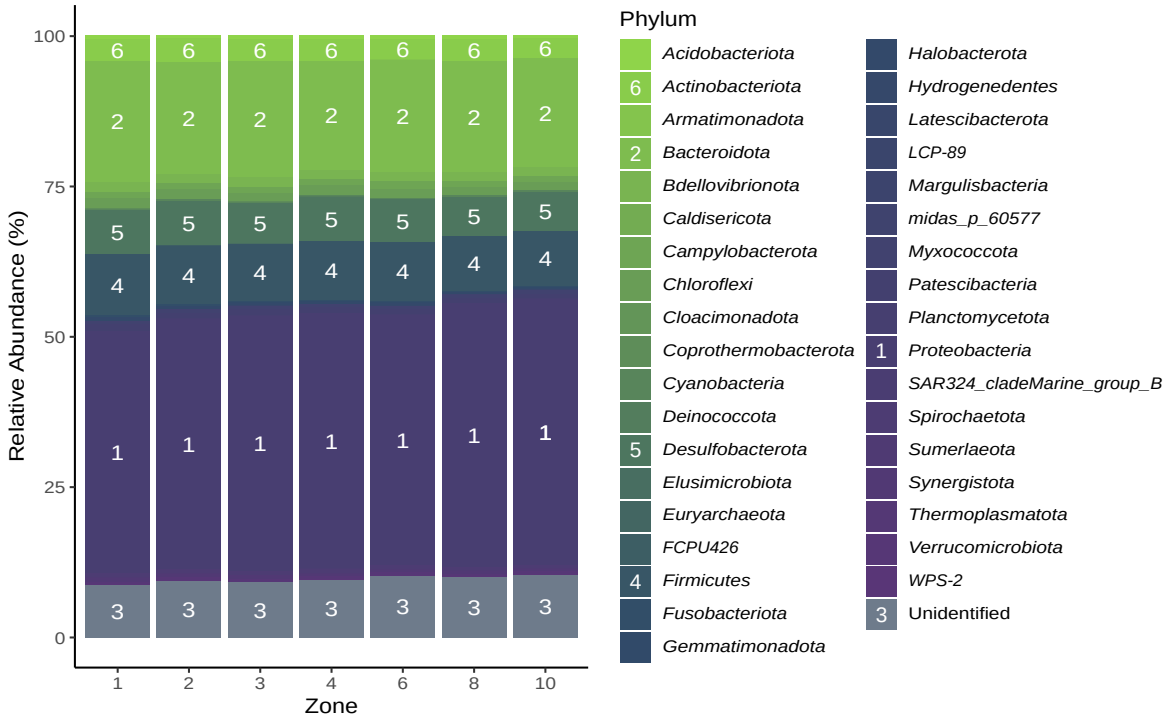
the obtained OTUs from the Hias reactor. No similar sequence hits were found, implying the 16S rDNA sequencing data from the Hias reactor does not contain sequences belonging to AOA. Similarly no comammox bacteria, nor any bacteria belonging to the *Nitrospira* class were identified in the taxonomic assignment data using the RDP nor MiDas database. Though neither AOA nor comammox bacteria were identified using 16S rDNA sequencing, it cannot be concluded they are not present in the reactor (see 4.2.2).



(a)



(b) (c)



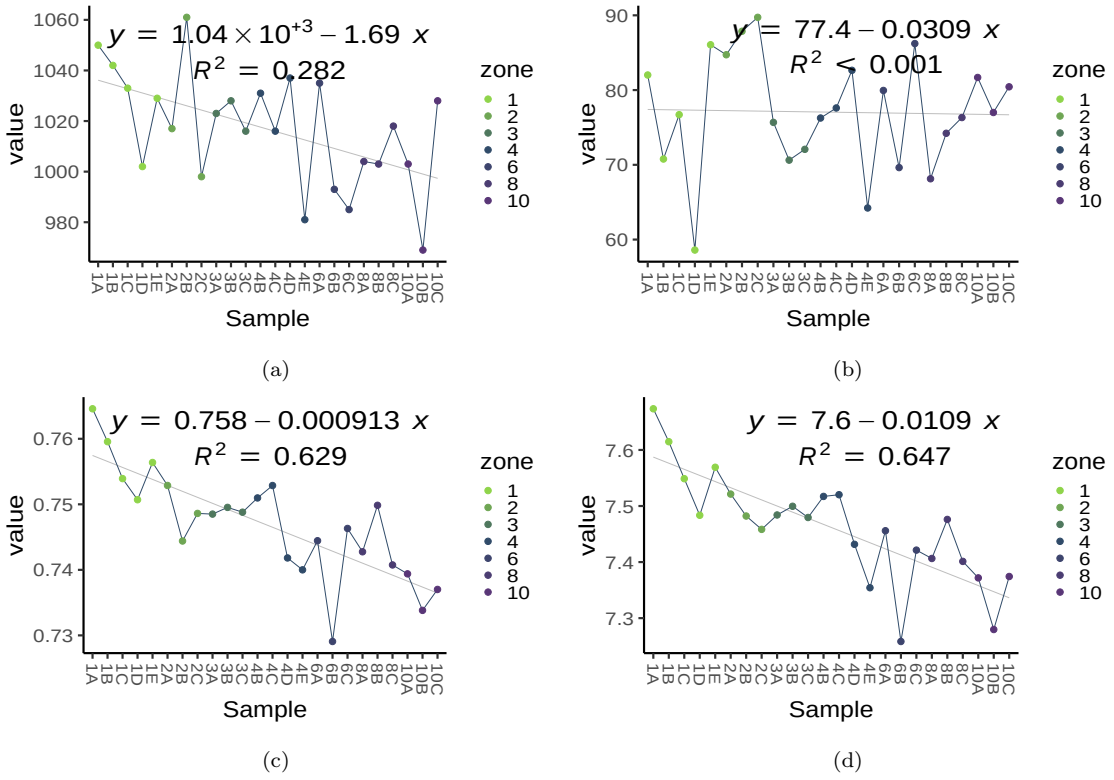
(d)

Figure 3.1: Relative abundance of Phyla and functional groups through the Hias reactor based on identification of OTUs obtained through 16S rDNA sequencing using the MiDas 4 database [25]. 3.1a Realative abundance of top 10 most abundant Phyla through the Hias process, 3.1b change in relative abundance of functional groups of microorganisms through different zones of the Hias reactor; ammonium oxidising bacteria (AOB), glycogen accumulating organisms (GAO), nitrite oxidising bacteria (NOB) and phosphate accumulating organisms (PAO), 3.1c change in relative abundance of heterotrophic bacteria through different zones of the Hias reactor, 3.1d change in relative abundance on the Phylum level through different zones of the Hias reactor (numbers indicate most abundant species for ease of interpretation).

3.1.2 The microbial composition of the Hias biofilm remains relatively unchanged throughout the different reactor zones

The relative abundance of phyla in the Hias reactor was calculated per reactor zone by averaging the relative abundance in samples obtained from the same zone. Plotting these data shows minimal change in relative abundance of phyla throughout different zones of the Hias reactor (Figure 3.1d). In order to evaluate these trends more objectively, QIIME was used to determine alpha and beta diversity metrics and R was used for statistical analysis of these data. Alpha metrics were plotted per sample and linear regression was used to determine what if any trends exist in the data (Figure 3.2). Linear regression revealed no significant trend in change of alpha diversity metrics in the different samples. Analysis of variance (Anova) is used to determine the influence that independent variables have on the dependent variable in a regression study. In this case anova was used to determine whether alpha metric values of a sample are related to which zone the sample was taken from. This analysis is visually represented in Figure 3.3. In anova, the F-value indicates the ratio between variation due to a variable and variation due to error, in other words

how much a variable is responsible for the observed change compared to natural variation. When assessing the two alpha metrics "total amount of species" and "phylogenetic distance", F-values are close to one (0.935 and 1.64 respectively), indicating the changes in alpha metric value are likely due to natural variation rather than differences in the reactor zones. Furthermore, the P-values (respectively 0.496 and 0.197) indicate these observations are likely to arise by random chance. Thus there is no significant change in total observed species, nor phylogenetic diversity throughout the Hias reactor zones. However, alpha metrics taking evenness into account (Shannon entropy and Pielou index), did change significantly throughout the reactor zones, with F-values around 5 (5.395 and 5.547 respectively) and P-values of respectively 0.00278 and 0.00241. This implies species richness remains relatively unchanged whereas the relative abundance does change throughout the Hias process.



(a) (b) (c) (d)

Figure 3.2: Alpha diversity metrics of 16S rDNA sequencing data of samples taken from different zones of the Hias reactor. Values were generated using QIIME [10]. 3.2a Total amount of species per sample, 3.2b phylogenetic diversity [29], 3.2c evenness as measured by Pielou index, 3.2d shannon entropy [36].

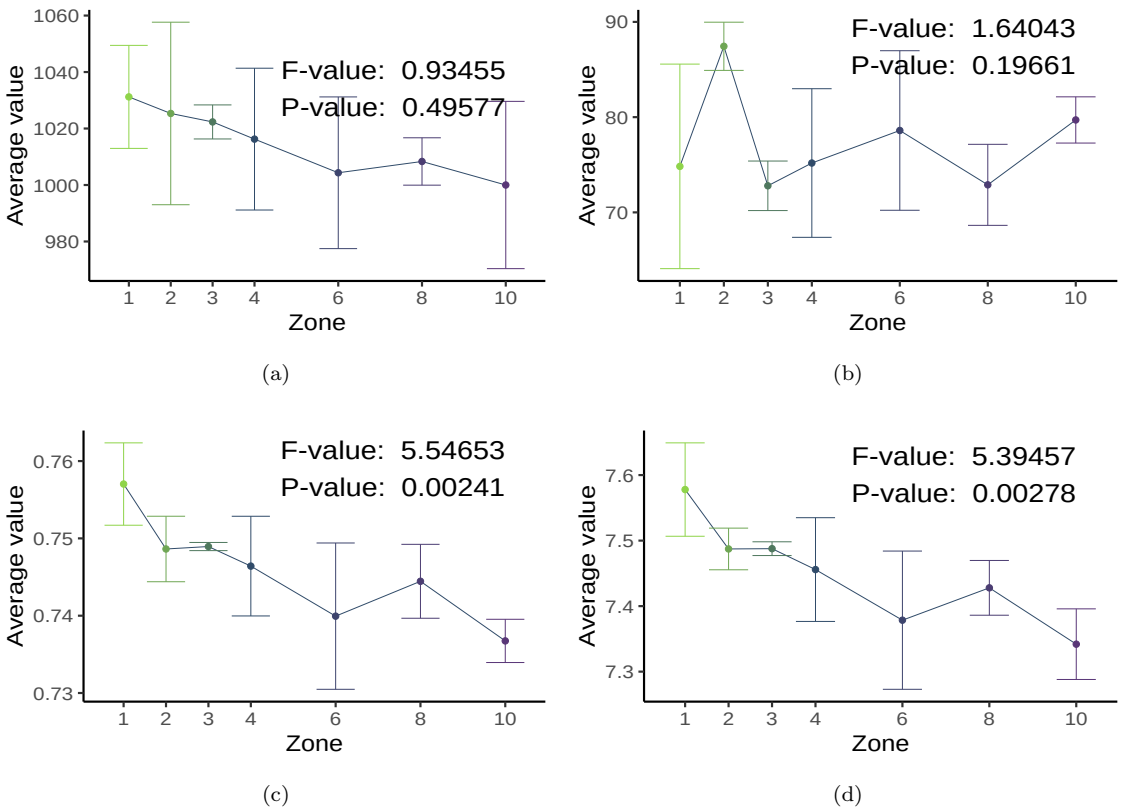


Figure 3.3: Alpha diversity metrics of 16S rDNA sequencing data of samples taken from different zones of the Hias reactor. Values were generated using QIIME [10]. 3.3a Total amount of species per sample, 3.3b phylogenetic diversity [29], 3.3c evenness as measured by Pielou index, 3.3d shannon entropy [36].

Alpha diversity metrics revealed a change in relative abundance throughout the Hias reactor zones. This change is not apparent in beta diversity metrics. Beta diversity metrics are represented as multidimensional distance matrices which are difficult to interpret. To address this, methods exist to map multidimensional data onto a lower dimensional representation for ease of interpretation. Non-metric multidimensional scaling (NMDS) was used to visually illustrate these distance matrix data on a two dimensional plot (Figure 3.4). Clustering of data points in distinct groups or point clouds would suggest strong similarity between those data points, thus a significant difference in beta diversity metric values between zones would result in clustering of data points by their respective zone. No clustering is observed however, indicating variation in samples can be attributed to natural variation. Adonis [70] was developed for statistical analysis of multidimensional distance matrix data and is comparable to anova. Analysis of the four beta diversity metrics used in this study (Bray-Curtis dissimilarity, Jaccard similarity, weighed and unweight UniFrac), all results in high P-values, indicating diversity metrics of samples from the Hias reactor are not statistically related to which reactor zone they were taken from. These findings imply beta diversity between samples is related to biological variation rather than related to the reactor zone.

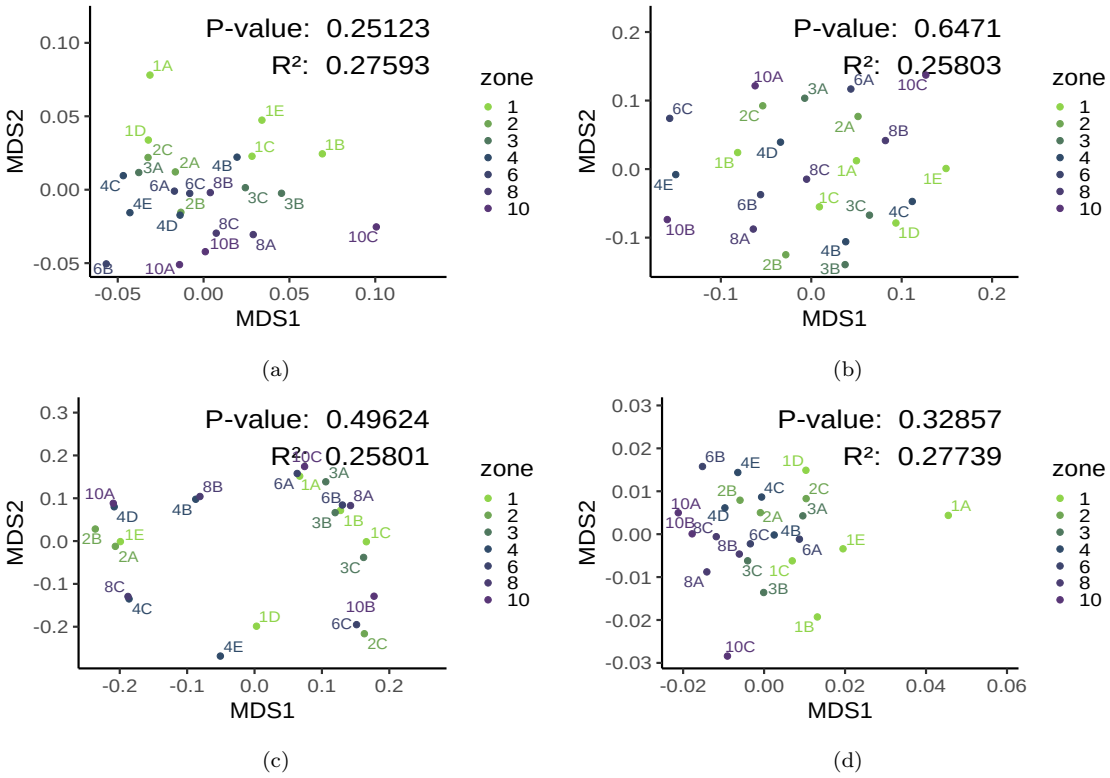


Figure 3.4: Beta diversity metrics of 16S rDNA sequencing data of samples collected from different zones of the HISA reactor, plotted in two dimensions using non-metric multidimensional scaling (NMDS). Values were generated using QIIME [10]. NMDS was preformed using the vegan package [70]. 3.4a Bray-Curtis dissimilarity index, 3.4b Jaccard similarity coefficient, 3.4c phylogenetic distance measured as unweighted (qualitative) UniFrac, 3.4d phylogenetic distance measured as weighted (quantitative) UniFrac.

In conclusion, most changes in diversity observed in the Hias reactor can be attributed to natural variation. In addition, the species richness has been shown to remain relatively unchanged, whereas their relative abundance (species evenness) does change to a significant degree. Namely average species evenness was observed to decrease from the first to the last reactor zone.

3.2 Optimisation and validation of qPCR

RT-qPCR can offer an advantage over 16S rDNA sequencing when it comes to identifying low abundance species or groups because of high target specificity and sensitivity. This method however requires optimisation and validation to ensure this target specificity of the used primers in the sample type being studied. Gradient PCR was preformed to determine the optimal annealing temperature for each of ten primer pairs used in this study for amplification of genes and transcripts of genes involved in nitrogen removal (Appendix A.1). Only gradient PCR reactions for four of the ten tested primer pairs resulted in clear amplification of PCR product(s) (Figure 3.5). Gradient PCR results in amplification of PCR-products of expected size for primer pairs amoA-1F/amoA-2R, 771F/957R, CTO189fA/B/CTO654r and

CTO189fC/CTO654r (hereafter referred to as amoA, 771, CTOAB and CTOC respectively).

Gradient PCR preformed using primer pairs 771, CTOAB and CTOC all resulted in a single product of expected size (220 bp, 465 bp and 465 bp respectively) (Figure 3.5). Based on the intensity of the observed bands after agarose gel electrophoresis, annealing temperature of 52 °C for the 771 primer pair and 58 °C for the CTOAB and CTOC primer pair were used for all further analysis in this study. Sanger sequencing of the PCR products confirms the products are of expected size (see Appendix A.2.1). BLASTing the obtained sequences of these PCR products against the NCBI nucleotide collection database shows the products are complementary to expected sequences; bacterial 16S ribosomal RNA gene for products obtained from CTOAB and CTOC primers and archaeon 16S ribosomal RNA gene for product obtained from 771 primers as expected (see Appendix A.4). Post dissociation analysis of qPCR preformed with the obtained annealing temperature for the respective primer pairs shows a single peak indicating a single unique PCR product as expected (see Appendix Figure A.2f, A.2c and A.2d).

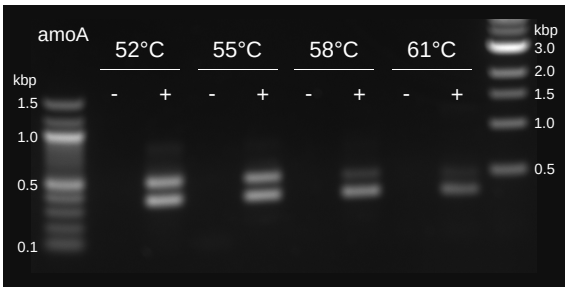
Gradient PCR with primer pair amoA resulted in amplification of a 491 bp product as expected. In addition to the PCR product of expected size, amplification of a product of shorter size was observed (Figure 3.5a). These findings correspond to post dissociation analysis of qPCR preformed using the same cDNA and primer pair under similar reaction conditions (see Appendix Figure A.2a). The two PCR products obtained using gradient PCR were sequenced using Sanger sequencing. BLASTing the obtained sequences against the NCBI nucleotide collection database reveals the 491 bp PCR product is complementary to *amoA* gene products as expected (Appendix Figure A.3a), the additional unexpected amoA PCR product is 349 bp in length and BLAST reveals the product is complementary to partial bacterial 16S ribosomal RNA gene sequences (Appendix Figure A.3b). Gradient PCR shows lower amplification of the expected 491 bp product with increased annealing temperature, therefore further analysis with this primer pair was conducted with an annealing temperature of 52 °C. The optimisation was repeated using higher quality RNA (see 4.1) treated with DNase I amp grade (Invitrogen, USA) instead of SuperScript IV ezDNase (Invitrogen, USA) to ensure complete gDNA digestion before cDNA synthesis. PCR amplification of the resulting higher quality cDNA using the amoA primer pair with an annealing temperature of 52 °C resulted in a single 491 bp product. These findings correspond to post dissociation analysis of qPCR using the same template and conditions, showing a single peak representing the PCR product (Appendix Figure A.4a). The product was sequenced and BLASTed against the NCBI nucleotide collection database and was shown to be complementary to *amoA* gene sequences as expected (Appendix Table A.5a).

Primer pair Nino_amoA_19F/Nino_amoA_252R (hereafter referred to as Nino) resulted in amplification of only a single fragment obtained with an annealing temperature of 62 °C, no amplification was observed when using lower annealing temperature. This product was longer than the expected size (around 900 bp as opposed to the expected 254 bp). qPCR using 52 °C annealing temperature, as suggested by the authors who developed the primer pair [22] did result in amplification however. Post dissociation analysis revealed a single peak (Appendix Figure A.2j), indicating a single PCR product. The identity of this product was unknown however, thus PCR was preformed using higher quality cDNA (as with the amoA primer pair) and an annealing temperature of 52 °C. The resulting PCR product was sequenced and only one sequence of good quality was obtained. BLAST revealed this se-

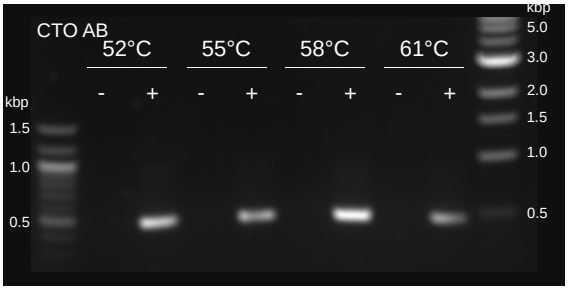
quence was complementary to a region of the 16S RNA gene (Appendix Table A.5), whereas the primer was designed to identify the comammox *amoA* gene. Thus the primer pair did not result in specific amplification of its target.

No amplification was observed in PCR reactions using the remaining primer pairs amoBMf/amoBMr, Arch-amoAF/Arch-amoAF, HSBeta396F/HSBeta742R, Scnir372F/Scnir845R, Amx368f/Amx820r, comamoA AF/comamoA SR and heteroamo 378f/heteroamo 634r (Appendix Figure A.1). Thus these primer pairs were not used in further analysis.

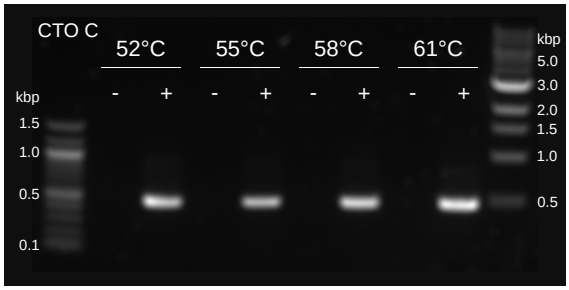
In conclusion, optimal annealing temperature for the five primer pairs resulting in amplification was determined. Primer pairs amoA, 771, CTOAB and CTOC resulted in amplification of their respective expected products after optimisation, whereas the Nino primer pair resulted in non-specific amplification.



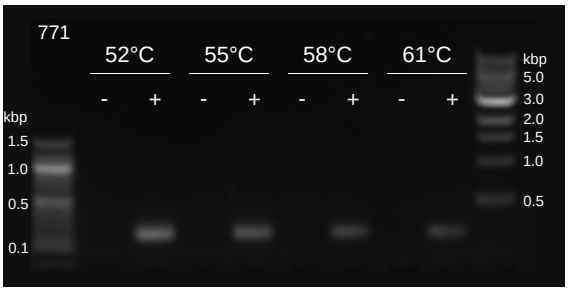
(a) Amplification using primer pair amoA-1F/amoA-2R targeting the *amoA* gene of ammonium oxidising bacteria (AOB)



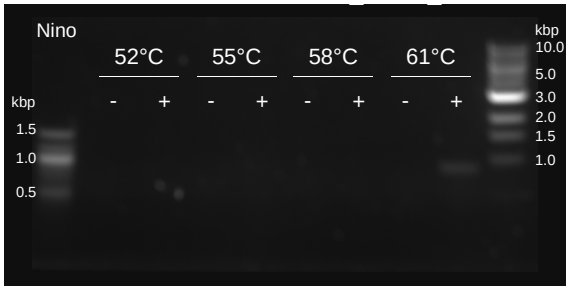
(b) Amplification using primer pair CTO189fA/B /CTO189r targeting the 16S rRNA gene of ammonium oxidising bacteria (AOB)



(c) Amplification using primer pair CTO189fC/CTO189r targeting the 16S rRNA gene of ammonium oxidising bacteria (AOB)



(d) Amplification using primer pair 771F/957R targeting the 16S rRNA gene of ammonium oxidising archaea (AOA)

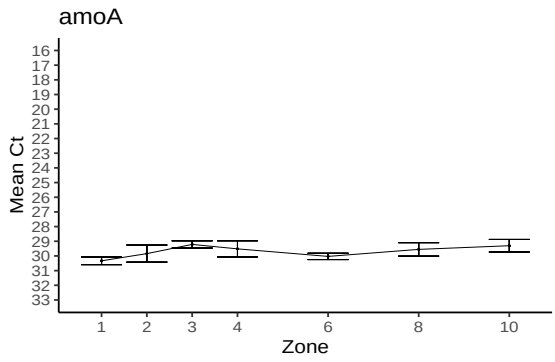


(e) Amplification using primer pair Nino_amoA_19F/Nino_amoA_252R targeting the *amoA* gene of complete ammonium oxidising (comammox) bacteria

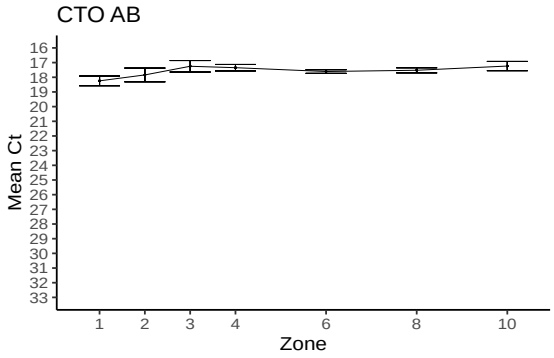
Figure 3.5: Gradient PCR amplification of cDNA derived from biofilm in waste water treatment using primers specific for different genes of prokaryotes involved in waste water nitrite conversion

3.3 Expression of genes associated with nitrogen removal throughout the Hias process

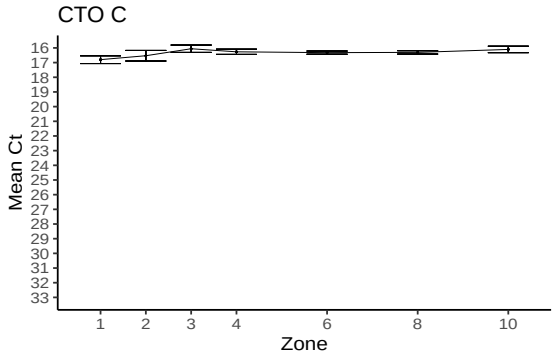
RT-qPCR was used to quantify the transcript level of genes relevant to nitrogen removal in nitrifying biofilm samples taken from different reactor zones in the Hias process. The purpose of this quantification was to evaluate the changes in those transcript levels throughout the reactor zones and how they relate to the ammonium levels measured in those zones at the time of sampling. Three biological replicates were taken from each zone studied (zone 1, 2, 3, 4, 6, 8 and 10). Total RNA was isolated, treated with DNase to remove genomic DNA and finally cDNA was synthesised in two technical replicates. The resulting cDNA used to preform qPCR using primers targeting the *amoA* and 16S rRNA gene of different groups of organisms involved in nitrogen removal. In addition a no reverse transcriptase (no-RT) control was included for each sample to determine whether genomic DNA was completely digested. As a result, for each reactor zone three biological samples were analysed in technical duplicate with an additional no-RT control using qPCR with five primer pairs, each having a unique target. RT-qPCR reactions of all five primer pairs resulted in clear amplification with little variation between reactor zones. No-RT controls did not result in amplification confirming gDNA digestion. Post dissociation plots reveal single peaks as expected, indicating amplification of a single product (Appendix A.5). Thus RT-qPCR indicates the presence of transcripts of the *amoA* gene of AOB and comammox bacteria as well as transcripts of the 16S rRNA gene of AOB and AOA. Mean cycle threshold (Ct) value and standard deviation for each zone were calculated and plotted for every primer pair (Figure 3.6). Anova was used to determine the statistical significance of the variation in transcript level measured as mean cycle threshold value for each primer pair, through the different reactor zones of the Hias process. No significant variation in transcript level was observed for primer pairs amoA, CTOAB, CTOC 771 or Nino, with respective F-values of 1.588, 3.043, 3.043, 0.019 and 18.70 and P-values of 0.263, 0.142, 0.144, 0.897 and 0.008. Thus it can be concluded the observed transcript levels in different reactor zones did not change significantly.



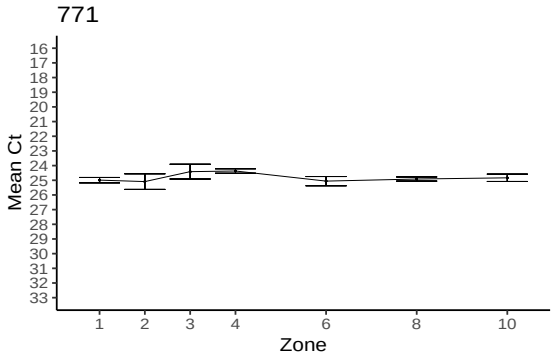
(a) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: amoA, Primers: amoA-1F & amoA-2R



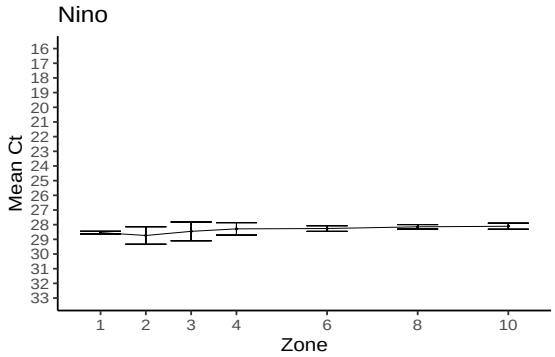
(b) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: 16s rRNA, Primers: CTO189fA/B & CTO189r



(c) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: 16s rRNA, Primers: CTO189fC & CTO189r



(d) Micro organisms: Ammonium Oxidising Archaea, Functional Gene: 16s rRNA, Primers: 771F & 957R

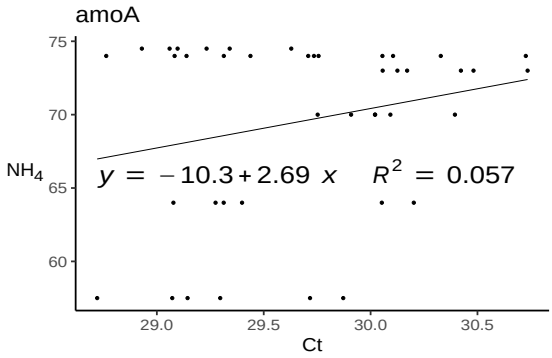


(e) Micro organisms: Complete Ammonium Oxidising Bacteria, Functional Gene: amoA, Nino_amoA_19F & Nino_amoA_252R

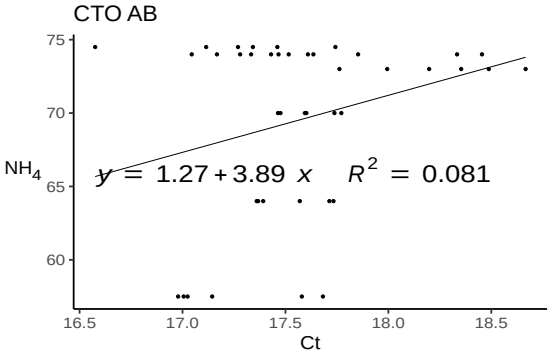
Figure 3.6: Graphs of Mean cycle threshold (Ct) value of transcripts associated with nitrification throughout different reaction zones of the Hias process with standard deviation error bars

While taking samples used for RT-qPCR analysis as well as 16S rDNA sequencing, technicians at Hias waste water treatment plant took water samples to measure the concentration of relevant chemical parameters including ammonium levels (Appendix A.7). While Ct values don't change significantly, measured ammonium levels decrease from 74 mg/l in zone 4 to 57.5 mg/l in zone 10. The lack of significant variation in transcript level suggests there may be no correlation between transcript levels of *amoA* nor 16S rRNA of ammonium oxi-

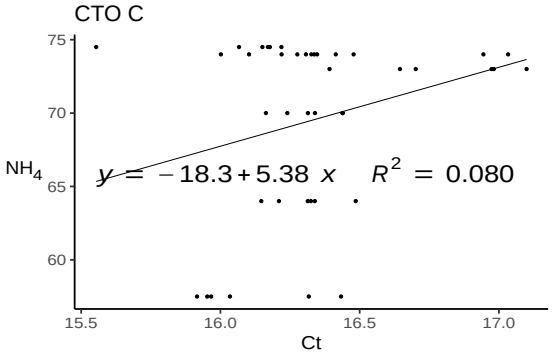
dising organisms and NH_4 levels in the Hias process. Linear regression of Ct-value to NH_4 levels confirms there is no correlation between transcript level and ammonium concentration (Figure 3.7, Appendix A.6).



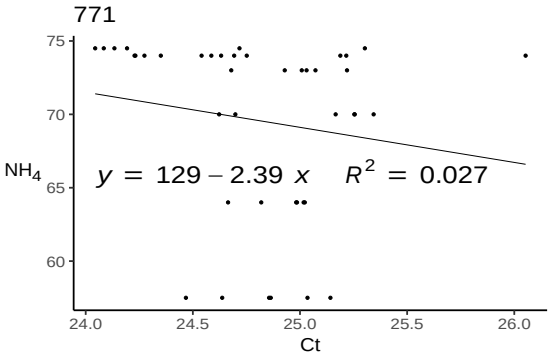
(a) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: amoA, Primers: amoA-1F & amoA-2R



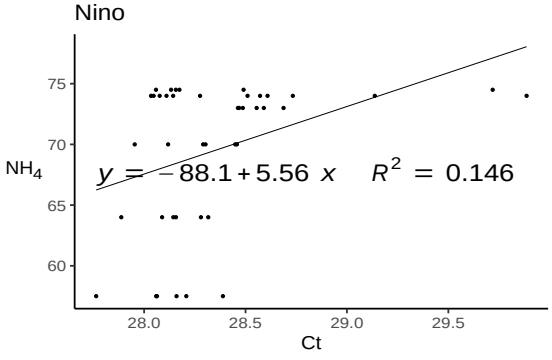
(b) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: 16s rRNA, Primers: CTO189fA/B & CTO189r



(c) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: 16s rRNA, Primers: CTO189fC & CTO189r



(d) Micro organisms: Ammonium Oxidising Archaea, Functional Gene: 16s rRNA, Primers: 771F & 957R



(e) Micro organisms: Complete Ammonium Oxidising Bacteria, Functional Gene: amoA, Nino_amoA_19F & Nino_amoA_252R

Figure 3.7: Graphic summary of correlation between transcript levels of genes involved in nitrogen removal, measured as qPCR Ct value, and NH_4 concentration throughout the Hias process

4. Discussion

4.1 The importance of sample handling and high quality RNA and DNA isolation in microbial studies

When studying microbial communities, it is essential that samples used in analysis accurately represent the community at time of sampling. Measures must be taken to prevent the introduction of bias when sampling and isolating nucleic acid, thus ensuring for example that DNA and RNA of all species is represented in a quantity that accurately represents their occurrence in the biological system. The Hias reactor is divided into ten holding zones to ensure a more constant rate of biofilm carrier flow [41]. Despite this measure, there remains variation in retention time of biofilm carriers in the reactor zones. To minimise bias in sampling, at least three biofilm carriers were taken from the middle of each reactor zone of interest at random, representing biological parallels. To ensure no bias is introduced after sampling, it is essential all changes to the sample are minimised between the time of sampling and nucleic acid extraction. RNA is particularly prone to degradation which may cause loss of important information. Furthermore, if biological activity is not halted at time of sampling, organisms may respond to changing environmental conditions as a result of the sampling by altering gene expression. RNAlater is widely used to prevent degradation of RNA during storage of biofilm samples [17, 18]. Initially, an RNA isolation protocol was developed for Hias biofilm samples making use of RNAlater as a storage solution during sample transportation [100]. While good quality RNA has been obtained using this protocol, samples obtained during winter resulted in highly degraded RNA. Because of the presence of an extracellular polysaccharide matrix (EPS), RNA extraction from biofilm samples presents challenges similar to those of RNA extraction from plant material; namely interference of polysaccharide substances with nucleic acid extraction and purification as well as DNase activity and PCR amplification. Liquid nitrogen is widely used to preserve RNA in plants and make the material brittle to accommodate homogenisation [88]. Thus an RNA isolation protocol was developed using liquid nitrogen for sample storage instead of RNAlater. The resulting RNA was of high quality (see Appendix A.3). Furthermore, Passow et al. [72] found RNAlater may introduce bias in expression studies by eliciting a biological response whereas liquid nitrogen storage does not. In addition to preventing RNA degradation, this method addresses another potential point for introduction of bias; to ensure a representative sample of nucleic acid from a biofilm sample, all biofilm material must be released from the plastic biofilm carrier. Different methods were tested including using mechanical shearing force to release biofilm material from the plastic carrier (Appendix A.8). The mechanical force used in this method meant a trade off between nucleic acid integrity and ensuring complete release of biofilm material from the carrier. The brittleness introduced by liquid nitrogen on the other hand enables simple mechanical homogenisation using mortar and pestle. This method results in minimal degradation and more complete homogenisation. Because of the improved

homogenisation, high-quality DNA could be isolated from the same homogenised material, in addition to high-quality RNA. When using this method for homogenisation, no significant heat is generated, further helping to prevent degradation and because samples never thaw before nucleic acid extraction, samples can be easily stored for an extended period at -80 °C. Thus the liquid nitrogen method developed for handling Hias biofilm samples resulted in high-quality RNA and DNA isolated from the same homogenised sample while minimising introduction of bias and extending sample storage opportunities.

4.2 Assessment of functional groups of micro organisms in the Hias reactor

4.2.1 Phosphorus accumulation

The samples analysed in this study were taken in winter. Water temperature at time of sampling was 9 °C. Previous research has illustrated a lower phosphorus removal potential associated with lower water temperatures in the Hias process [82, 85]. Chemical data shows at time of sampling much soluble carbon (measured as soluble chemical oxygen demand, SCOD) remained in the waste water when entering the aerobic part of the reactor. Saltnes, Sørensen, and Eikås [85] showed this may affect phosphorus uptake as a result of increased competition from heterotrophic bacteria. Despite the remaining SCOD and lower water temperature, high phosphorus removal was observed. Heterotrophic bacteria did not significantly increase in relative abundance in aerobic zones, indicating there may not be a significant effect of competition for carbon sources with heterotrophic bacteria. However, PAO relative abundance did decrease in aerobic zones. PAO reached highest abundance in zone 3 (1.36 %) and lowest abundance in zone 10 (1.09 %). Glycogen accumulating organisms (GAO) were roughly 6 times more abundant than PAO. Rudi et al. [82] showed PAO to be around 30 times more abundant than GAO in Hias biofilm in 2019, noting at the time that phosphorus removal potential (PAO abundance) might be excessive as a result of fluctuating phosphorus levels in the waste water influent. A potential explanation for the lower observed abundance is reactor operation under more stable influent as compared to conditions in 2019. Despite the lower abundance of PAO, chemical data still shows sufficient phosphorus removal to meet the discharge limits.

4.2.2 Nitrogen removal

The observed nitrogen removal efficiency (measured as decrease in NH_4 levels, Appendix A.7) was relatively low as compared to nitrogen removal potential observed under optimal conditions [85]. As with phosphorus removal, low nitrogen removal is known to be associated with lower water temperatures [82, 85]. The Hias process is presumed to employ incomplete nitrification (AOB but not NOB) in combination with denitrification to convert NH_4 to gaseous nitrogen [85]. Despite this, NOB have been found to have a similar abundance to AOB (AOB < 0.25 %, NOB < 0.20 %). Both RT-qPCR and 16S rDNA sequencing confirm the presence of AOB and RT-qPCR reveals probable activity (as measured by *amoA* transcripts). The activity of NOB has not been assessed due to a lack of primers targeting

NOB in this study. Based on the data presented in this study, the importance of NOB in the Hias process cannot be determined, however further research could assess their importance in nitrogen removal. Other organisms may contribute to nitrogen removal besides AOB and NOB. No anammox bacteria have been identified in the Hias process. Primers targeting both AOA and comammox bacteria however, resulted in amplification when preforming qPCR with cDNA from the Hias reactor. However, 16S rDNA sequencing with DNA from the exact same samples did not identify AOA nor comammox bacteria, despite both being represented in both databases used for taxonomic assignment. Sanger sequencing of PCR products obtained using primers targeting comammox *amoA* indicated the amplification was non-specific , explaining the incongruence. However, RT-qPCR using AOA targeting primers did result in specific amplification of expected sequences. Furthermore AOA specific primers targeted 16S rRNA, thus RT-qPCR was able to pick up AOA 16S rRNA whereas 16S rDNA sequencing did not pick up AOA 16S rDNA. Biases in 16S rDNA sequencing (as touched on before, 1.6.1) have been shown to affect predominantly low abundance targets, making qPCR more reliable in this case [23]. Thus it is likely AOA are present in the Hias reactor, while their contribution to nitrogen removal remains unclear. The presence of comammox bacteria however, cannot be concluded based on the data presented. In connection with other research, whole genome shotgun sequencing, as well as meta-transcriptomic sequencing was preformed using the same DNA and RNA obtained from the Hias reactor (data not shown in this study). Further investigation of these data could give insight into the presence of AOA and comammox bacteria as well as their activity levels throughout the reactor.

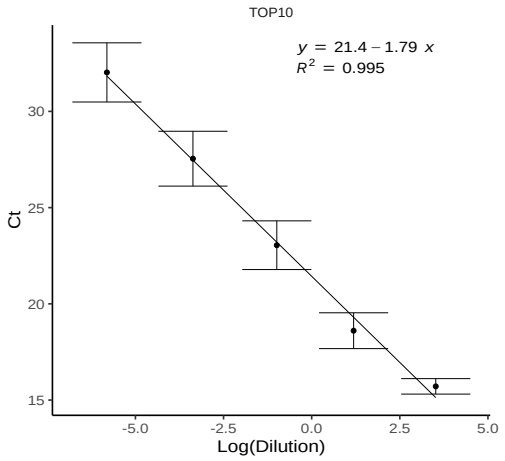
4.3 The lack of change in biofilm microbial composition and implications for regulation of biological processes

Measured NH_4 levels start decreasing at the transition from the initial anaerobic zones to the aerated part of the reactor. This observation is easily explained by the aerobic nature of ammonium oxidation. Diversity metrics indicate the microbial composition of Hias biofilm remains relatively unchanged throughout the reactor zones, whereas the relative abundance of species present may change; Rudi et al. [82] found decreased alpha diversity and increased beta diversity in response to low temperature. Despite beta diversity being observed to be higher at low temperatures, no significant change in beta diversity was shown to be related to which Hias reactor zone the sample was taken from. Alpha metrics however, revealed a decrease in species evenness throughout the different reactor zones. AOB abundance was lowest in reactor zone 3 (0.186 %) and highest in zone 10 (0.244 %). Reactor zone 3 is the last anaerobic zone before the transition to the aerobic part of the Hias reactor whereas zone 10 contains the biofilm carriers which have been exposed to aerobic conditions for the longest time anywhere in the reactor. These findings appear consistent with theoretical knowledge of the aerobic growth of AOB. It cannot be definitively determined whether the increased relative abundance is a result of increased AOB growth or relative decreased growth of other groups of micro organisms. However, the other functional groups assessed in this study did not follow any such discernible trends, thus indicating the increased AOB abundance after aerobic growth may be biologically relevant. RT-qPCR is a more targeted

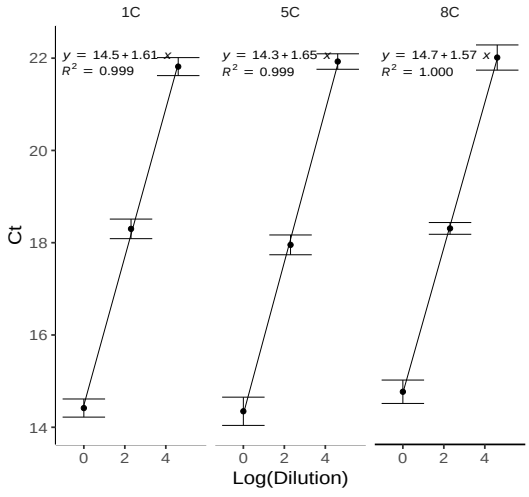
approach and allows for clear quantification of expression levels. Transcript levels may give more information about biological activity of AOB and how it changes throughout the Hias process.

4.4 Normalisation of qPCR data

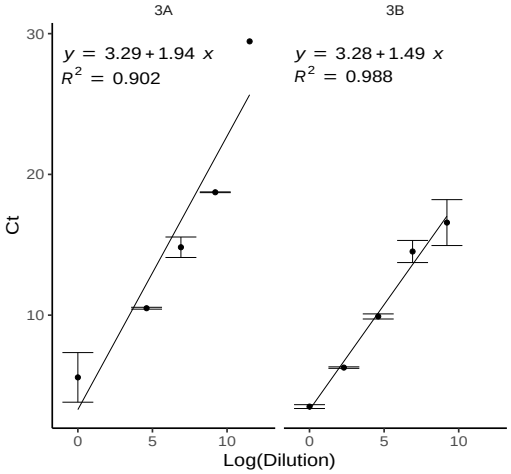
Before being able to draw conclusions from RT-qPCR data, it is essential to validate the methods used, and importantly, using the same sample type as used for analysis, as effects of for example PCR inhibitors can bias results by affecting PCR efficiency. One important part of any RT-qPCR study is normalisation; accounting and correcting for differences in the input to the reaction, for example nucleic acid concentration, as well as composition with regards to inhibitors. This correction can be done by measuring the expression levels of a gene known to be expressed continually and equally under the changing experimental conditions (refernce gene). According to the MIQE guidelines, reference genes used for normalisation of qPCR data should be stably expressed and there should be high correlation between their expression and total mRNA levels. Additionally, the use of a single reference gene is only acceptable if good evidence for invariant expression under the studied experimental condition is given [7]. The high complexity of the Hias biofilm means meeting these requirements is very challenging. Amplifying a reference gene with equal efficiency in a complex microbial community without introducing bias is essential for studying microbial communities and, despite much research, remains a challenge. Yu et al. [107] developed universal primers based on the V3 and V4 regions of prokaryotic 16S rDNA, matching approximately 97.4% of bacterial and 93.8% of archaeal rRNA sequences in the database used for validation [95]. This approach however, is biased towards prokaryotes with greater 16S gene copy number. Tools exist to correct for these biases when assessing microbiome sequence data, however they have been shown to have low predictive accuracy [60] . Despite these limitations, 16S rRNA is commonly used for normalisation of microbiome data. When assessing qPCR data however, it is also important to assess PCR efficiency for targets used as reference genes [7]. Standard curves generated using a serial dilution of both *E. coli* gDNA as well as gDNA and cDNA obtained from the Hias reactor show PCR efficiency between 227.69 % and 368.97 % (Figure 4.1). These excessive efficiency percentage can both be due to the influence of inhibitors as well as excessive cDNA target concentration. Amplification with PRK primers resulted in very low Ct values, indicating excessive target concentration. The presence of EPS in biofilm samples is known to have PCR inhibitory effects [88].



(a) Standard curve of qPCR efficiency of amplification of TOP10 *E. coli* genomic DNA using universal prokaryotic primers



(b) Standard curve of qPCR efficiency of amplification of Hias biofilm genomic DNA using universal prokaryotic primers



(c) Standard curve of qPCR efficiency of amplification of Hias biofilm copy DNA using universal prokaryotic primers

Figure 4.1: Determination of qPCR efficiency of universal prokaryotic primers with different template

Amplification of Hias samples using PRK primers resulted in very low Ct values, which may fall outside of the linear amplification range, therefore template used was diluted 1:1000 x more than template used for amplifying gene products of interest. This dilution introduces technical error. Figure 4.2 shows normalisation with the universal PRK primers results in high standard deviation, likely caused by technical variation introduced during dilution. An additional strategy for normalisation is normalisation against total RNA concentration used in the cDNA synthesis reaction. (cDNA concentration cannot be accurately measured directly as reaction components such as dNTPs used in the synthesis reaction will affect concentration measurements). The method used for this concentration measurement however did not have sufficient accuracy as reflected in increased spread in average values and stan-

dard deviation of Ct values after normalisation (Figure 4.2). As a result of these findings, no normalisation was used for analysis of qPCR Ct data. To ensure reliability of the RT-qPCR method, total RNA into the RT-qPCR reaction was normalised to ensure the minimal variation possible as a result of differing nucleotide amounts in the PCR reaction. In addition, special care was taken to ensure minimal technical error. The low variation in Ct value for technical replicates indicates a good amount of reliability of the obtained RT-qPCR data. Though efforts were made to minimise sources of technical error, the lack of normalisation must be taken into account when interpreting data, as it may result in small but significant expression differences going undetected. Additionally, significant differences in Ct value may not be reflective of biological significance. Furthermore, Ct data was not normally distributed, which could render statistical analysis more inaccurate (Appendix A.6.3).

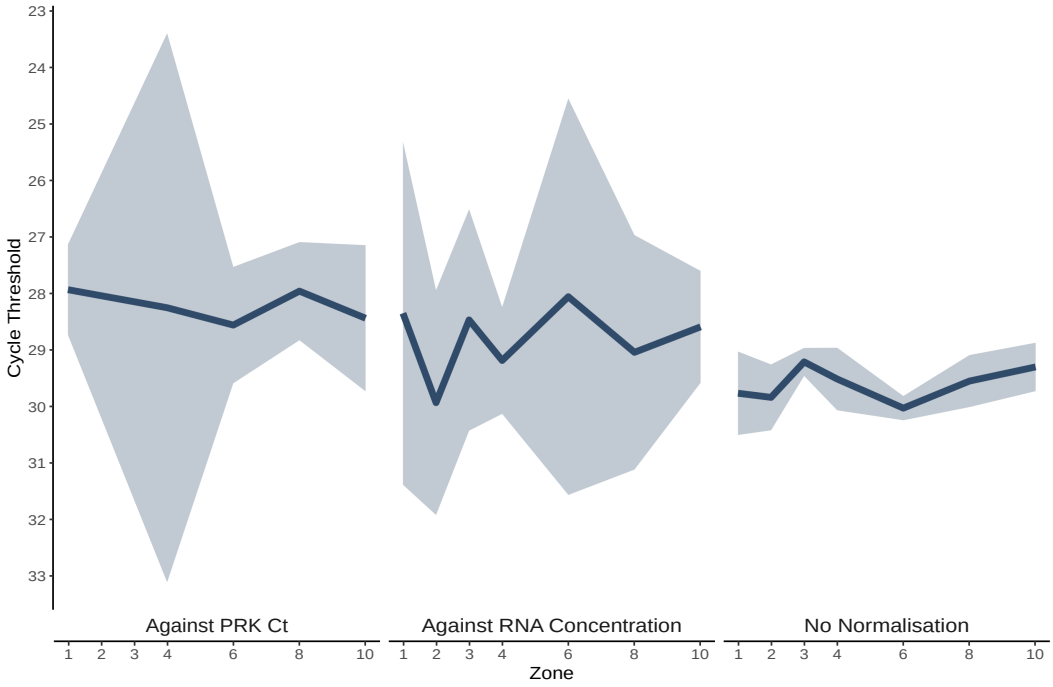


Figure 4.2: Comparison of mean Ct value and standard deviation after different normalisation methods shown respectively as a full line and ribbon. Used methods were normalisation against universal prokaryotic primers (PRK) and total RNA concentration measured before cDNA synthesis, as well as no normalisation.

Taking this into account, it can be concluded that no significant change in Ct value of transcripts related to ammonium oxidation was detected and there was no correlation between NH_4 concentration and detected levels of transcripts. This means there either was no significant change in actual transcript levels in response to NH_4 concentration or the change was too small to be detected by the methods used.

4.5 Lack of change in *amoA* transcripts in response to NH_4 concentration

The increase in AOB abundance observed in 16s rDNA sequencing data is not reflected in RT-qPCR data. The lack of normalisation of RT-qPCR data could be to blame for changes in abundance going unnoticed. Sequencing data shows the change in relative abundance is low (0.058 % change in relative abundance at the extremes). These changes may be to small to result in significant change in transcript levels as measured by RT-qPCR with the methods used. In addition, no correlation between *amoA* transcript levels and NH_4 concentration was observed. Though typically *amoA* mRNA concentration increases in response to high NH_4 concentration after a period of starvation, Bollmann et al. [5] found mRNA concentrations to be constant during growth. In addition, Stein, Arp, and Hyman [92] found *de novo* protein synthesis rather than transcript level to be the mechanism for increasing AMO activity in response to NH_3 in *Nitrosomonas europaea*. Sheikh and Klotz [87] reported continuous *amoA* expression in *Nitrosococcus oceani*. In addition Sheikh and Klotz [87] revealed complex regulation mechanisms for the entire *amo* gene cluster including multiple promotors (two constitutive and one regulated by presence of ammonia) as well as regulation based on mRNA degradation. Sheikh and Klotz [87] concludes *amoA* transcripts may be more stable in environments with low ammonium levels. While many report a lack of change in *amoA* transcript levels, Aoi et al. [3] did report a significant *amoA* mRNA level change in response to ammonium levels in *Nitrosomonas europaea*. AOB identified through 16S rDNA sequencing in this study all belonged to the *Nitrosomonas* genus. Thus a similar response in transcript levels could be expected. Ammonium levels used to induce transcription by Aoi et al. [3] were much higher (600 mg/l) than ammonium levels observed in the Hias reactor. In the Hias reactor, ammonium levels only decreased by 17 mg/l. In addition, Aoi et al. [3] reports a long *amoA* mRNA half life (on the order of hours). The time in which Hias biofilm is exposed to lower ammonium concentrations may be too short for mRNA levels to decrease to a detectable degree.

Much remains unknown about the regulation of AMO in complex microbial communities and how it may be affected by environmental conditions. As mentioned before, meta-transcriptomic sequencing was performed in connection with this study (data not shown). In light of further reseach, insight in these meta-transcriptomics data, not only of *amoA* but also of *amoB* and *amoC* as well as genes potentially involved in formation of the AmoABC trimer, could give more insight into regulation of ammonium oxidation in the Hias process. Additionally, studies of AMO on the protein level, rather than the transcript level alone, may aid in this understanding. Proteomics studies of the AMO protein may however present challenges related to solubilisation and separation as AMO is membrane bound [96].

Other groups of organisms rather than AOB and HNB could also contribute to the decreasing NH_4 concentration. Salbitani and Carfagna [84] conclude micro algae could be used for effective nitrogen removal. Though eukaryotes have not been studied as a part of this thesis, no significant micro algae growth has been observed in the Hias reactor, rendering their contribution to nitrogen removal in the Hias process questionable. Rudi et al. [82] studied eukaryotes in the Hias process but no functional assignment was preformed.

In conclusion, RT-qPCR did not reflect the observed increase in relative abundance of AOB, possibly as a result of hampered sensitivity through lack of normalisation. Secondly, while no correlation between NH_3 and *amoA* levels was observed, regulation at different biological levels, such as protein synthesis, may give further insight and thus could warrant further research.

5. Conclusion

A sample handling method was developed using liquid nitrogen for sample storage. This method allowed extended storage of biofilm samples. The brittleness introduced by the cold temperature facilitated complete release of biofilm material from the carrier as well as mechanical homogenisation of the sample. This homogenised material was used for both RNA and DNA isolation and resulted in high yield and low degradation level.

RNA and DNA obtained through this method were used to assess microbiota composition and gene expression of genes related to nitrogen removal as well as how, if at all, they change throughout the Hias process. Assessment of alpha and beta diversity metrics showed the variation of species in samples from different reactor zones in the Hias process is probably a result of natural variation. In other words, the microbial composition of the Hias biofilm does not change significantly throughout the Hias process. However, the species evenness did decrease through the Hias process to a significant degree, implying changes in relative abundance of one or more species in the biofilm.

Chemical data suggests low nitrogen removal, which can be explained by low water temperature at time of sampling. Despite low water temperature, phosphorus removal remained sufficient to meet discharge requirements. *Proteobacteria* were the most abundant phylum (> 40 % relative abundance) and the functional group of heterotrophic bacteria accounted for over 30 % of the observed microbiota. No trend was observed in the abundance of heterotrophic bacteria as it relates to which part of the Hias reactor was aerated. PAO on the other hand, reached highest abundance at the end of the anaerobic part of the reactor, where they enjoy a growth advantage and their abundance steadily decreased throughout the aerated part of the reactor by 0.27 % relative abundance.

AOB and NOB were comparable in abundance, with AOB being slightly more abundant (< 0.25 % and < 0.20 % abundance respectively). No trend was observed in relative abundance of NOB. The activity of NOB has not been evaluated in this study, but further research could shine light on their involvement in the Hias process. Chemical data suggests ammonium oxidation activity. AOB abundance increased in aerated zones reaching highest abundance in zone 10 (0.244 %) while their relative abundance was lowest in zone 3 (0.186 %). While relative abundance increased, no significant change in transcript level of AOB *amoA* nor 16S rRNA was observed in RT-qPCR analysis. This may be due to a lack of sensitivity of the used method as no normalisation of RT-qPCR data was used. Furthermore, previous research has shown *amoA* mRNA to linger for up to several days after a period of growth. Much remains unknown about AMO regulation. Further research of the Hias process on the transcript and protein level may reveal insights into control of AMO activity.

In addition to AOB and NOB, RT-qPCR suggests the presence of AOA. Indications of the presence of comammox bacteria were not conclusive. The data shows no indications of the presence of anammox bacteria.

Bibliography

[1] Elinor T. Adman, J.W. Godden, and Stewart Turley. "The Structure of Copper-nitrite Reductase from Achromobacter cycloclastes at Five pH Values, with NO-2 Bound and with Type II Copper Depleted". In: *Journal of Biological Chemistry* 270.46 (Nov. 1995), pp. 27458–27474. DOI: 10.1074/jbc.270.46.27458.

[2] Teruki Amano et al. "Detection of Anammox Activity and Diversity of Anammox Bacteria-Related 16S rRNA Genes in Coastal Marine Sediment in Japan". In: *Microbes and Environments* 22.3 (2007), pp. 232–242. DOI: 10.1264/jsme2.22.232.

[3] Y. Aoi et al. "Quantitative analysis of amoA mRNA expression as a new biomarker of ammonia oxidation activities in a complex microbial community". In: *Letters in Applied Microbiology* 39.6 (Dec. 2004), pp. 477–482. DOI: 10.1111/j.1472-765x.2004.01585.x.

[4] Daniel J. Arp and Lisa Y. Stein. "Metabolism of Inorganic N Compounds by Ammonia-Oxidizing Bacteria". In: *Critical Reviews in Biochemistry and Molecular Biology* 38.6 (Jan. 2003), pp. 471–495. DOI: 10.1080/10409230390267446.

[5] Annette Bollmann et al. "Influence of Starvation on Potential Ammonia-Oxidizing Activity and amoA mRNA Levels of Nitrosospira briensis". In: *Applied and Environmental Microbiology* 71.3 (Mar. 2005), pp. 1276–1282. DOI: 10.1128/aem.71.3.1276-1282.2005.

[6] Violaine Bonnefoy and John A. Demoss. "Nitrate reductases in Escherichia coli". In: *Antonie van Leeuwenhoek* 66.1-3 (1994), pp. 47–56. DOI: 10.1007/bf00871632.

[7] Stephen A Bustin et al. "The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments". In: *Clinical Chemistry* 55.4 (Apr. 2009), pp. 611–622. DOI: 10.1373/clinchem.2008.112797.

[8] Laia Calvó and L.Jesús Garcia-Gil. "Use of amoB as a new molecular marker for ammonia-oxidizing bacteria". In: *Journal of Microbiological Methods* 57.1 (Apr. 2004), pp. 69–78. DOI: 10.1016/j.mimet.2003.11.019.

[9] Christiam Camacho et al. "BLAST: architecture and applications". In: *BMC Bioinformatics* 10.1 (Dec. 2009). DOI: 10.1186/1471-2105-10-421.

[10] J Gregory Caporaso et al. "QIIME allows analysis of high-throughput community sequencing data". In: *Nature Methods* 7.5 (Apr. 2010), pp. 335–336. DOI: 10.1038/nmeth.f.303.

[11] Jonathan D. Caranto and Kyle M. Lancaster. "Nitric oxide is an obligate bacterial nitrification intermediate produced by hydroxylamine oxidoreductase". In: *Proceedings of the National Academy of Sciences* 114.31 (July 2017), pp. 8217–8222. DOI: 10.1073/pnas.1704504114.

[12] Gilda Carvalho et al. “Differential distribution of ammonia- and nitrite-oxidising bacteria in flocs and granules from a nitrifying/denitrifying sequencing batch reactor”. In: *Enzyme and Microbial Technology* 39.7 (Nov. 2006), pp. 1392–1398. DOI: 10.1016/j.enzmictec.2006.03.024.

[13] Hong Chen et al. “Iron Robustly Stimulates Simultaneous Nitrification and Denitrification Under Aerobic Conditions”. In: *Environmental Science & Technology* 52.3 (Jan. 2018), pp. 1404–1412. DOI: 10.1021/acs.est.7b04751.

[14] Jianwei Chen and Marc Strous. “Denitrification and aerobic respiration, hybrid electron transport chains and co-evolution”. In: *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1827.2 (Feb. 2013), pp. 136–144. DOI: 10.1016/j.bbabio.2012.10.002.

[15] Tadeo Moreno Chicano et al. “Structural and functional characterization of the intracellular filament-forming nitrite oxidoreductase multiprotein complex”. In: *Nature Microbiology* 6.9 (July 2021), pp. 1129–1139. DOI: 10.1038/s41564-021-00934-8.

[16] James R. Cole et al. “Ribosomal Database Project: data and tools for high throughput rRNA analysis”. In: *Nucleic Acids Research* 42.D1 (Nov. 2013), pp. D633–D642. DOI: 10.1093/nar/gkt1244.

[17] John J. Cotter et al. “Oxygen-Mediated Regulation of Biofilm Development Is Controlled by the Alternative Sigma Factor B in Staphylococcus epidermidis”. In: *Applied and Environmental Microbiology* 75.1 (Jan. 2009), pp. 261–264. DOI: 10.1128/aem.00261-08.

[18] Jaime A. Cury and Hyun Koo. “Extraction and purification of total RNA from Sreptococcus mutans biofilms”. In: *Analytical Biochemistry* 365.2 (June 2007), pp. 208–214. DOI: 10.1016/j.ab.2007.03.021.

[19] Francesca Cutruzzolà et al. “NO Production by Pseudomonas aeruginosa cd1 Nitrite Reductase”. In: *IUBMB Life* 55.10 (Jan. 2004), pp. 617–621. DOI: 10.1080/15216540310001628672.

[20] Glen T. Daigger and Helen X. Littleton. “Simultaneous Biological Nutrient Removal: A State-of-the-Art Review”. In: *Water Environment Research* 86.3 (Mar. 2014), pp. 245–257. DOI: 10.2175/106143013x13736496908555.

[21] Holger Daims, Sebastian Lücker, and Michael Wagner. “A New Perspective on Microbes Formerly Known as Nitrite-Oxidizing Bacteria”. In: *Trends in Microbiology* 24.9 (Sept. 2016), pp. 699–712. DOI: 10.1016/j.tim.2016.05.004.

[22] Holger Daims et al. “Complete nitrification by Nitrospira bacteria”. In: *Nature* 528.7583 (Nov. 2015), pp. 504–509. DOI: 10.1038/nature16461.

[23] Matthias Dreier et al. “High-throughput qPCR and 16S rRNA gene amplicon sequencing as complementary methods for the investigation of the cheese microbiota”. In: *BMC Microbiology* 22.1 (Feb. 2022). DOI: 10.1186/s12866-022-02451-y.

[24] Shoupeng Duan, Yanyan Zhang, and Shaokui Zheng. "Heterotrophic nitrifying bacteria in wastewater biological nitrogen removal systems: A review". In: *Critical Reviews in Environmental Science and Technology* (Feb. 2021), pp. 1–37. DOI: 10.1080/10643389.2021.1877976.

[25] Morten Kam Dahl Dueholm et al. "MiDAS 4: A global catalogue of full-length 16S rRNA gene sequences and taxonomy for studies of bacterial communities in wastewater treatment plants". In: *Nature Communications* 13.1 (Apr. 2022). DOI: 10.1038/s41467-022-29438-7.

[26] Robert C. Edgar. "Search and clustering orders of magnitude faster than BLAST". In: *Bioinformatics* 26.19 (Aug. 2010), pp. 2460–2461. DOI: 10.1093/bioinformatics/btq461.

[27] Oliver Einsle et al. "Structure of cytochrome c nitrite reductase". In: *Nature* 400.6743 (July 1999), pp. 476–480. DOI: 10.1038/22802.

[28] S A Ensign, M R Hyman, and D J Arp. "In vitro activation of ammonia monooxygenase from Nitrosomonas europaea by copper". In: *Journal of Bacteriology* 175.7 (Apr. 1993), pp. 1971–1980. DOI: 10.1128/jb.175.7.1971-1980.1993.

[29] Daniel P. Faith. "Conservation evaluation and phylogenetic diversity". In: *Biological Conservation* 61.1 (1992), pp. 1–10. DOI: 10.1016/0006-3207(92)91201-3.

[30] M. Laura Fernández, Darío A. Estrin, and Sara E. Bari. "Theoretical insight into the hydroxylamine oxidoreductase mechanism". In: *Journal of Inorganic Biochemistry* 102.7 (July 2008), pp. 1523–1530. DOI: 10.1016/j.jinorgbio.2008.01.032.

[31] C. A. Francis et al. "Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean". In: *Proceedings of the National Academy of Sciences* 102.41 (Sept. 2005), pp. 14683–14688. DOI: 10.1073/pnas.0506625102.

[32] Bo Fu et al. "Characterization of microbial community in an aerobic moving bed biofilm reactor applied for simultaneous nitrification and denitrification". In: *World Journal of Microbiology and Biotechnology* 26.11 (Mar. 2010), pp. 1981–1990. DOI: 10.1007/s11274-010-0382-y.

[33] Emily Garner et al. "Next generation sequencing approaches to evaluate water and wastewater quality". In: *Water Research* 194 (Apr. 2021), p. 116907. DOI: 10.1016/j.watres.2021.116907.

[34] Andrew J. Gates et al. "Electrocatalytic reduction of nitrate and selenate by NapAB". In: *Biochemical Society Transactions* 39.1 (Jan. 2011), pp. 236–242. DOI: 10.1042/bst0390236.

[35] D.S. Goodsell. "Cytochrome bc1". In: *RCSB Protein Data Bank* (May 2011). DOI: 10.2210/rcsb_pdb/mom_2011_5. URL: https://pdb101.rcsb.org/motm/137.

[36] Michael Grabchak et al. "The generalized Simpson's entropy is a measure of biodiversity". In: *PLOS ONE* 12.3 (Mar. 2017). Ed. by Stefan J. Green, e0173305. DOI: 10.1371/journal.pone.0173305.

[37] CP Leslie Grady Jr et al. *Biological wastewater treatment*. CRC press, 2011.

[38] Susanne Günther, Michael Grunert, and Susann Müller. "Overview of recent advances in phosphorus recovery for fertilizer production". In: *Engineering in Life Sciences* 18.7 (May 2018), pp. 434–439. DOI: 10.1002/elsc.201700171.

[39] Harry R. Harhangi et al. "Hydrazine Synthase, a Unique Phylomarker with Which To Study the Presence and Biodiversity of Anammox Bacteria". In: *Applied and Environmental Microbiology* 78.3 (Feb. 2012), pp. 752–758. DOI: 10.1128/aem.07113-11.

[40] J. Heil, H. Vereecken, and N. Brüggemann. "A review of chemical reactions of nitrification intermediates and their role in nitrogen cycling and nitrogen trace gas formation in soil". In: *European Journal of Soil Science* 67.1 (Nov. 2015), pp. 23–39. DOI: 10.1111/ejss.12306.

[41] HIAS IKS. "METHOD FOR BIOLOGICAL PURIFICATION OF WASTE WATER". NO20140660. 2015.

[42] T. Hino et al. "Structural Basis of Biological N2O Generation by Bacterial Nitric Oxide Reductase". In: *Science* 330.6011 (Nov. 2010), pp. 1666–1670. DOI: 10.1126/science.1195591.

[43] Daisuke Hira et al. "Anammox organism KSU-1 expresses a NirK-type copper-containing nitrite reductase instead of a NirS-type with cytochromecd1". In: *FEBS Letters* 586.11 (May 2012), pp. 1658–1663. DOI: 10.1016/j.febslet.2012.04.041.

[44] Bin Ji et al. "Aerobic denitrification: A review of important advances of the last 30 years". In: *Biotechnology and Bioprocess Engineering* 20.4 (Aug. 2015), pp. 643–651. DOI: 10.1007/s12257-015-0009-0.

[45] Yangyang Jia et al. "Rare Taxa Exhibit Disproportionate Cell-Level Metabolic Activity in Enriched Anaerobic Digestion Microbial Communities". In: *mSystems* 4.1 (Feb. 2019). Ed. by Catherine Lozupone. DOI: 10.1128/msystems.00208-18.

[46] Mart KAMP et al. "Involvement of the hydrophobic patch of azurin in the electron-transfer reactions with cytochrome c551 and nitrite reductase". In: *European Journal of Biochemistry* 194.1 (Nov. 1990), pp. 109–118. DOI: 10.1111/j.1432-1033.1990.tb19434.x.

[47] Boran Kartal et al. "Molecular mechanism of anaerobic ammonium oxidation". In: *Nature* 479.7371 (Oct. 2011), pp. 127–130. DOI: 10.1038/nature10453.

[48] Than Khin and Ajit P. Annachhatre. "Novel microbial nitrogen removal processes". In: *Biotechnology Advances* 22.7 (Sept. 2004), pp. 519–532. DOI: 10.1016/j.biotechadv.2004.04.003.

[49] John B. Kirkpatrick et al. "Concurrent activity of anammox and denitrifying bacteria in the Black Sea". In: *Frontiers in Microbiology* 3 (2012). DOI: 10.3389/fmicb.2012.00256.

[50] Roger Knowles. "Denitrification". In: *Microbiological reviews* 46.1 (1982), pp. 43–70.

[51] Hanna Koch, Maartje A. H. J. van Kessel, and Sebastian Lücker. "Complete nitrification: insights into the ecophysiology of comammox Nitrospira". In: *Applied Microbiology and Biotechnology* 103.1 (Nov. 2018), pp. 177–189. DOI: 10.1007/s00253-018-9486-3.

[52] M. Koutný. "From no-confidence to nitric oxide acknowledgement: A story of bacterial nitric-oxide reductase". In: *Folia Microbiologica* 45.3 (June 2000), pp. 197–203. DOI: 10.1007/bf02908943.

[53] G A Kowalchuk et al. "Analysis of ammonia-oxidizing bacteria of the beta subdivision of the class Proteobacteria in coastal sand dunes by denaturing gradient gel electrophoresis and sequencing of PCR-amplified 16S ribosomal DNA fragments". In: *Applied and Environmental Microbiology* 63.4 (Apr. 1997), pp. 1489–1497. DOI: 10.1128/aem.63.4.1489-1497.1997.

[54] P. Lam et al. "Revising the nitrogen cycle in the Peruvian oxygen minimum zone". In: *Proceedings of the National Academy of Sciences* 106.12 (Mar. 2009), pp. 4752–4757. DOI: 10.1073/pnas.0812444106.

[55] Xudong Lang et al. "Aerobic denitrifiers with petroleum metabolizing ability isolated from caprolactam sewage treatment pool". In: *Bioresource Technology* 290 (Oct. 2019), p. 121719. DOI: 10.1016/j.biortech.2019.121719.

[56] Yingyu Law et al. "Nitrous oxide emissions from wastewater treatment processes". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1593 (May 2012), pp. 1265–1277. DOI: 10.1098/rstb.2011.0317.

[57] Ziyuan Lin et al. "The variation on nitrogen removal mechanisms and the succession of ammonia oxidizing archaea and ammonia oxidizing bacteria with temperature in biofilm reactors treating saline wastewater". In: *Bioresource Technology* 314 (Oct. 2020), p. 123760. DOI: 10.1016/j.biortech.2020.123760.

[58] Xinying Liu et al. "Heterotrophic Nitrifiers Dominate Reactors Treating Incineration Leachate with High Free Ammonia Concentrations". In: *ACS Sustainable Chemistry & Engineering* 6.11 (Oct. 2018), pp. 15040–15049. DOI: 10.1021/acssuschemeng.8b03512.

[59] YuXiang Liu et al. "Heterotrophic nitrogen removal by Acinetobacter sp. Y1 isolated from coke plant wastewater". In: *Journal of Bioscience and Bioengineering* 120.5 (Nov. 2015), pp. 549–554. DOI: 10.1016/j.jbiosc.2015.03.015.

[60] Stilianos Louca, Michael Doebeli, and Laura Wegener Parfrey. "Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem". In: *Microbiome* 6.1 (Feb. 2018). DOI: 10.1186/s40168-018-0420-9.

[61] S. Lucker et al. "A Nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria". In: *Proceedings of the National Academy of Sciences* 107.30 (July 2010), pp. 13479–13484. DOI: 10.1073/pnas.1003860107.

[62] Yabing Meng, Li-Nan Huang, and Fangang Meng. "Metagenomics Response of Anaerobic Ammonium Oxidation (anammox) Bacteria to Bio-Refractory Humic Substances in Wastewater". In: *Water* 11.2 (Feb. 2019), p. 365. DOI: 10.3390/w11020365.

[63] Conrado Moreno-Vivian et al. "Prokaryotic Nitrate Reduction: Molecular Properties and Functional Distinction among Bacterial Nitrate Reductases". In: *Journal of Bacteriology* 181.21 (Nov. 1999), pp. 6573–6584. DOI: 10.1128/jb.181.21.6573-6584.1999.

[64] Elisabeth V. Münch, Paul Lant, and Jürg Keller. "Simultaneous nitrification and denitrification in bench-scale sequencing batch reactors". In: *Water Research* 30.2 (Feb. 1996), pp. 277–284. DOI: 10.1016/0043-1354(95)00174-3.

[65] Michael E.P. Murphy, Stewart Turley, and Elinor T. Adman. "Structure of Nitrite Bound to Copper-containing Nitrite Reductase from Alcaligenes faecalis". In: *Journal of Biological Chemistry* 272.45 (Nov. 1997), pp. 28455–28460. DOI: 10.1074/jbc.272.45.28455.

[66] Francesco Musiani et al. "The model structure of the copper-dependent ammonia monooxygenase". In: *JBIC Journal of Biological Inorganic Chemistry* 25.7 (Sept. 2020), pp. 995–1007. DOI: 10.1007/s00775-020-01820-0.

[67] Tristan Nicke et al. "Maturation of the cytochrome cd1 nitrite reductase NirS from Pseudomonas aeruginosa requires transient interactions between the three proteins NirS, NirN and NirF". In: *Bioscience Reports* 33.3 (June 2013). DOI: 10.1042/bsr20130043.

[68] Torsten Ochsenreiter et al. "Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR". In: *Environmental Microbiology* 5.9 (Sept. 2003), pp. 787–797. DOI: 10.1046/j.1462-2920.2003.00476.x.

[69] A OEHMEN et al. "Advances in enhanced biological phosphorus removal: From micro to macro scale". In: *Water Research* 41.11 (June 2007), pp. 2271–2300. DOI: 10.1016/j.watres.2007.02.030.

[70] Jari Oksanen et al. *vegan: Community Ecology Package*. R package version 2.6-2. 2022. URL: https://CRAN.R-project.org/package=vegan.

[71] Byoung-Joon Park et al. "Cultivation of Autotrophic Ammonia-Oxidizing Archaea from Marine Sediments in Coculture with Sulfur-Oxidizing Bacteria". In: *Applied and Environmental Microbiology* 76.22 (Nov. 2010), pp. 7575–7587. DOI: 10.1128/aem.01478-10.

[72] Courtney N. Passow et al. "Nonrandom RNAseq gene expression associated with RNAlater and flash freezing storage methods". In: *Molecular Ecology Resources* 19.2 (Dec. 2018), pp. 456–464. DOI: 10.1111/1755-0998.12965.

[73] Sofia R. Pauleta, Simone Dell'Acqua, and Isabel Moura. "Nitrous oxide reductase". In: *Coordination Chemistry Reviews* 257.2 (Jan. 2013), pp. 332–349. DOI: 10.1016/j.ccr.2012.05.026.

[74] Isobel V. Pearson et al. "A Mutant of Paracoccus denitrificans with Disrupted Genes Coding for Cytochrome c 550 and Pseudoazurin Establishes These Two Proteins as the In Vivo Electron Donors to Cytochrome cd 1 Nitrite Reductase". In: *Journal of Bacteriology* 185.21 (Nov. 2003), pp. 6308–6315. DOI: 10.1128/jb.185.21.6308-6315.2003.

[75] M. Pijuan et al. "Aerobic phosphorus release linked to acetate uptake: Influence of PAO intracellular storage compounds". In: *Biochemical Engineering Journal* 26.2-3 (Nov. 2005), pp. 184–190. DOI: 10.1016/j.bej.2005.04.014.

[76] Adam R Platt, Robert W. Woodhall, and Alfred L. George. "Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol". In: *BioTechniques* 43.1 (July 2007), pp. 58–62. DOI: 10.2144/000112499.

[77] E. Pruesse et al. "SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB". In: *Nucleic Acids Research* 35.21 (Nov. 2007), pp. 7188–7196. DOI: 10.1093/nar/gkm864.

[78] Massimo Raboni, Andrea G. Capodaglio, and Petr Hlavínek. "Physico-chemical technologies for nitrogen removal from wastewaters: a review". In: *Ambiente e Agua - An Interdisciplinary Journal of Applied Science* 10.3 (July 2015). DOI: 10.4136/ambi-agua.1618.

[79] Sundas Rani et al. "Detection and Diversity of the Nitrite Oxidoreductase Alpha Subunit (nxrA) Gene of Nitrospina in Marine Sediments". In: *Microbial Ecology* 73.1 (Nov. 2016), pp. 111–122. DOI: 10.1007/s00248-016-0897-3.

[80] J H Rotthauwe, K P Witzel, and W Liesack. "The ammonia monooxygenase structural gene amoA as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations". In: *Applied and Environmental Microbiology* 63.12 (Dec. 1997), pp. 4704–4712. DOI: 10.1128/aem.63.12.4704-4712.1997.

[81] Samarpita Roy et al. "Recent advances in understanding the ecophysiology of enhanced biological phosphorus removal". In: *Current Opinion in Biotechnology* 67 (Feb. 2021), pp. 166–174. DOI: 10.1016/j.copbio.2021.01.011.

[82] Knut Rudi et al. "Microbial ecological processes in MBBR biofilms for biological phosphorus removal from wastewater". In: *Water Science and Technology* 79.8 (Apr. 2019), pp. 1467–1473. DOI: 10.2166/wst.2019.149.

[83] Dimitra Sakoula et al. "Enrichment and physiological characterization of a novel comammox Nitrospira indicates ammonium inhibition of complete nitrification". In: *The ISME Journal* 15.4 (Nov. 2020), pp. 1010–1024. DOI: 10.1038/s41396-020-00827-4.

[84] Giovanna Salbitani and Simona Carfagna. "Ammonium Utilization in Microalgae: A Sustainable Method for Wastewater Treatment". In: *Sustainability* 13.2 (Jan. 2021), p. 956. DOI: 10.3390/su13020956.

[85] T. Saltnes, G. Sørensen, and S. Eikås. "Biological nutrient removal in a continuous biofilm process". In: *Water Practice and Technology* 12.4 (Dec. 2017), pp. 797–805. DOI: 10.2166/wpt.2017.083.

[86] Thomas S. B. Schmidt, João F. Matias Rodrigues, and Christian Mering. "Limits to robustness and reproducibility in the demarcation of operational taxonomic units". In: *Environmental Microbiology* 17.5 (Sept. 2014), pp. 1689–1706. DOI: 10.1111/1462-2920.12610.

[87] Amal F. El Sheikh and Martin G. Klotz. "Ammonia-dependent differential regulation of the gene cluster that encodes ammonia monooxygenase inNitrosococcus oceaniATCC 19707". In: *Environmental Microbiology* 10.11 (Nov. 2008), pp. 3026–3035. DOI: 10.1111/j.1462-2920.2008.01766.x.

[88] Huazhong Shi and Ray Bressan. “RNA Extraction”. In: *Arabidopsis Protocols*. Humana Press, pp. 345–348. DOI: 10.1385/1-59745-003-0:345.

[89] Jörg Simon and Martin G. Klotz. “Diversity and evolution of bioenergetic systems involved in microbial nitrogen compound transformations”. In: *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1827.2 (Feb. 2013), pp. 114–135. DOI: 10.1016/j.bbabio.2012.07.005.

[90] Pamela M. Smith, Jennifer L. Fox, and Dennis R. Winge. “Reprint of: Biogenesis of the cytochrome bc1 complex and role of assembly factors”. In: *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1817.6 (June 2012), pp. 872–882. DOI: 10.1016/j.bbabio.2012.03.003.

[91] Aina Soler-Jofra, Julio Pérez, and Mark C.M. van Loosdrecht. “Hydroxylamine and the nitrogen cycle: A review”. In: *Water Research* 190 (Feb. 2021), p. 116723. DOI: 10.1016/j.watres.2020.116723.

[92] L Y Stein, D J Arp, and M R Hyman. “Regulation of the Synthesis and Activity of Ammonia Monooxygenase in Nitrosomonas europaea by Altering pH To Affect NH(inf3) Availability”. In: *Applied and Environmental Microbiology* 63.11 (Nov. 1997), pp. 4588–4592. DOI: 10.1128/aem.63.11.4588-4592.1997.

[93] Lisa Y Stein. “Insights into the physiology of ammonia-oxidizing microorganisms”. In: *Current Opinion in Chemical Biology* 49 (Apr. 2019), pp. 9–15. DOI: 10.1016/j.cbpa.2018.09.003.

[94] A. Szabó et al. “Significance of Design and Operational Variables in Chemical Phosphorus Removal”. In: *Water Environment Research* 80.5 (May 2008), pp. 407–416. DOI: 10.2175/106143008x268498.

[95] Shunsuke Takahashi et al. “Development of a Prokaryotic Universal Primer for Simultaneous Analysis of Bacteria and Archaea Using Next-Generation Sequencing”. In: *PLoS ONE* 9.8 (Aug. 2014). Ed. by Kostas Bourtzis, e105592. DOI: 10.1371/journal.pone.0105592.

[96] Sandra Tan, Hwee Tong Tan, and Maxey C. M. Chung. “Membrane proteins and membrane proteomics”. In: *PROTEOMICS* 8.19 (Oct. 2008), pp. 3924–3932. DOI: 10.1002/pmic.200800597.

[97] B L Trumpower. “Cytochrome bc1 complexes of microorganisms”. In: *Microbiological Reviews* 54.2 (June 1990), pp. 101–129. DOI: 10.1128/mr.54.2.101-129.1990.

[98] Tomáš Větrovský and Petr Baldrian. “The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses”. In: *PLoS ONE* 8.2 (Feb. 2013). Ed. by Josh Neufeld, e57923. DOI: 10.1371/journal.pone.0057923.

[99] Erik Vijgenboom, Julie E. Busch, and Gerard W. Canters. “In vivo studies disprove an obligatory role of azurin in denitrification in Pseudomonas aeruginosa and show that azu expression is under control of RpoS and ANR”. In: *Microbiology* 143.9 (Sept. 1997), pp. 2853–2863. DOI: 10.1099/00221287-143-9-2853.

[100] Didrik Villard et al. “Spatial fractionation of phosphorus accumulating biofilm: stratification of polyphosphate accumulation and dissimilatory nitrogen metabolism”. In: *Biofouling* (Feb. 2022), pp. 1–11. DOI: 10.1080/08927014.2022.2044475.

[101] Mingyuan Wang et al. “Newly designed primer pair revealed dominant and diverse comammox amoA gene in full-scale wastewater treatment plants”. In: *Bioresource Technology* 270 (Dec. 2018), pp. 580–587. DOI: 10.1016/j.biortech.2018.09.089.

[102] Shanyun Wang et al. “Anammox Bacterial Abundance, Activity, and Contribution in Riparian Sediments of the Pearl River Estuary”. In: *Environmental Science & Technology* 46.16 (Aug. 2012), pp. 8834–8842. DOI: 10.1021/es3017446.

[103] Xiujie Wang et al. “Simultaneous nitrification and denitrification by a novel isolated Pseudomonas sp. JQ-H3 using polycaprolactone as carbon source”. In: *Bioresource Technology* 288 (Sept. 2019), p. 121506. DOI: 10.1016/j.biortech.2019.121506.

[104] Yubo Wang, Han Gao, and George Wells. “Integrated omics analyses reveal differential gene expression and potential for cooperation between denitrifying polyphosphate and glycogen accumulating organisms”. In: *Environmental Microbiology* 23.6 (May 2021), pp. 3274–3293. DOI: 10.1111/1462-2920.15486.

[105] Andrzej Woznica et al. “Stimulatory Effect of Xenobiotics on Oxidative Electron Transport of Chemolithotrophic Nitrifying Bacteria Used as Biosensing Element”. In: *PLoS ONE* 8.1 (Jan. 2013). Ed. by Ivo G. Boneca, e53484. DOI: 10.1371/journal.pone.0053484.

[106] Pablo Yarza et al. “Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences”. In: *Nature Reviews Microbiology* 12.9 (Aug. 2014), pp. 635–645. DOI: 10.1038/nrmicro3330.

[107] Youngseob Yu et al. “Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction”. In: *Biotechnology and Bioengineering* 89.6 (2005), pp. 670–679. DOI: 10.1002/bit.20347.

[108] Quan Yuan, Pengfei Liu, and Yahai Lu. “Differential responses of nirK- and nirS-carrying bacteria to denitrifying conditions in the anoxic rice field soil”. In: *Environmental Microbiology Reports* 4.1 (Nov. 2011), pp. 113–122. DOI: 10.1111/j.1758-2229.2011.00311.x.

[109] Tao Zhang et al. “Ammonium nitrogen removal from coking wastewater by chemical precipitation recycle technology”. In: *Water Research* 43.20 (Dec. 2009), pp. 5209–5215. DOI: 10.1016/j.watres.2009.08.054.

[110] Zheng Zhang et al. “A greedy algorithm for aligning DNA sequences”. In: *Journal of Computational biology* 7.1-2 (2000), pp. 203–214. DOI: 10.1089/10665270050081478.

[111] Guibing Zhu et al. “Biological Removal of Nitrogen from Wastewater”. In: *Reviews of Environmental Contamination and Toxicology*. Springer New York, 2008, pp. 159–195. DOI: 10.1007/978-0-387-71724-1_5.

[112] W G Zumft. “Cell biology and molecular basis of denitrification”. In: *Microbiology and Molecular Biology Reviews* 61.4 (Dec. 1997), pp. 533–616. DOI: 10.1128/mmbr.61.4.533-616.1997.

A. Appendix: qPCR

A.1 Primers

Table A.1: overvieuw of primers used

Microorganisms	Functional genes	Specific primers	Sequence (5'-3')	References	Size PCR product
AOB	amoA	amoA-1F	GGGGTTTCTACTGGTGGT	[80]	491
		amoA-2R	CCCCTCKGSAAAGCCTTCTTC		
	amoB	amoBMf	TGG TAY GAC ATK AWA TGG	[8]	500
		amoBMr	RCG SGG CAR GAA CAT SGG		
	16s rRNA (V-2/V-3 variable domains)	CTO189fA/B	GGA GRA AAG CAG GGG ATC G	[53]	465
		CTO189fC	GGA GGA AAG TAG GGG ATC G		
		CTO654r	CTA GCY TTG TAG TTT CAA ACG C		
AOA	amoA	Arch-amoAF	STAATGGTCTGGCTTAGACG	[31]	635
		Arch-amoAR	GCGGCCATCCATCTGTATGT		
	16s rRNA	771F	ACGGTGAGGGATGAAAGCT	[68]	220
		957R	CGGCGTTGACTCCAATTG		
Anaerobic AOB	hzsB	HSBeta396F	ARGGHTGGGGHAGYTGGAAG	[102]	unknown
		HSBeta742R	GTYCCHACRTCATGVGTCTG		
	nirS	Scnir372F	TGT AGC CAG CAT TGT AGC GT	[54]	unknown
		Scnir845R	TCA AGC CAG AC CAT TTG CT		
	16s rRNA	Amx368f	TTCGCAATGCCCGAAAGG	[102]	476*
		Amx820r	AAAACCCCTCTACTTAGTGCCC		
Comammox bacteria	amoA	Nino_amoA_19F	ATAATCAAAGCCGCCAAGTTGC	[22]	254
		Nino_amoA_252R	AACGGCTGACGATAATTGACC		
		comamoA AF	AGGNGAYTGGGAYTTCTGG	[101]	436
		comamoA SR	CCGVACATACATRAAGCCCAT		
HNB	amoA	heteroamo 378f	GTTGCAGGACATGCTGGTCTTCG	[58]	unknown
		heteroamo 634r	CATCGGCCAAGGATCGAGGC		

** size not disclosed by author but determined in silico*

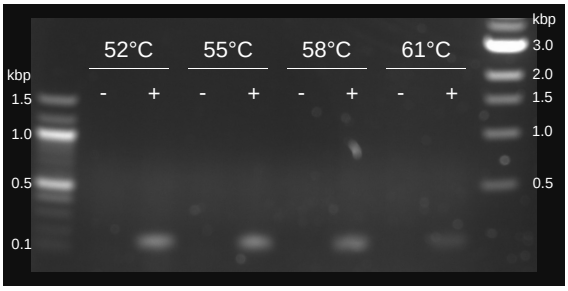
AOB: ammonia-oxidizing bacteria; AOA: ammonia-oxidizing archaea; anaerobic AOB: anaerobic ammonia-oxidizing bacteria; comammox bacteria: complete ammonia-oxidizing bacteria [24]

A.2 Optimisation and Validation of qPCR

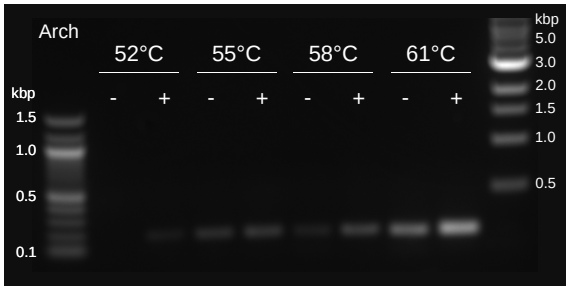
Table A.2: Table of concentration and purity of RNA extracted from biofilm fromdifferent zones of the waste water treatment plant, used for optimisation and validation of qPCR

Sample ID	ng/ul	260/280	260/230
1A	53.40	2.35	2.61
3A	43.45	1.97	1.52
3B	49.67	2.16	1.71
7A	58.77	2.06	2.03

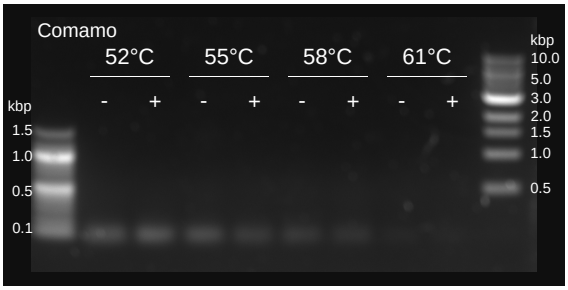
(a) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: amoB, Primers: amoBMf & amoB-Mr



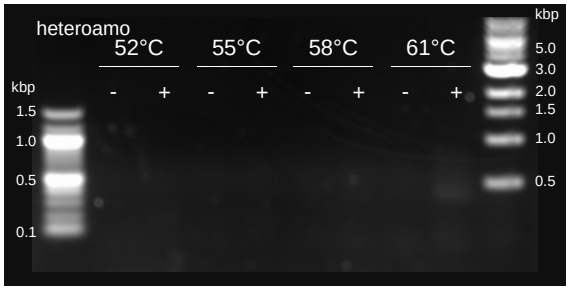
(b) Micro organisms: Anaerobe Ammonium Oxidising Bacteria, Functional Gene: 16s rRNA, Primers: Amx396F & Amx820r



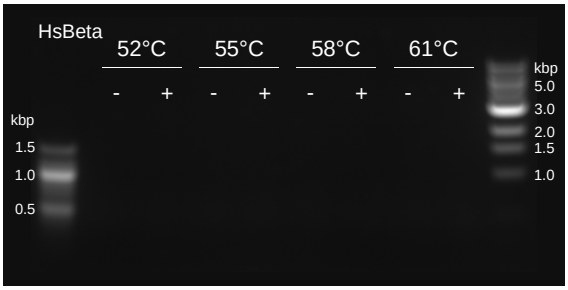
(c) Micro organisms: Ammonium Oxidising Archaea, Functional Gene: amoA, Primers: Arch-amoAF & Arch-amoAR



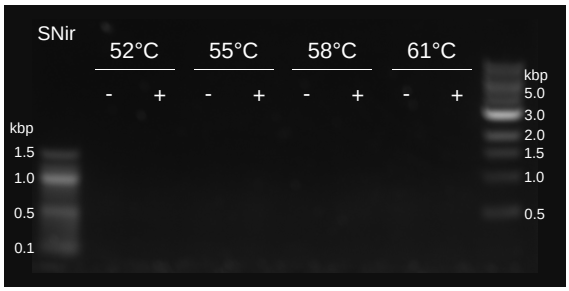
(d) Micro organisms: Complete Ammonium Oxidising Archaea, Functional Gene: amoA, Primers: comamoA AF & comamoA SR



(e) Micro organisms: Heterotrophic Nitrifying Bacteria, Functional Gene: amoA, heteroamo 378f & heteroamo 634r

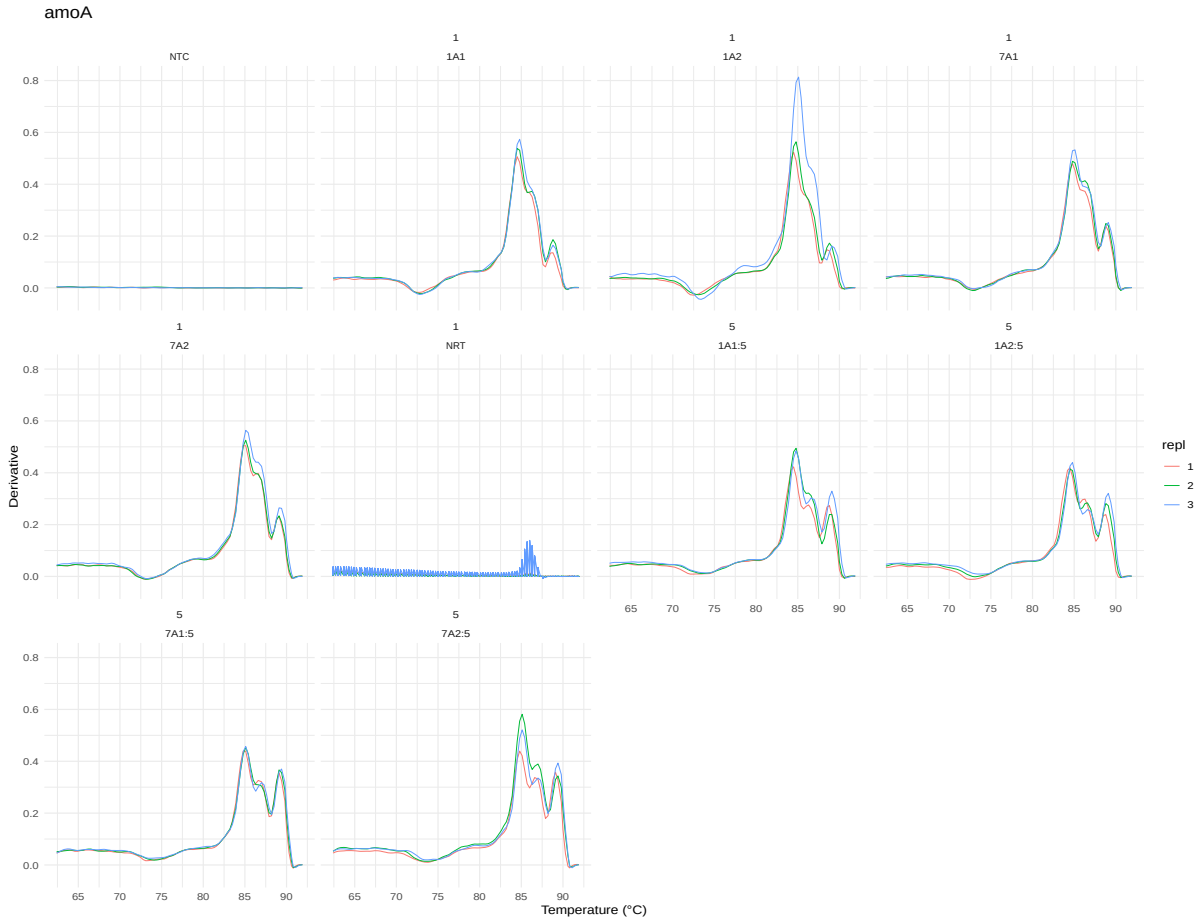


(f) Micro organisms: Anaerobe Ammonium Oxidising Bacteria, Functional Gene: hzsB, HSBeta396F & HS-Beta742R

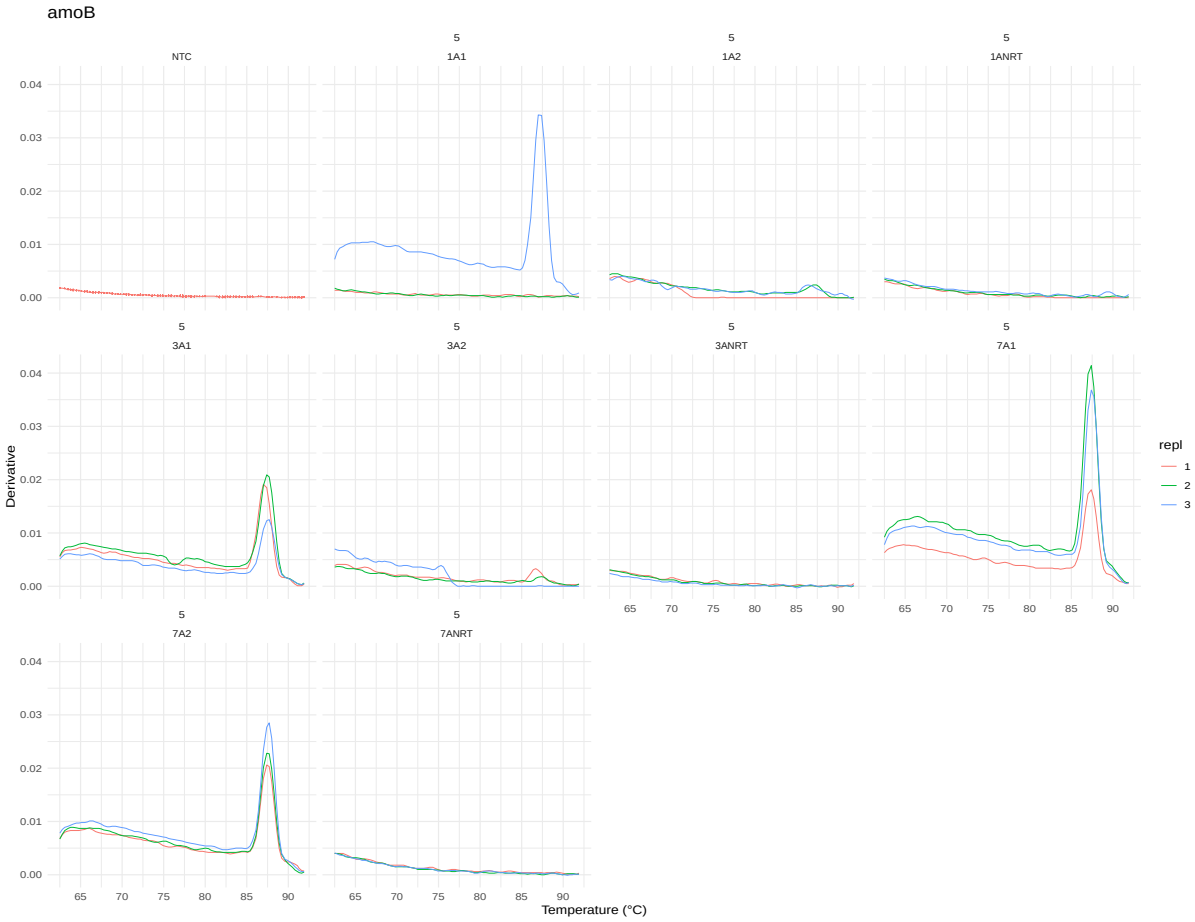


(g) Micro organisms: Anaerobe Ammonium Oxidising Bacteria, Functional Gene: nirS, Scnir372F & Scnir845R

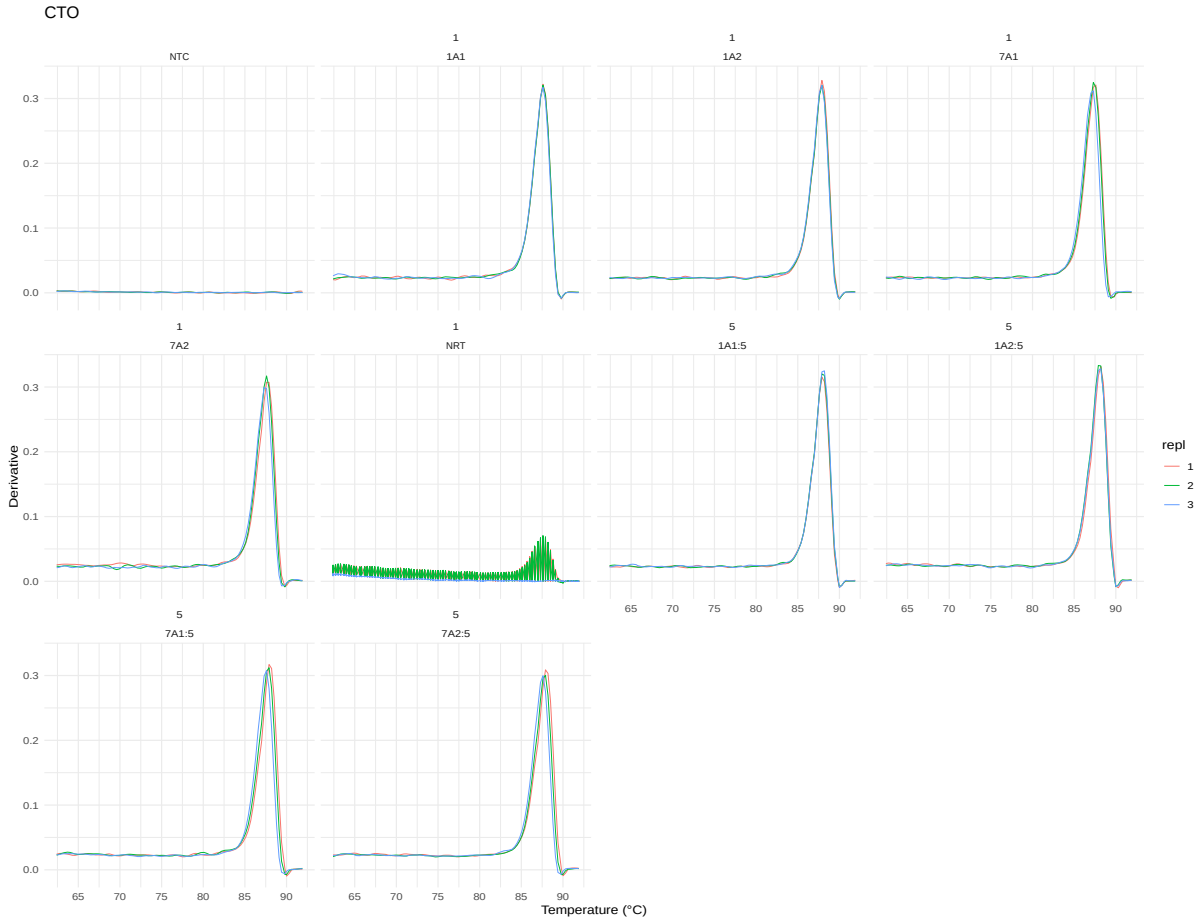
Figure A.1: Gradient PCR amplification of cDNA derived from biofilm in waste water treatment using primers specific for different genes of prokaryotes involved in waste water nitrite conversion



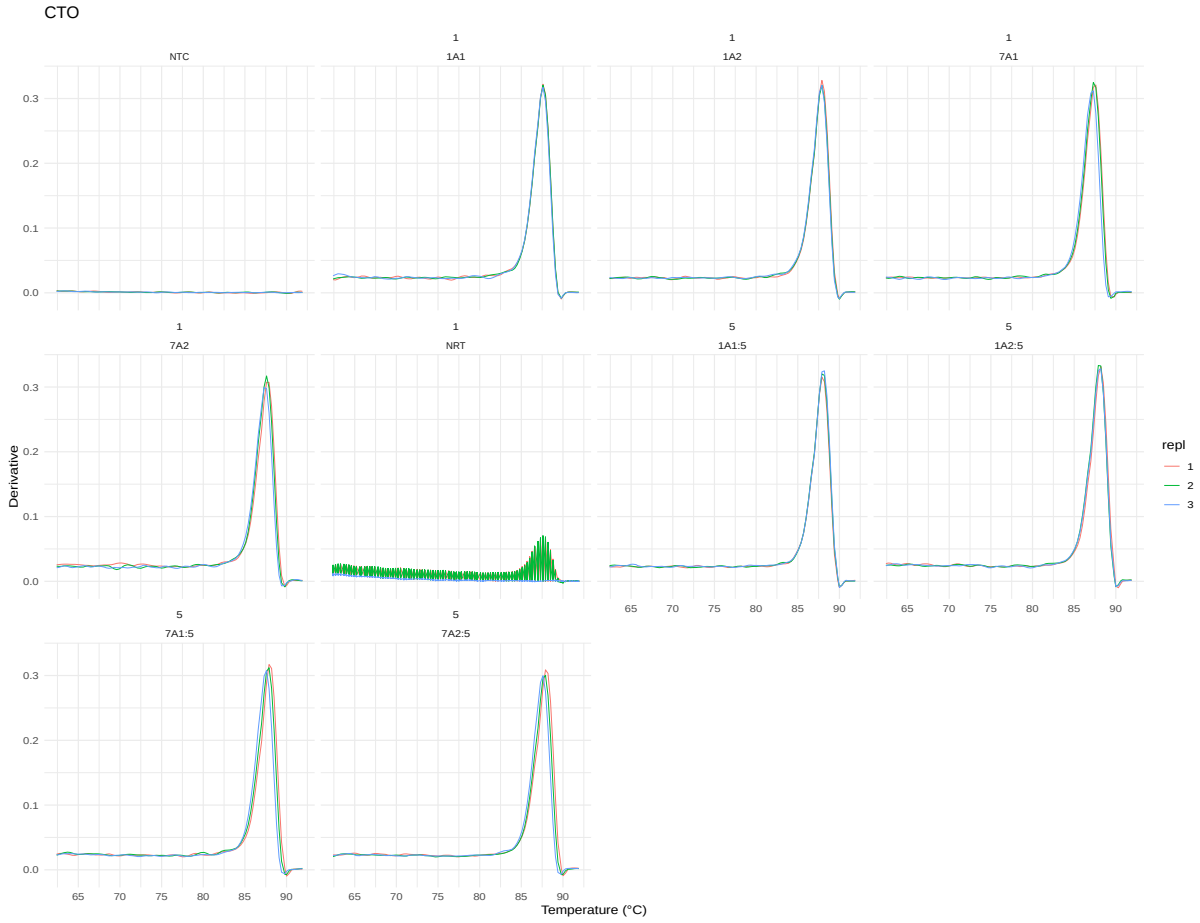
(a) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: amoA, Primers: amoA-1F & amoA-2R



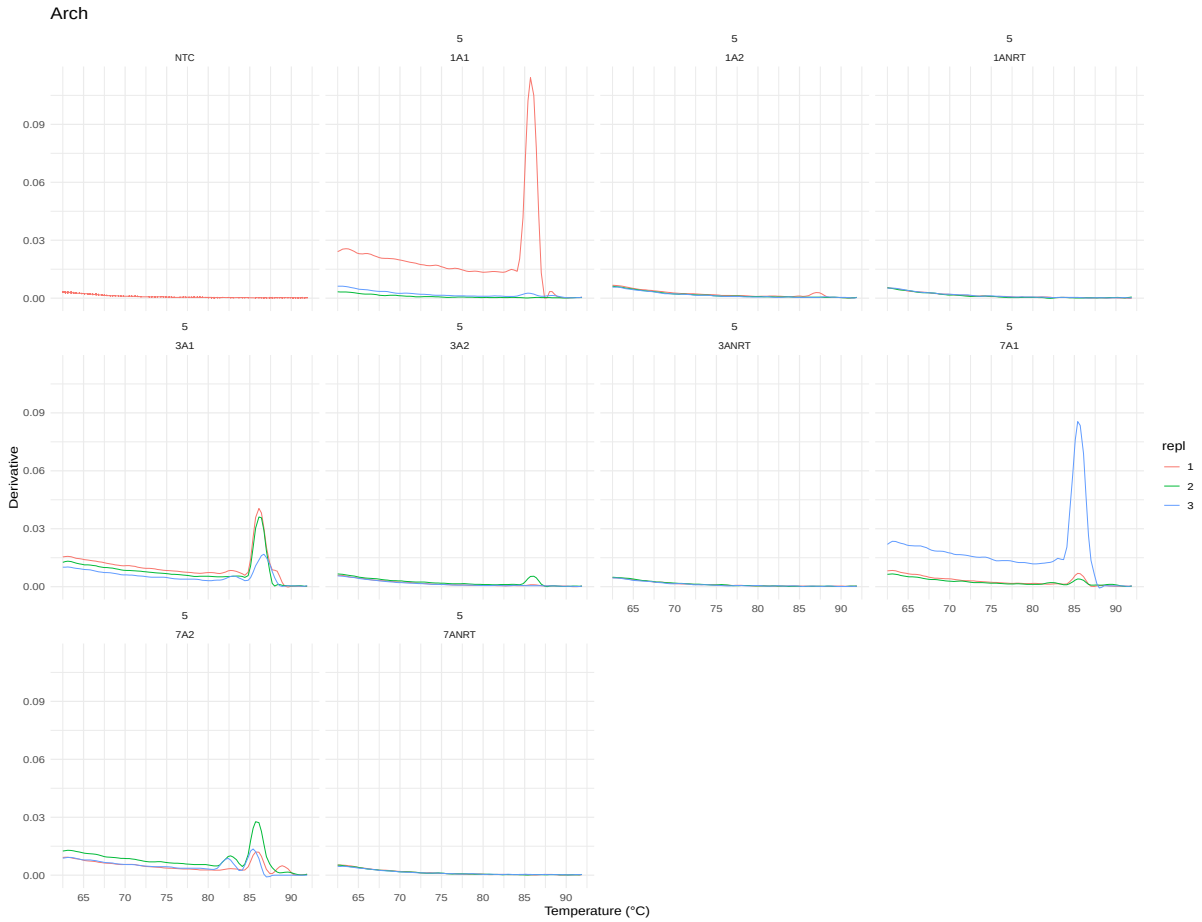
(b) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: amoB, Primers: amoBMf & amoBMr



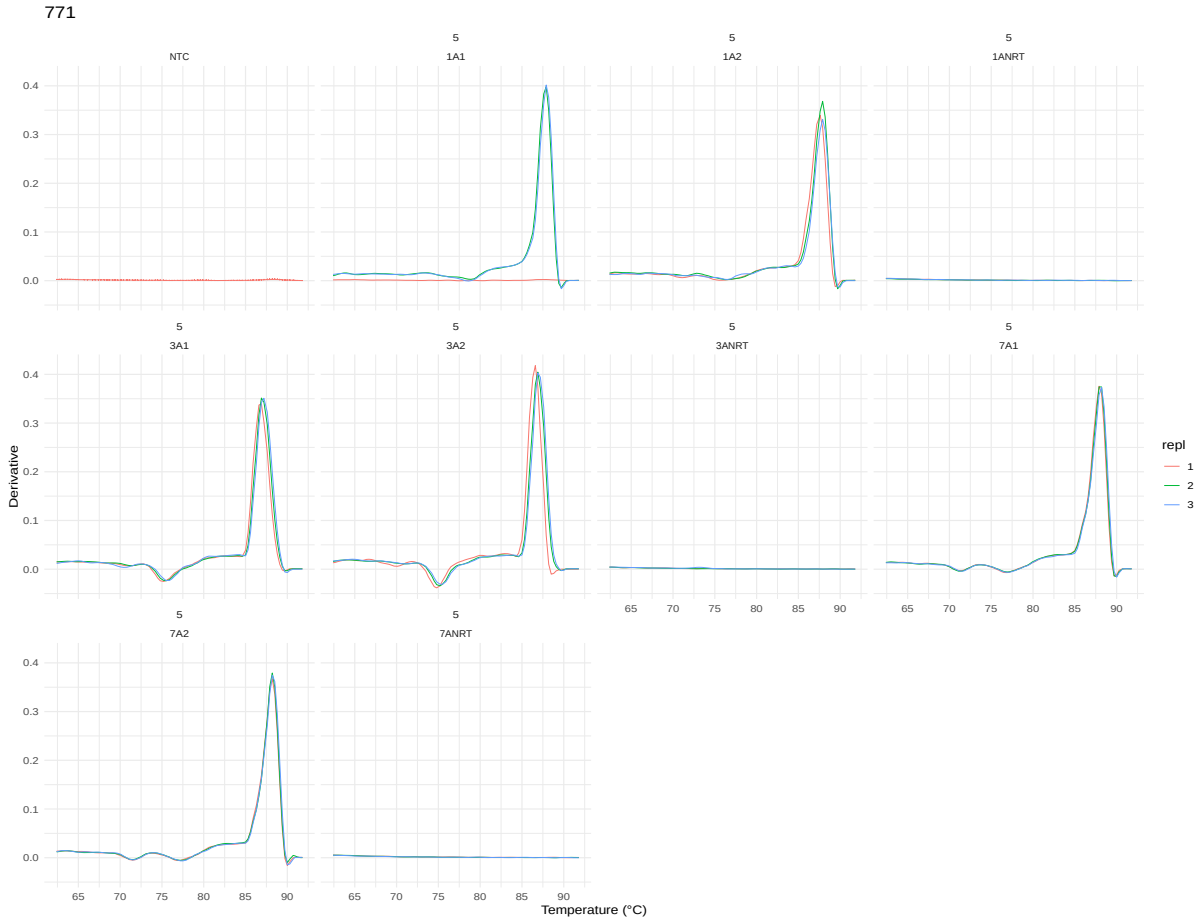
(c) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: 16s rRNA, Primers: CTO189fA/B & CTO189fC & CTO189r



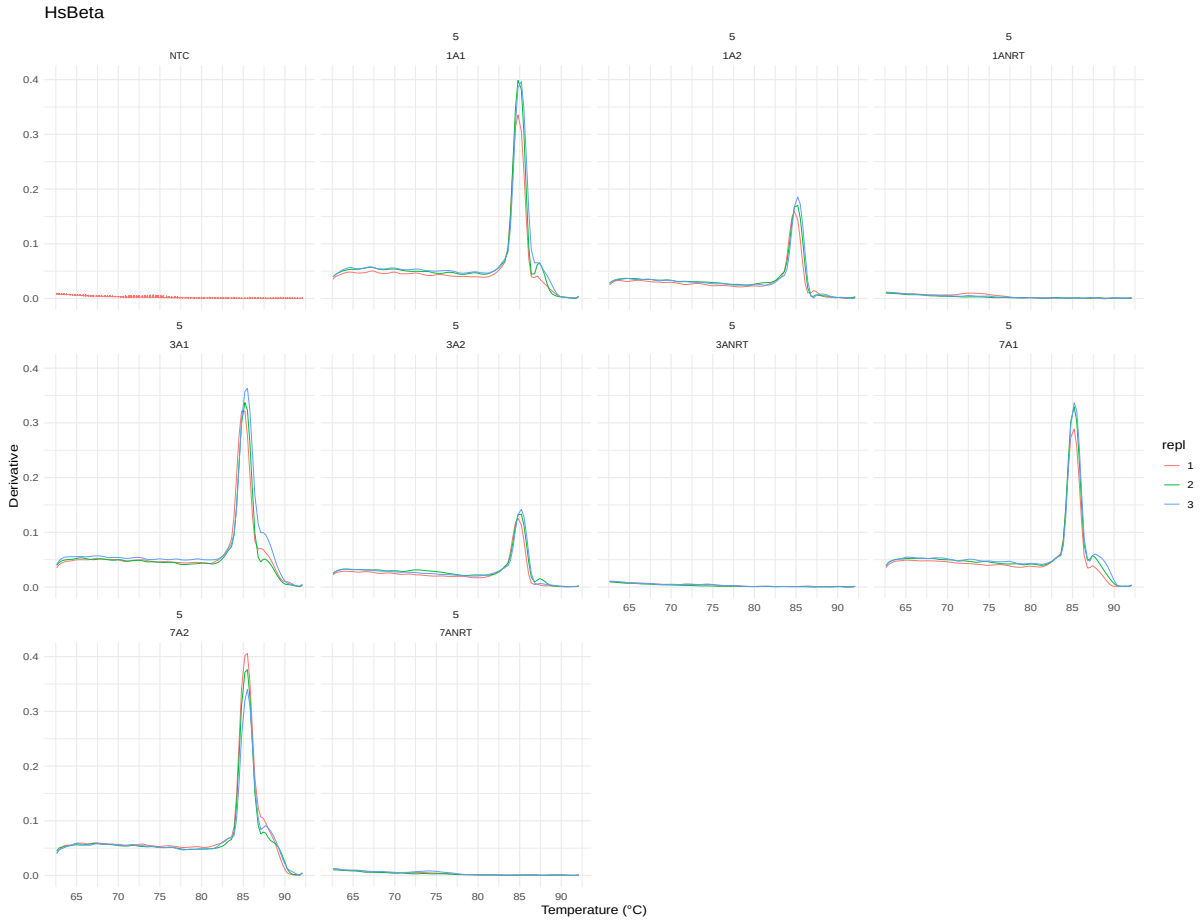
(d) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: 16s rRNA, Primers: CTO189fA/B & CTO189fC & CTO189r



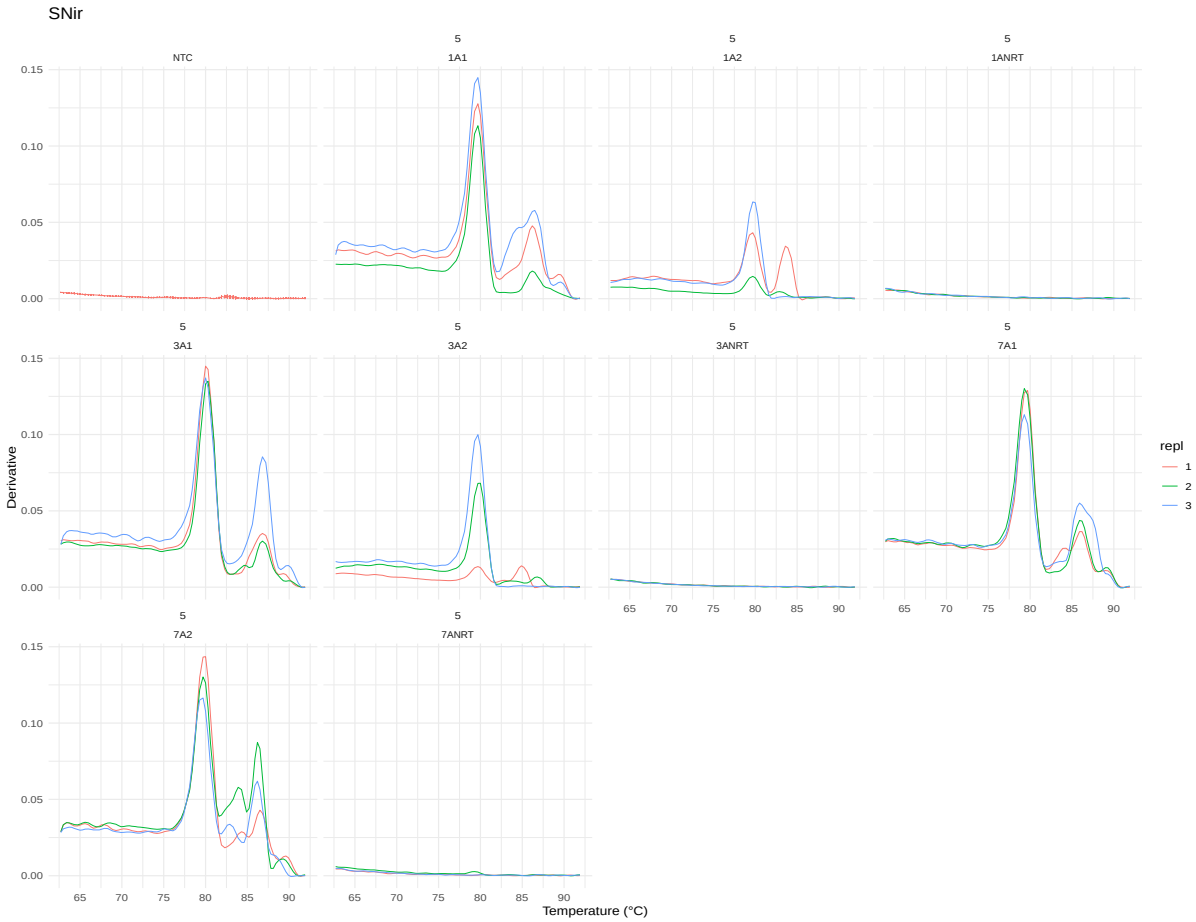
(e) Micro organisms: Ammonium Oxidising Archaea, Functional Gene: amoA, Primers: Arch-amoAF & Arch-amoAR



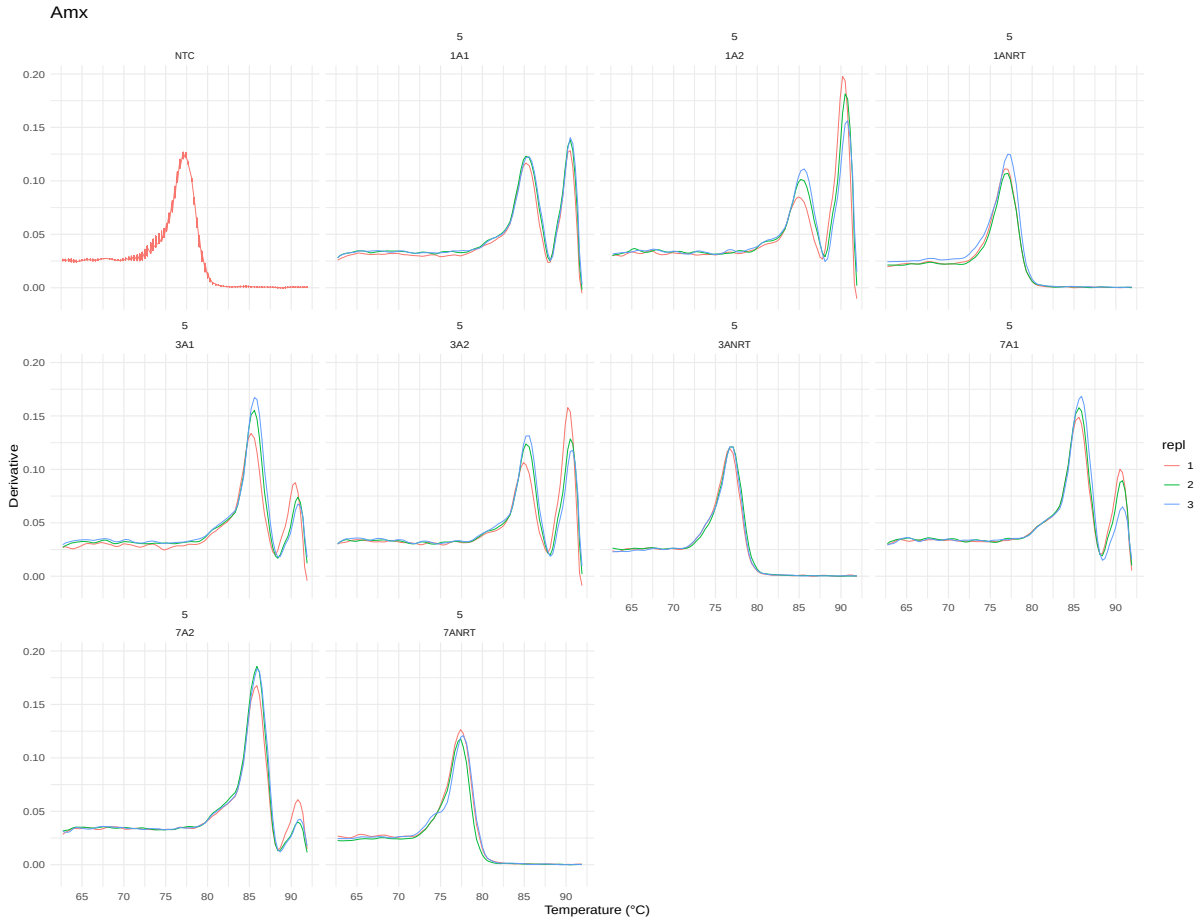
(f) Micro organisms: Ammonium Oxidising Archaea, Functional Gene: 16s rRNA, Primers: 771F & 957R



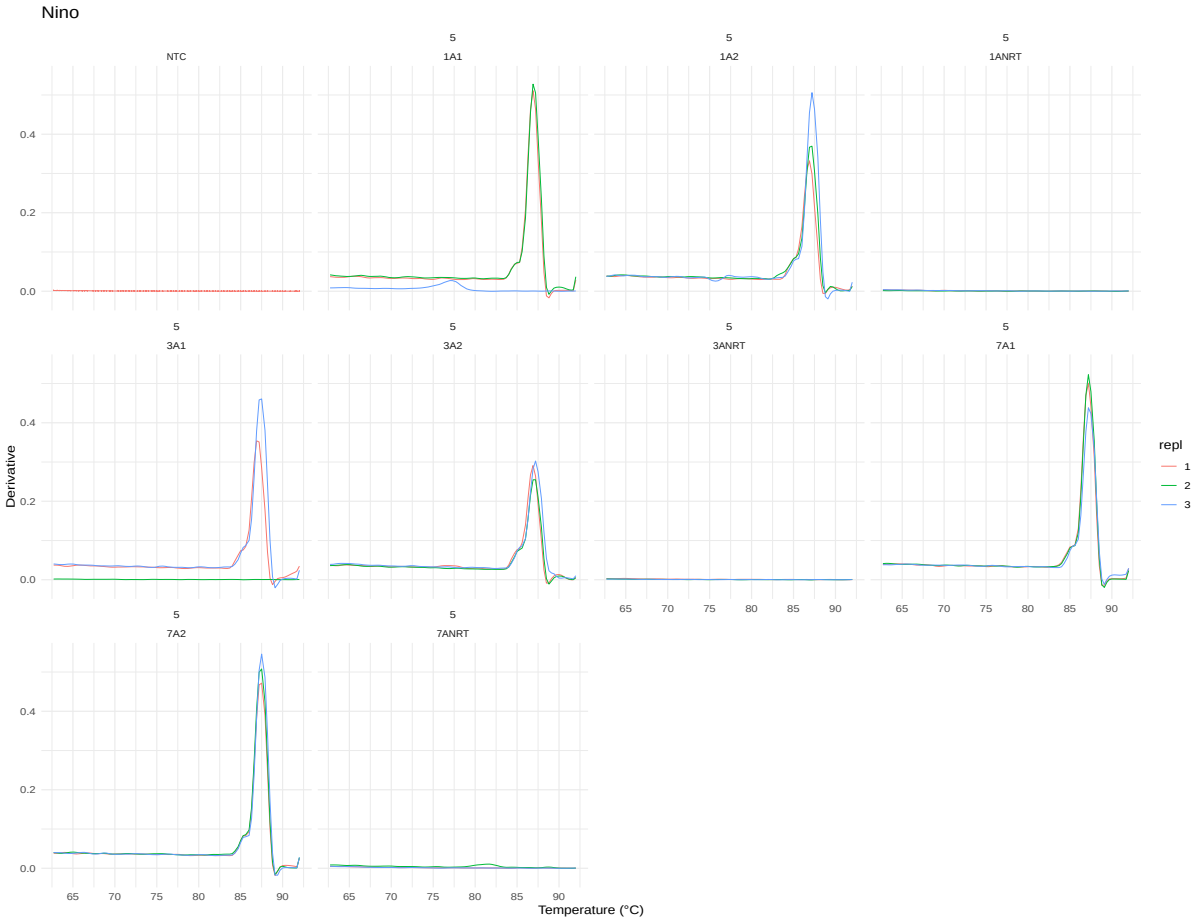
(g) Micro organisms: Anaerobic Ammonium Oxidising Bacteria, Functional Gene: hzsB, Primers: HSBeta396F & HSBeta742R



(h) Micro organisms: Anaerobic Ammonium Oxidising Bacteria, Functional Gene: nirS, Primers: Scnir372F & Scnir845R

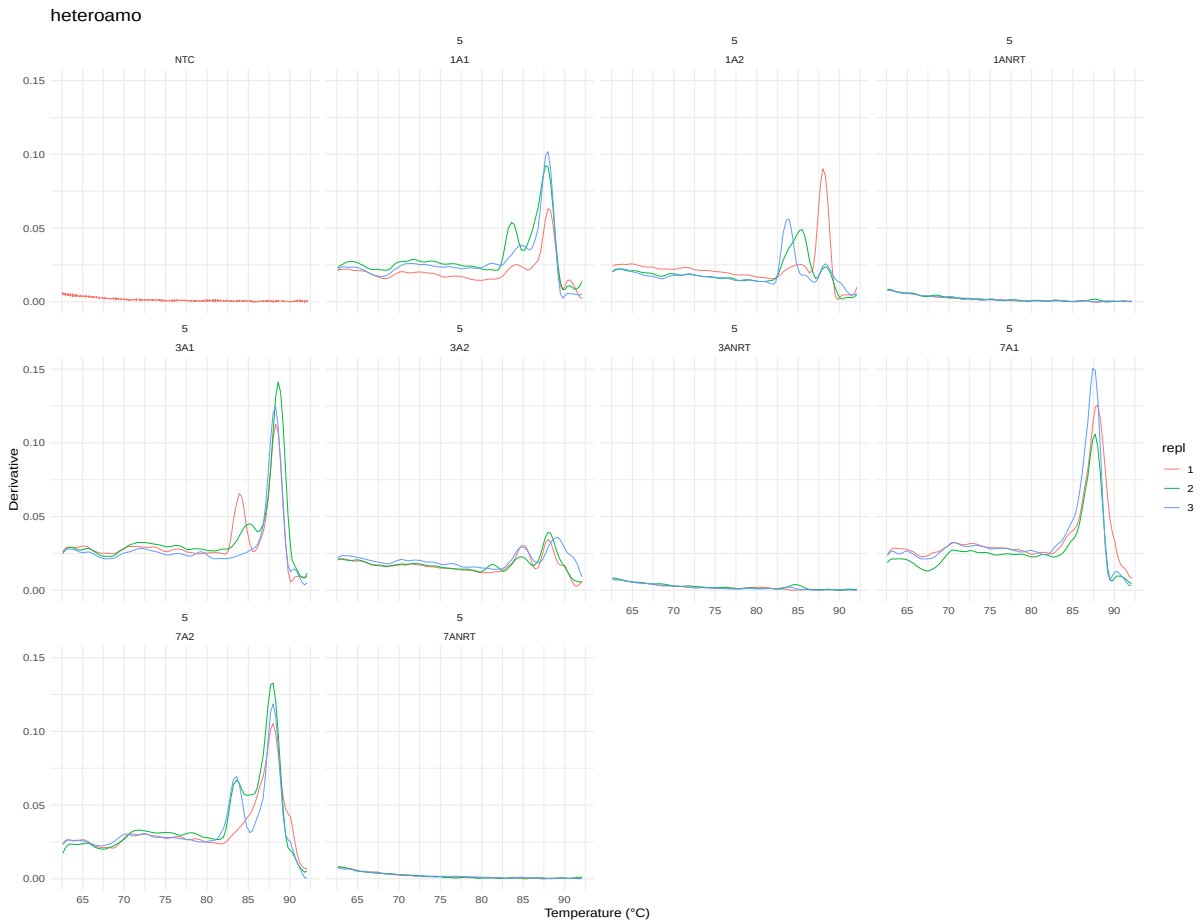


(i) Micro organisms: Anaerobic Ammonium Oxidising Bacteria, Functional Gene: 16s rRNA, Primers: Amx368f & Amx820r



(j) Micro organisms: Complete Ammonium Oxidising Bacteria, Functional Gene: amoA, Primers: Nino_amoA_19F & Nino_amoA_252R

(k) Micro organisms: Complete Ammonium Oxidising Bacteria, Functional Gene: amoA, Primers: comamoA AF & comamoA SR



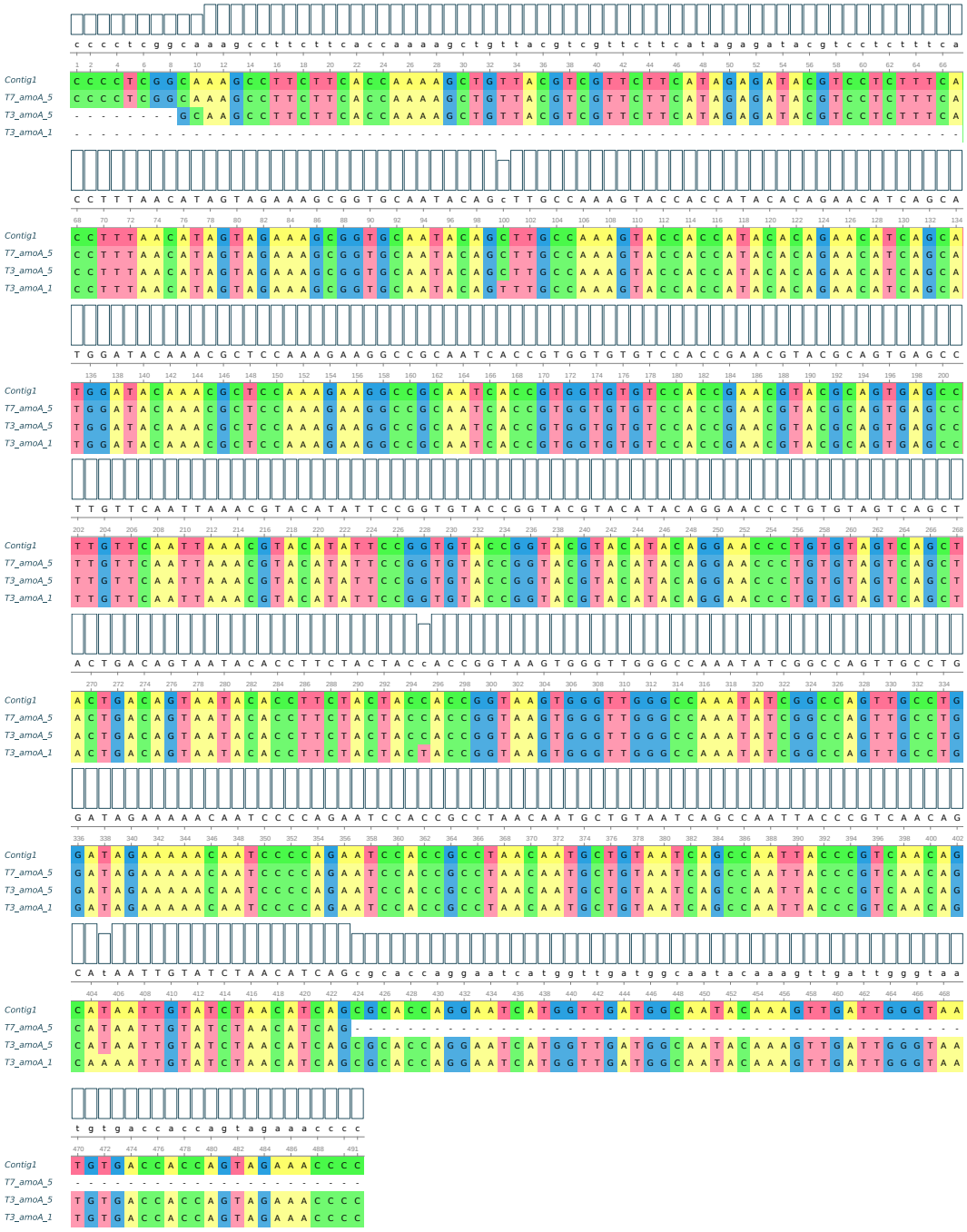
(l) Micro organisms: Heterotrophic Nitrifying Bacteria, Functional Gene: amoA, Primers: heteroamo 378f & heteroamo 634r

Figure A.2: Dissociation curves of qPCR reactions preformed with different primer pairs on cDNA made from RNA extracted from biofilm from waste water treatment reactor. No template control (NTC), No reverse transcriptase (NRT)

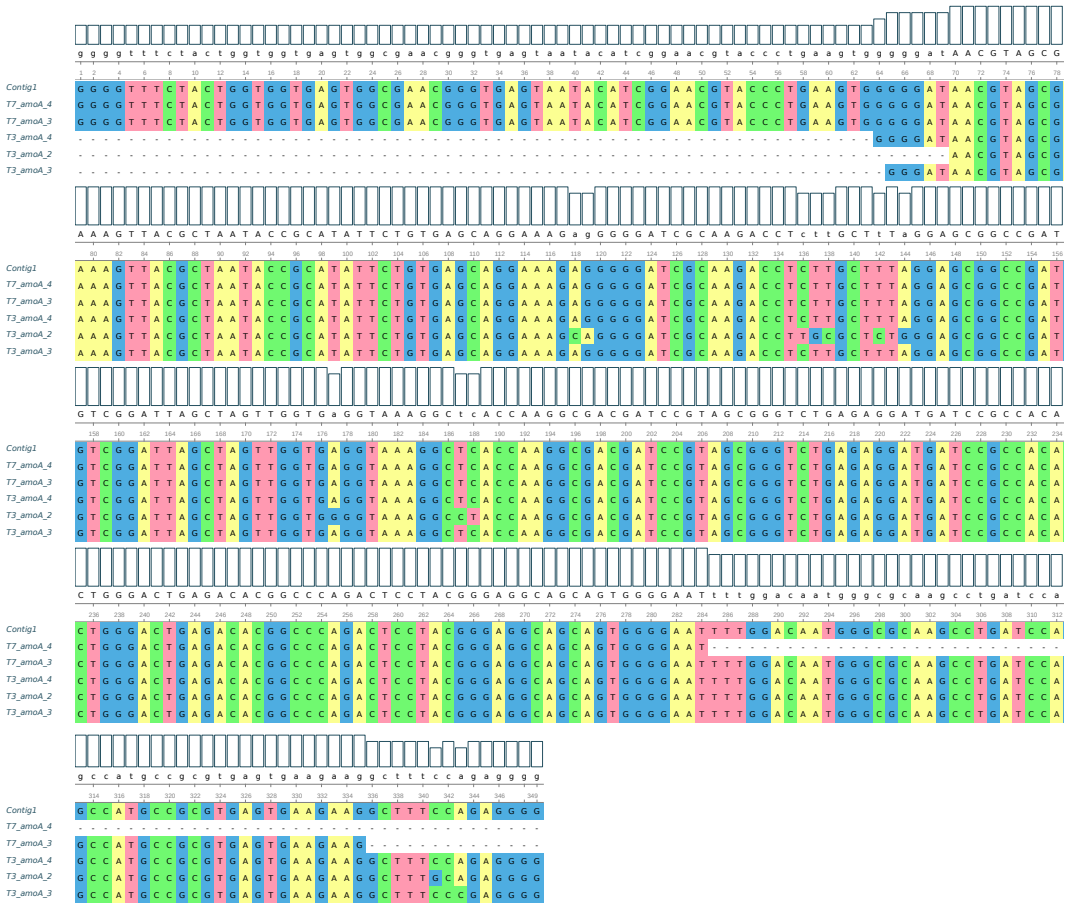
A.2.1 Sanger Sequencing

Table A.3: Settings used for CAP3 alignment

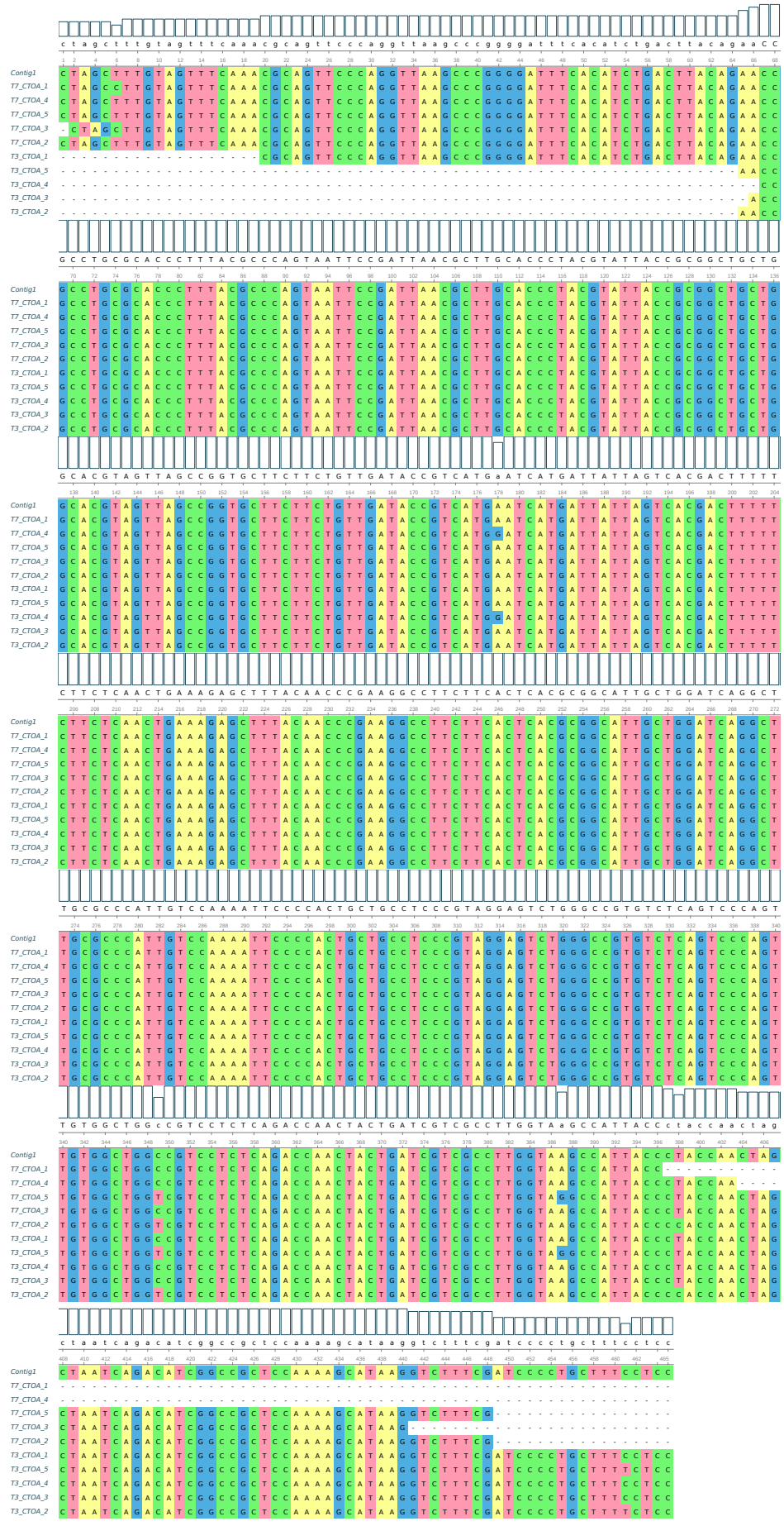
Parameter	Setting
Base quality cutoff for clipping (-c)	12
Clipping range (.y)	100
Base quality cutoff for differences (-b)	20
Max qscore sum at differences (-d)	200
Match score factor (-m)	2
Mismatch score factor (-n)	-5
Gap penalty factor (-g)	6
Overlap similarity score cutoff (-s)	900
Overlap length cutoff (-o)	40
Overlap percent identity cutoff (-p)	90
Max number of word matches (-t)	300
Band expansion size (-a)	20
Max gap length in any overlap (-f)	20
Assembly reverse reads (-r)	Yes



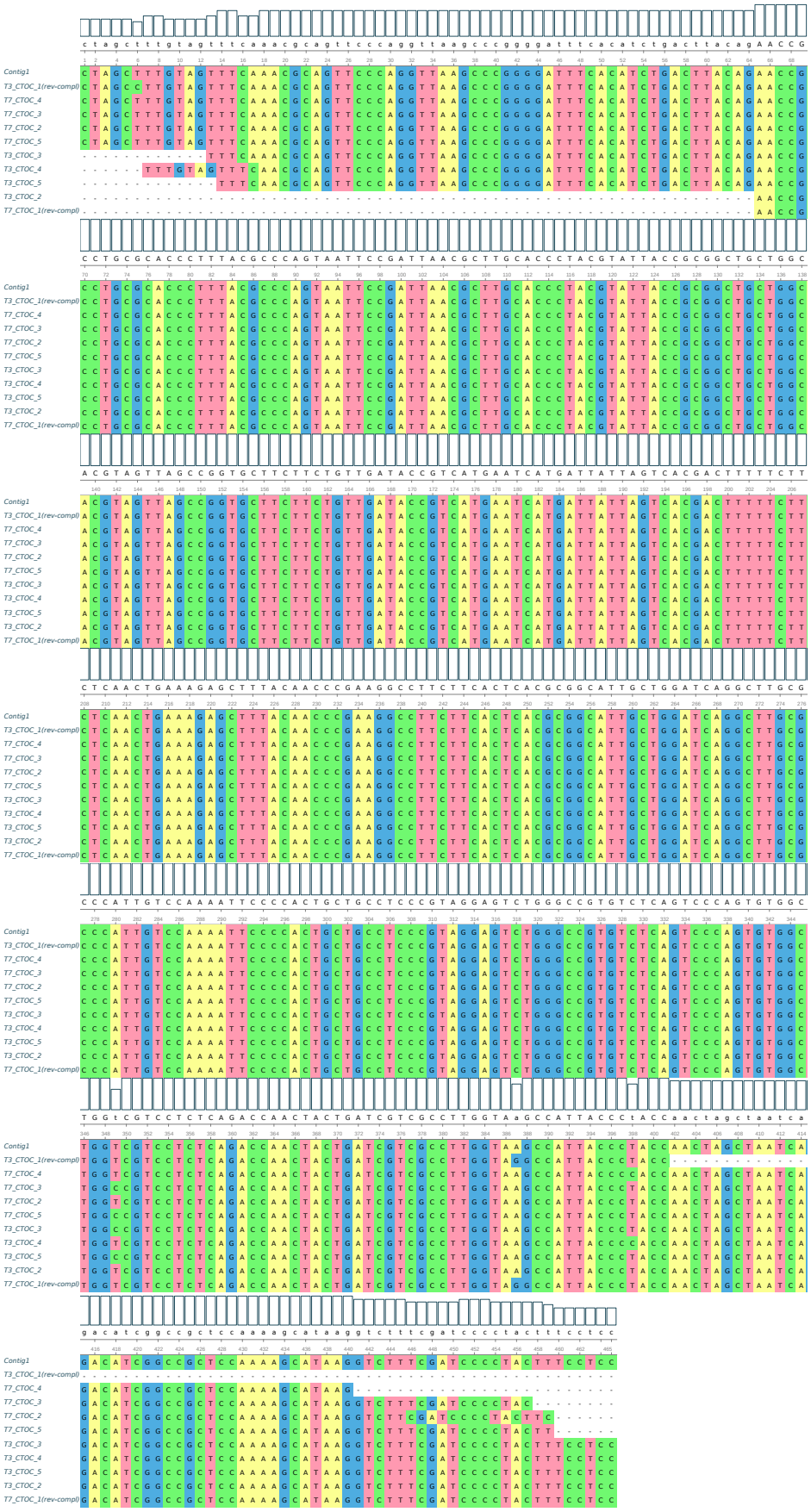
(a) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: amoA, Insert from colonies: 1 & 5



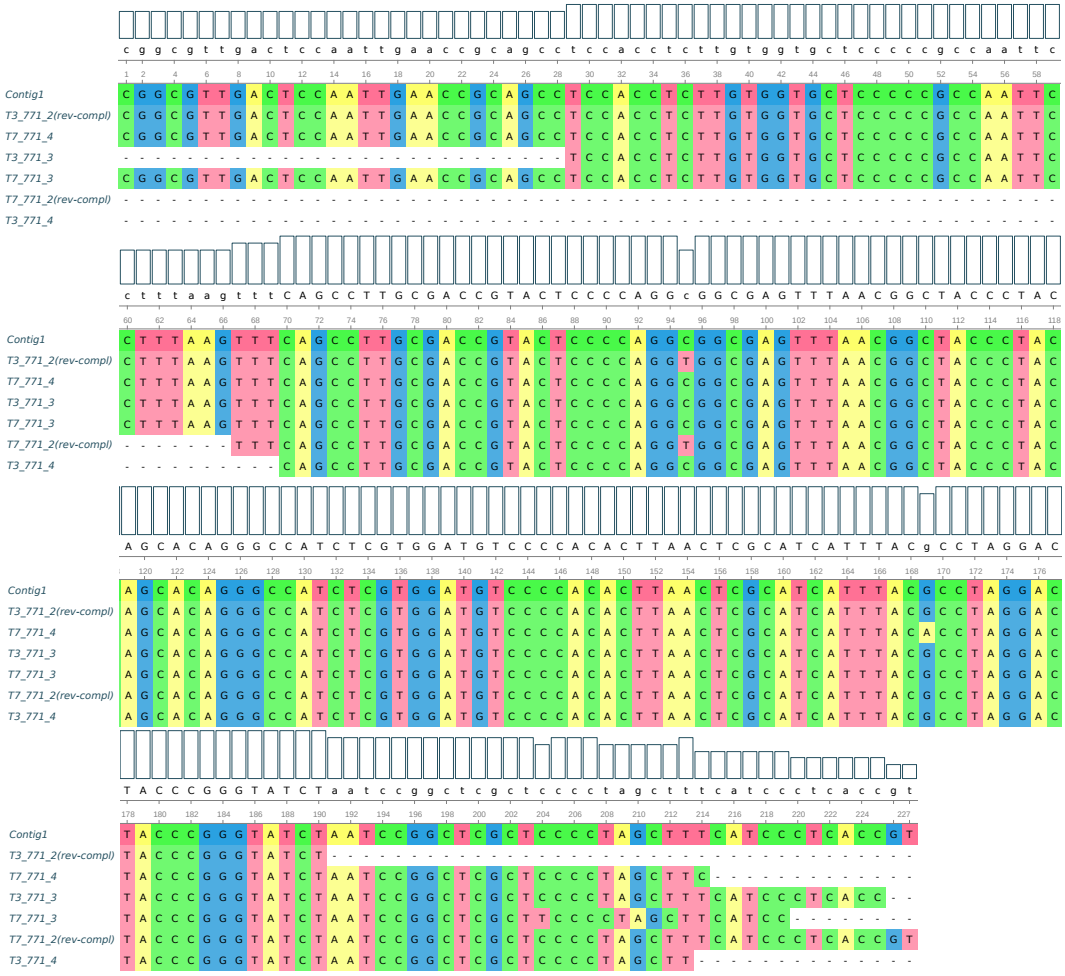
(b) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: amoA, Insert from colonies: 2, 3 & 4



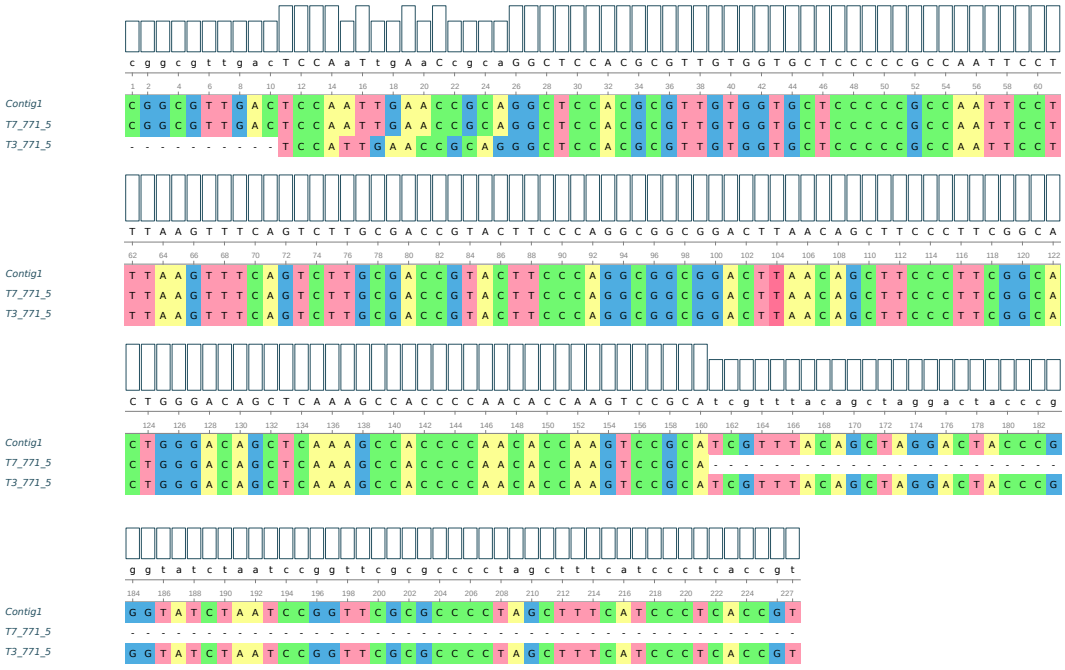
(c) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: 16s rRNA, Primers: CTO189fA/B & CTO189r



(d) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: 16s rRNA, Primers: CTO189fC & CTO189r



(e) Micro organisms: Ammonium Oxidising Archaea, Functional Gene: 16s rRNA, Insert from colonies: 2, 3 & 4



(f) Micro organisms: Ammonium Oxidising Archaea, Functional Gene: 16s rRNA, Insert from colonies: 5

Figure A.3: Alignment of trimmed sanger sequencing reads of inserts from colonies transformed with plasmid containing PCR product inserts for validation of those inserts.

Table A.4: Top 5 BLAST hits for consensus contigs blasted against the NCBI database

(a) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: amoA, Insert from colonies: 1 & 5

Accession	def	E-value	gaps	identities
MF117381	AmoA (amoA) gene	0	0/491 (0%)	489/491 (99.59%)
MH316795	ammonia monooxygenase subunit A (amoA) gene	0	0/491 (0%)	488/491 (99.39%)
KC735628	Uncultured Nitrosomonadaceae AmoA (amoA) gene	0	0/491 (0%)	488/491 (99.39%)
KC735627	Uncultured Nitrosomonadaceae AmoA (amoA) gene	0	0/491 (0%)	488/491 (99.39%)
MW368023	ammonia monooxidase (amoA) gene	0	0/491 (0%)	488/491 (99.39%)

(b) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: amoA, Insert from colonies: 2, 3 & 4

Accession	def	E-value	gaps	identities
GU537209	Uncultured bacterium partial 16S ribosomal RNA gene	2.72551e-164	0/335 (0%)	332/335 (99.1%)
LC489036	Uncultured bacterium partial 16S ribosomal RNA gene	2.72551e-164	0/337 (0%)	333/337 (98.81%)
LR643506	Uncultured bacterium partial 16S ribosomal RNA gene	2.72551e-164	0/337 (0%)	333/337 (98.81%)
LR645588	Uncultured bacterium partial 16S ribosomal RNA gene	2.72551e-164	0/337 (0%)	333/337 (98.81%)
LR646951	Uncultured bacterium partial 16S ribosomal RNA gene	2.72551e-164	0/337 (0%)	333/337 (98.81%)

(c) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: 16s rRNA, Primers: CTO189fA/B & CTO189r

Accession	def	E-value	gaps	identities
FM997795	Uncultured Nitrosomonas sp. partial 16S rRNA gene	0	0/465 (0%)	463/465 (99.57%)
KT724492	Uncultured bacterium 16S ribosomal RNA gene	0	0/465 (0%)	462/465 (99.35%)
KT724481	Uncultured bacterium 16S ribosomal RNA gene	0	0/465 (0%)	461/465 (99.14%)
HQ158724	Uncultured bacterium 16S ribosomal RNA gene	0	0/465 (0%)	461/465 (99.14%)
HQ158626	Uncultured bacterium 16S ribosomal RNA gene	0	0/465 (0%)	461/465 (99.14%)

(d) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: 16s rRNA, Primers: CTO189fC & CTO189r

Accession	def	E-value	gaps	identities
KT724492	Uncultured bacterium 16S ribosomal RNA gene	0	0/465 (0%)	464/465 (99.78%)
HQ158724	Uncultured bacterium 16S ribosomal RNA gene	0	0/465 (0%)	463/465 (99.57%)
HQ158626	Uncultured bacterium 16S ribosomal RNA gene	0	0/465 (0%)	463/465 (99.57%)
HQ221939	Uncultured ammonia-oxidizing bacterium 16S ribo...	0	0/465 (0%)	463/465 (99.57%)
HQ221934	Uncultured ammonia-oxidizing bacterium 16S ribo...	0	0/465 (0%)	463/465 (99.57%)

(e) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: 16s rRNA, Primers: CTO189fC & CTO189r

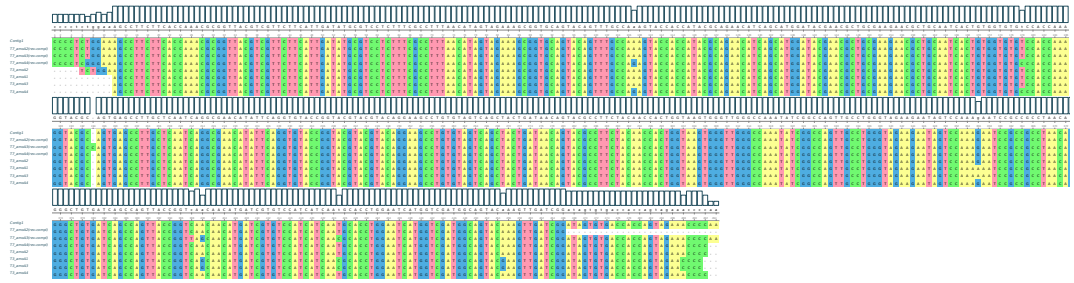
(f) Micro organisms: Ammonium Oxidising Archaea, Functional Gene: 16s rRNA, Insert from colonies: 2, 3 & 4

Accession	def	E-value	gaps	identities
KJ635823	Uncultured archaeon 16S ribosomal RNA gene	2.871090e-110	0/227 (0%)	227/227 (100%)
KF607836	Uncultured archaeon 16S ribosomal RNA gene	4.261070e-108	0/227 (0%)	226/227 (99.56%)
KF60.418	Uncultured archaeon 16S ribosomal RNA gene	6.323990e-106	0/227 (0%)	225/227 (99.12%)
KF60.406	Uncultured archaeon 16S ribosomal RNA gene	6.323990e-106	0/227 (0%)	225/227 (99.12%)
KF60.404	Uncultured archaeon 16S ribosomal RNA gene	6.323990e-106	0/227 (0%)	225/227 (99.12%)

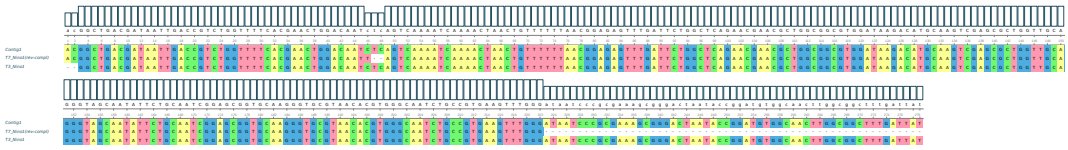
(g) Micro organisms: Ammonium Oxidising Archaea, Functional Gene: 16s rRNA, Insert from colonies: 5

Accession	def	E-value	gaps	identities
KX938459	Uncultured archaeon 16S ribosomal RNA gene	9.385630e-104	0/227 (0%)	224/227 (98.68%)
DQ186739	Uncultured archaeon 16S ribosomal RNA gene	9.385630e-104	0/227 (0%)	224/227 (98.68%)
CT573766	Uncultured bacterium partial 16S rRNA gene	9.385630e-104	0/227 (0%)	224/227 (98.68%)
KX940628	Uncultured archaeon 16S ribosomal RNA gene	1.392950e-101	0/227 (0%)	223/227 (98.24%)
KM221256	Uncultured archaeon 16S ribosomal RNA gene	1.392950e-101	0/227 (0%)	223/227 (98.24%)

because sequencing of *amoA* PCR products yielded uncertain results during the qPCR optimisation stage, the primer pair was validated again using the cDNA that showed amplification using qPCR. Additionally these new cDNA samples resulted in amplification with the Nino primer pair, the resulting product was also sequenced.



(a) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: amoA, Insert from colonies: 1, 2, 3 & 4



(b) Micro organisms: Comammox Bacteria, Functional Gene: amoA, Insert from colonies: 1

Figure A.4: Alignment of trimmed sanger sequencing reads of inserts from colonies transformed with plasmid containing PCR product inserts for validation of those inserts.

Table A.5: Top 5 BLAST hits for consensus contigs blasted against the NCBI database

(a) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: amoA, Insert from colonies: 1, 2, 3, & 4

Accession	def	E-value	gaps	identities
MF104089	AmoA (amoA) gene	0	0/491 (0%)	491/491 (100%)
MF104075	AmoA (amoA) gene	0	0/491 (0%)	491/491 (100%)
MF104065	AmoA (amoA) gene	0	0/491 (0%)	491/491 (100%)
MF104057	AmoA (amoA) gene	0	0/491 (0%)	491/491 (100%)
MF103932	AmoA (amoA) gene	0	0/491 (0%)	491/491 (100%)

(b) Micro organisms: Comammox Bacteria, Functional Gene: amoA, Insert from colonies: 1

Accession	def	E-value	gaps	identities
GU496349	Uncultured bacterium partial 16S ribosomal RNA gene	5.28157e-71	0/163 (0%)	161/163 (98.77%)
LN804079	Uncultured Verrucomicrobia partial 16S ribosomal RNA gene	7.83854e-69	0/163 (0%)	160/163 (98.16%)
EF688163	Uncultured bacterium partial 16S ribosomal RNA gene	3.12168e-61	5/180 (2.778%)	168/180 (93.33%)
KX239080	Uncultured bacterium partial 16S ribosomal RNA gene	5.64413e-58	2/177 (1.13%)	165/177 (93.22%)
KC758951	Uncultured bacterium partial 16S ribosomal RNA gene	1.08861e-79	0/183 (0%)	180/183 (98.36%)

A.3 Quality control of RNA used for analysis

The quality of RNA extracted using the liquid nitrogen sample handling protocol (2.1, 2.2) was assessed by Novogene RNA QC Group using Nanodrop, Agilent 2100 and Agarose Gel Electrophoresis. The Novogene Sample Quality Control Report shows high quality RNA.

- Qc Report
  - Project General Information
  - QC Methods
  - QC Results
  - Sample classification standard table

Novogene 诺禾致源

A.3.1 RNA QC Report

Report Date:2022-02-16 21:37:16

Project General Information
QC Methods
QC Results

- QC Result Summary
- Agarose Gel Electrophoresis Results
- The Integrity test result picture

Sample classification standard table

1.Project General Information

Project Name	NVUK2022012004-NO-INN-Johansen-Microbial-21-Shotgun-6G-Metatranscriptome-12G-WOBI	Contract No.	X204SC22012591-Z02
Client	Wenche Johansen	Institute/Company	Høgskolen i innlandet (INN)
Received By	CHANGYUAN7038	Received Date	2022-02-15
Tested By	RNA QC Group	Test Date	2022-02-15
Verified By	LIUHUIFANG4075	Verified Date	2022-02-16

2.QC Methods

Sample Type
DNA;√ RNA; smRNA; Tissue; Library ; Others
Assay Type
Sample Quantitation
Qubit Fluorometer Agarose Gel Electrophoresis Quantitation √ Nanodrop √ Agilent 2100 Agilent 5400
Sample Integrity
√ Agilent 2100 Agarose Gel Electrophoresis Agilent 5400
Sample Purity
Nanodrop √ Agarose Gel Electrophoresis √ Agilent 2100 Agilent 5400

3.QC Results

3.1 QC Results Summary

No.	Sample Name	Nucleic Acid ID	Concentration(ng/ul)	Volume(ul)	Total amount(ug)	Integrity value	Sample QC Results
1	R1A	FKRN220023336-1A	285.000	19.00	5.41500	8.10	Pass
2	R1B	FKRN220023337-1A	134.000	19.00	2.54600	7.70	Pass
3	R1C	FKRN220023338-1A	264.000	16.00	4.22400	7.40	Pass
4	R2A	FKRN220023339-1A	258.000	19.00	4.90200	7.50	Pass
5	R2B	FKRN220023340-1A	154.000	19.00	2.92600	7.60	Pass
6	R2C	FKRN220023341-1A	318.000	19.00	6.04200	7.60	Pass
7	R3A	FKRN220023342-1A	279.000	19.00	5.30100	7.30	Pass
8	R3B	FKRN220023343-1A	240.000	19.00	4.56000	7.20	Pass
9	R3C	FKRN220023344-1A	190.000	19.00	3.61000	7.40	Pass
10	R4A	FKRN220023345-1A	248.000	19.00	4.71200	7.30	Pass
11	R4B	FKRN220023346-1A	272.000	19.00	5.16800	7.80	Pass
12	R4C	FKRN220023347-1A	261.000	19.00	4.95900	7.60	Pass
13	R6A	FKRN220023348-1A	441.000	19.00	8.37900	7.70	Pass
14	R6B	FKRN220023349-1A	198.000	19.00	3.76200	7.40	Pass
15	R6C	FKRN220023350-1A	258.000	19.00	4.90200	7.50	Pass
16	R8A	FKRN220023351-1A	324.000	19.00	6.15600	7.70	Pass
17	R8B	FKRN220023352-1A	122.000	19.00	2.31800	8.10	Pass
18	R8C	FKRN220023353-1A	135.000	18.00	2.43000	7.80	Pass
19	R10A	FKRN220023354-1A	204.000	19.00	3.87600	7.80	Pass
20	R10B	FKRN220023355-1A	206.000	19.00	3.91400	7.50	Pass
21	R10C	FKRN220023356-1A	347.000	19.00	6.59300	7.50	Pass

3.2 Electrophoresis Results

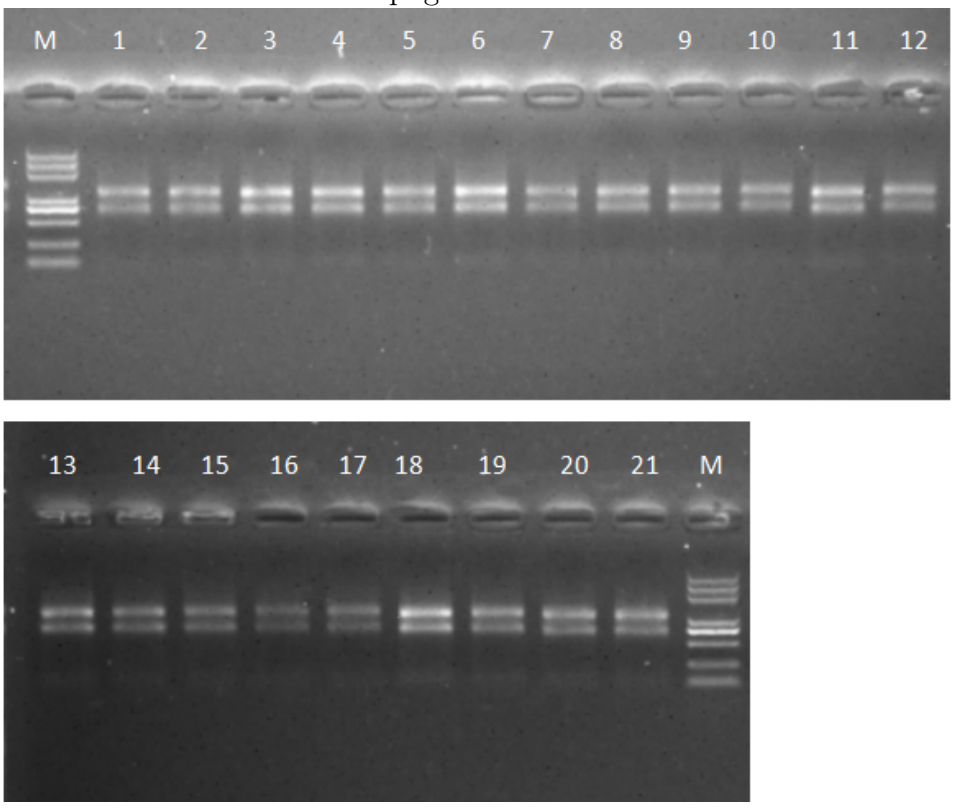
3.2.1 Electrophoresis Condition

GelConc. 1% Voltage 180v Run Time 16min

Note Agarose gel electrophoresis will not be conducted on Eukaryotic transcriptome library samples

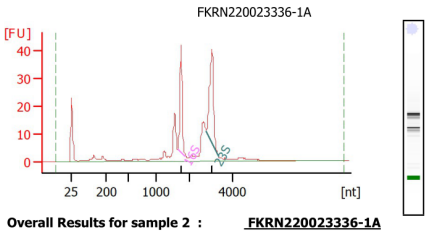
3.2.2 Electrophoresis Results

- 5A402BBB-8A9C-4AAA-ACBC-12166F19FBF9.png 5A402BBB-8A9C-4AAA-ACBC-12166F19FBF9.bb



3.3 The Integrity test result picture

FKRN220023336-1A



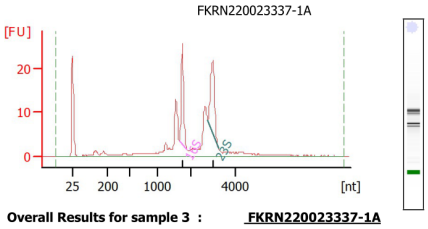
Overall Results for sample 2 : FKRN220023336-1A

RNA Area:	254.5
RNA Concentration:	95 ng/µl
rRNA Ratio [23s / 16s]:	1.3
RNA Integrity Number (RIN):	8.1 (B.02.10, Anomaly Threshold(s) manually adapted)
Result Flagging Color:	
Result Flagging Label:	RIN: 8.10

Fragment table for sample 2 : FKRN220023336-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,648	1,880	43.1	16.9
23S	2,761	3,331	56.2	22.1

FKRN220023337-1A



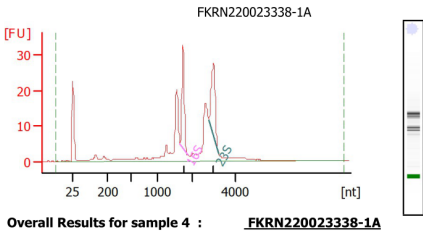
Overall Results for sample 3 : FKRN220023337-1A

RNA Area:	178.6
RNA Concentration:	67 ng/µl
rRNA Ratio [23s / 16s]:	1.1
RNA Integrity Number (RIN):	7.7 (B.02.10, Anomaly Threshold(s) manually adapted)
Result Flagging Color:	
Result Flagging Label:	RIN: 7.70

Fragment table for sample 3 : FKRN220023337-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,644	1,874	25.9	14.5
23S	2,759	3,245	28.5	16.0

FKRN220023338-1A



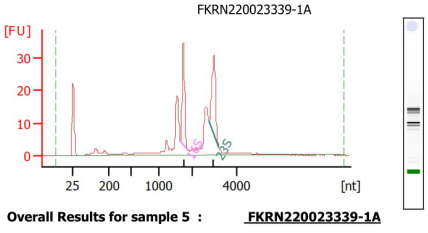
Overall Results for sample 4 : FKRN220023338-1A

RNA Area:	234.9
RNA Concentration:	88 ng/µl
rRNA Ratio [23s / 16s]:	1.1
RNA Integrity Number (RIN):	7.4 (B.02.10, Anomaly Threshold(s) manu
Result Flagging Color:	
Result Flagging Label:	RIN: 7.40

Fragment table for sample 4 : FKRN220023338-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,645	1,867	32.4	13.8
23S	2,772	3,230	34.4	14.6

FKRN220023339-1A



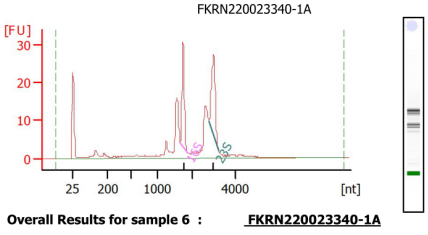
Overall Results for sample 5 : FKRN220023339-1A

RNA Area:	230.2
RNA Concentration:	86 ng/µl
rRNA Ratio [23s / 16s]:	1.1
RNA Integrity Number (RIN):	7.5 (B.02.10, Anomaly Threshold(s) manually adapted)
Result Flagging Color:	
Result Flagging Label:	RIN: 7.50

Fragment table for sample 5 : FKRN220023339-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,638	1,868	34.7	15.1
23S	2,758	3,201	37.7	16.4

FKRN220023340-1A



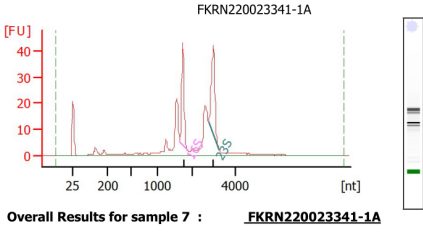
Overall Results for sample 6 : FKRN220023340-1A

RNA Area:	205.7
RNA Concentration:	77 ng/µl
rRNA Ratio [23s / 16s]:	1.1
RNA Integrity Number (RIN):	7.6 (B.02.10, Anomaly Threshold(s) manually adapted)
Result Flagging Color:	
Result Flagging Label:	RIN: 7.60

Fragment table for sample 6 : FKRN220023340-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,654	1,875	30.2	14.7
23S	2,786	3,230	33.7	16.4

FKRN220023341-1A



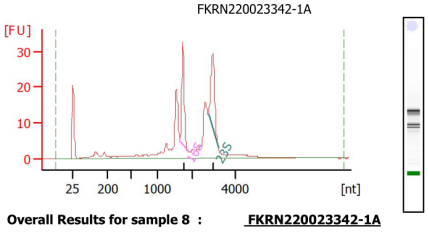
Overall Results for sample 7 : FKRN220023341-1A

RNA Area:	282.7
RNA Concentration:	106 ng/µl
rRNA Ratio [23s / 16s]:	1.2
RNA Integrity Number (RIN):	7.6 (B.02.10, Anomaly Threshold(s) manu
Result Flagging Color:	
Result Flagging Label:	RIN: 7.60

Fragment table for sample 7 : FKRN220023341-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,637	1,867	43.5	15.4
23S	2,759	3,243	50.3	17.8

FKRN220023342-1A



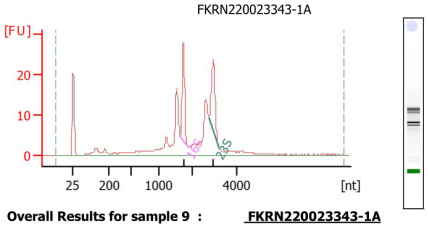
Overall Results for sample 8 : FKRN220023342-1A

RNA Area:	247.7
RNA Concentration:	93 ng/µl
rRNA Ratio [23s / 16s]:	1.1
RNA Integrity Number (RIN):	7.3 (B.02.10, Anomaly Threshold(s) manually adapted)
Result Flagging Color:	
Result Flagging Label:	RIN: 7.30

Fragment table for sample 8 : FKRN220023342-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,645	1,867	32.4	13.1
23S	2,759	3,216	34.6	14.0

FKRN220023343-1A



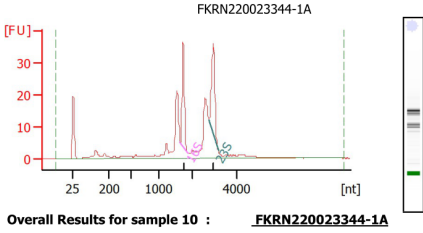
Overall Results for sample 9 : FKRN220023343-1A

RNA Area:	214.6
RNA Concentration:	80 ng/µl
rRNA Ratio [23s / 16s]:	1.0
RNA Integrity Number (RIN):	7.2 (B.02.10, Anomaly Threshold(s) manually adapted)
Result Flagging Color:	
Result Flagging Label:	RIN: 7.20

Fragment table for sample 9 : FKRN220023343-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,630	1,859	28.0	13.1
23S	2,745	3,201	26.7	12.4

FKRN220023344-1A

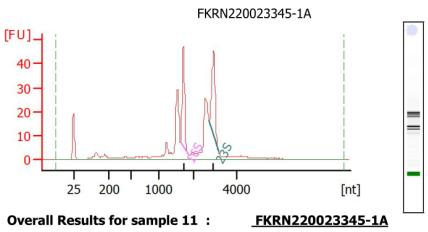


Overall Results for sample 10 : FKRN220023344-1A

RNA Area:	252.2
RNA Concentration:	95 ng/µl
rRNA Ratio [23s / 16s]:	1.2
RNA Integrity Number (RIN):	7.4 (B.02.10, Anomaly Threshold(s) manu
Result Flagging Color:	
Result Flagging Label:	RIN: 7.40

Fragment table for sample 10 : FKRN220023344-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,638	1,851	34.4	13.7
23S	2,745	3,174	40.4	16.0

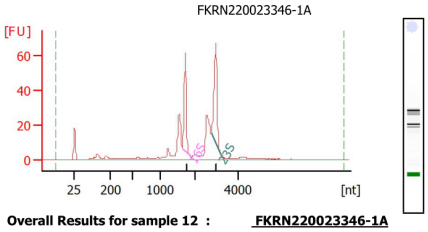


Overall Results for sample 11 : FKRN220023345-1A

RNA Area:	331.7
RNA Concentration:	124 ng/µl
rRNA Ratio [23s / 16s]:	1.1
RNA Integrity Number (RIN):	7.3 (B.02.10, Anomaly Threshold(s) manually adapted)
Result Flagging Color:	
Result Flagging Label:	RIN: 7.30

Fragment table for sample 11 : FKRN220023345-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,631	1,843	46.7	14.1
23S	2,744	3,146	49.5	14.9

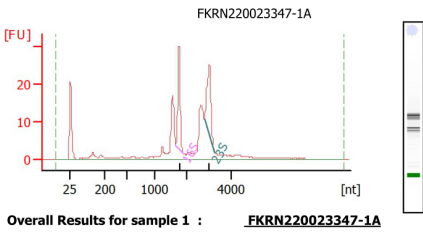


Overall Results for sample 12 : FKRN220023346-1A

RNA Area:	361.3
RNA Concentration:	136 ng/µl
rRNA Ratio [23s / 16s]:	1.2
RNA Integrity Number (RIN):	7.8 (B.02.10, Anomaly Threshold(s) manually adapted)
Result Flagging Color:	
Result Flagging Label:	RIN: 7.80

Fragment table for sample 12 : FKRN220023346-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,636	1,863	58.6	16.2
23S	2,743	3,208	71.8	19.9

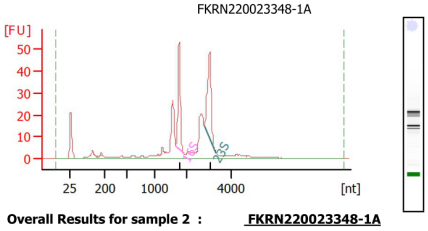


Overall Results for sample 1 : FKRN220023347-1A

RNA Area:	201.3
RNA Concentration:	87 ng/µl
rRNA Ratio [23s / 16s]:	1.0
RNA Integrity Number (RIN):	7.6 (B.02.10, Anomaly Threshold(s) manu
Result Flagging Color:	
Result Flagging Label:	RIN: 7.60

Fragment table for sample 1 : FKRN220023347-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,660	1,910	30.0	14.9
23S	2,786	3,280	31.1	15.4

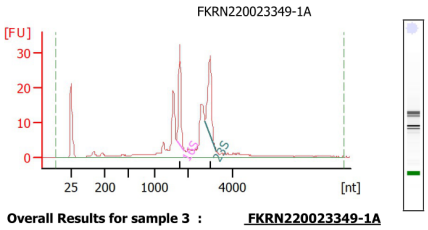


Overall Results for sample 2 : FKRN220023348-1A

RNA Area:	341.9
RNA Concentration:	147 ng/µl
rRNA Ratio [23s / 16s]:	1.1
RNA Integrity Number (RIN):	7.7 (B.02.10, Anomaly Threshold(s) manually adapted)
Result Flagging Color:	
Result Flagging Label:	RIN: 7.70

Fragment table for sample 2 : FKRN220023348-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,660	1,901	53.9	15.8
23S	2,800	3,319	61.4	18.0

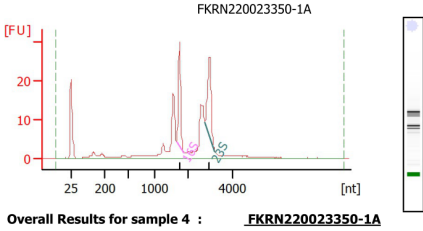


Overall Results for sample 3 : FKRN220023349-1A

RNA Area:	229.3
RNA Concentration:	99 ng/µl
rRNA Ratio [23s / 16s]:	1.1
RNA Integrity Number (RIN):	7.4 (B.02.10, Anomaly Threshold(s) manually adapted)
Result Flagging Color:	
Result Flagging Label:	RIN: 7.40

Fragment table for sample 3 : FKRN220023349-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,646	1,877	32.5	14.2
23S	2,772	3,263	34.2	14.9

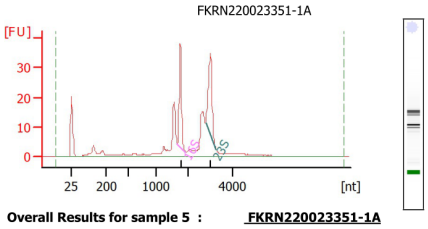


Overall Results for sample 4 : FKRN220023350-1A

RNA Area:	199.5
RNA Concentration:	86 ng/µl
rRNA Ratio [23s / 16s]:	1.0
RNA Integrity Number (RIN):	7.5 (B.02.10, Anomaly Threshold(s) manu
Result Flagging Color:	
Result Flagging Label:	RIN: 7.50

Fragment table for sample 4 : FKRN220023350-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,646	1,877	29.7	14.9
23S	2,772	3,223	29.9	15.0

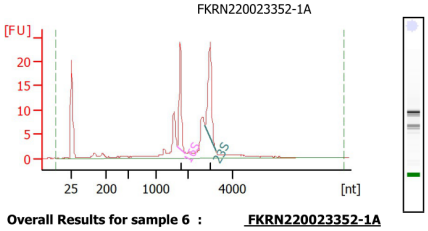


Overall Results for sample 5 : FKRN220023351-1A

RNA Area:	250.9
RNA Concentration:	108 ng/µl
rRNA Ratio [23s / 16s]:	1.1
RNA Integrity Number (RIN):	7.7 (B.02.10, Anomaly Threshold(s) manually adapted)
Result Flagging Color:	
Result Flagging Label:	RIN: 7.70

Fragment table for sample 5 : FKRN220023351-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,656	1,913	37.9	15.1
23S	2,771	3,261	43.1	17.2

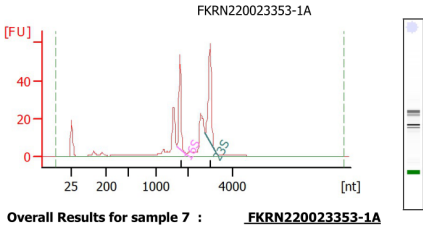


Overall Results for sample 6 : FKRN220023352-1A

RNA Area:	141.7
RNA Concentration:	61 ng/µl
rRNA Ratio [23s / 16s]:	1.2
RNA Integrity Number (RIN):	8.1 (B.02.10, Anomaly Threshold(s) manually adapted)
Result Flagging Color:	
Result Flagging Label:	RIN: 8.10

Fragment table for sample 6 : FKRN220023352-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,656	1,913	23.5	16.6
23S	2,758	3,221	28.9	20.4

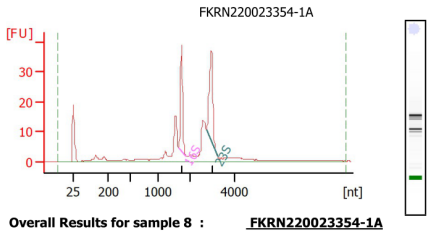


Overall Results for sample 7 : FKRN220023353-1A

RNA Area:	313.1
RNA Concentration:	135 ng/µl
rRNA Ratio [23s / 16s]:	1.3
RNA Integrity Number (RIN):	7.8 (B.02.10, Anomaly Threshold(s) manu
Result Flagging Color:	
Result Flagging Label:	RIN: 7.80

Fragment table for sample 7 : FKRN220023353-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,638	1,887	50.7	16.2
23S	2,758	3,235	65.5	20.9

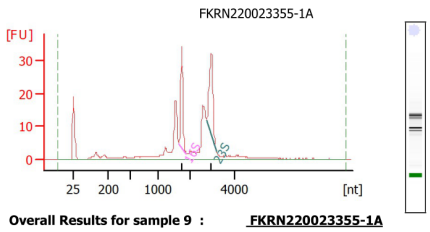


Overall Results for sample 8 : FKRN220023354-1A

RNA Area:	237.8
RNA Concentration:	102 ng/μl
rRNA Ratio [23s / 16s]:	1.2
RNA Integrity Number (RIN):	7.8 (B.02.10, Anomaly Threshold(s) manually adapted)
Result Flagging Color:	
Result Flagging Label:	RIN: 7.80

Fragment table for sample 8 : FKRN220023354-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,638	1,860	36.6	15.4
23S	2,718	3,221	45.2	19.0

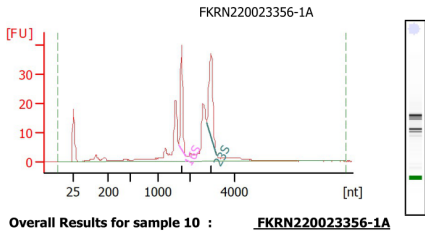


Overall Results for sample 9 : FKRN220023355-1A

RNA Area:	239.0
RNA Concentration:	103 ng/μl
rRNA Ratio [23s / 16s]:	1.2
RNA Integrity Number (RIN):	7.5 (B.02.10, Anomaly Threshold(s) manually adapted)
Result Flagging Color:	
Result Flagging Label:	RIN: 7.50

Fragment table for sample 9 : FKRN220023355-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,647	1,869	31.9	13.3
23S	2,731	3,169	38.7	16.2



Overall Results for sample 10 : FKRN220023356-1A

RNA Area:	268.7
RNA Concentration:	116 ng/μl
rRNA Ratio [23s / 16s]:	1.2
RNA Integrity Number (RIN):	7.5 (B.02.10, Anomaly Threshold(s) manu
Result Flagging Color:	
Result Flagging Label:	RIN: 7.50

Fragment table for sample 10 : FKRN220023356-1A

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,638	1,860	37.7	14.0
23S	2,731	3,182	45.6	17.0

4.Sample classification standard table

Note

The test conclusion is a comprehensive evaluation based on Novogene's sample quality requirements for library preparation and sequencing

The test result based on the "RNA Samples QC Criteria"explains whether the testing sample meets the requirement of library construction.

a) Pass: The sample is qualified and the library preparation can be prepared regularly;

b) Hold: The sample does not totally meet the requirements of library construction and sequencing. Novogene can try to construct library but the sequencing quality is not guaranteed;

c) Fail: It is highly recommended for clients to resend the sample or the library could be constructed at high risk and the sequencing quality is not guaranteed (not recommended).

A.4 Data Handling

The Applied Biosystems 7500 Fast Real-Time PCR System (USA) has no method for generating publication quality figures of qPCR data so R was used to generate these figures. In order to import the data exported from the 7500 Fast Dx software (Applied Biosystems, USA) into the R software, relevant data had to be filtered. A Bash script was developed to handle data exported form the 7500 Fast DX software and integrate R scripts to export publication quality figures in a streamlined way. In addition the pandas python package was used to automatically convert csv files into LaTeXtables.

```
#!/bin/bash
# Author: Sverre Branders

# Get file name from user
# read -p "Input file name: " FILE
if [ ! -f $FILES[1] ] 2> /dev/null
then
  while [ ! -f $FILES[1] ] 2> /dev/null; do

printf "Please select file:\n"
select f in ./*; do test -n "$f" && break; echo ">>> Invalid Selection"; done
#  !(DisCurve-byPrimer.R|DisCurve.R|generate-latex.py|layout-to-table.py|
    process_date.sh|StdCurve.R)
FILE=${f%.*}
FILE=${FILE#./}

# Remove -d or -l form file name
if [[ $FILE == *-d ]]
then
  FILE=${FILE%-d}
elif [[ $FILE == *-l ]]
then
  FILE=${FILE%-l}
fi
FILES=$(ls | grep $FILE)

if [ ! -f $FILES[1] ] 2> /dev/null
then
  echo "Invalid! Please use another file"
fi
done
fi

echo "Found the following files:
$FILES"

# Make Directories is not allready there
if [ ! -d processed/raw_files/ ]; then
  mkdir -p processed/raw_files/;
```

```
fi
# CP backup of files to the folder
cp $FILES processed/raw_files/
# CleanUp files
# Prompt user
read -r -p "Do you wish to automatically clean up the chosen files? [Y/n] "
    response
case "$response" in
    [nN][oO]|[nN])
        echo "Continuing without cleaning files"
        ;;
    *)
# removing junk from $FILE.csv
  if [ -f "$FILE.csv" ]
  then
    CLEAN=1
    cat "$FILE.csv" | head -28 > "$FILE-info.txt"
    cat "$FILE.csv" | sed 1,28d > "$FILE-cat.csv"
    echo "Cleaning $FILE.csv"
  elif [ ! -f "$FILE.csv" ]
  then
    echo "\033[31m $FILE.csv not found! \e[0m"
  fi

# removing junk from $FILE-l.txt and converting to csv
   if [ -f "$FILE-l.txt" ]
   then
     cat "$FILE-l.txt" | sed 1,6d > "temp-l-cat.csv"
     echo "Cleaning $FILE-l.txt"
     cat "temp-l-cat.csv" | sed 's/\t/,/g' > "$FILE-l-cat.csv"
     echo "Converting to csv"
     rm "temp-l-cat.csv"
     echo "Writing to $FILE-l-cat.csv"
   elif [ ! -f "$FILE-l.txt" ]
   then
     echo "$FILE-l.txt not found!"
   fi

# splitting $FILE-d.csv into different files
  if [ -f "$FILE-d.csv" ]
  then
    csplit "$FILE-d.csv" --elide-empty-files --prefix "$FILE-d-split" --suffix
   -format="%d.csv" '/^Well*/' '{*}' &> /dev/null
    echo "Splitting $FILE-d.csv"

    cat "$FILE-d-split0.csv" | sed 1d > "$FILE-d-split-temp.csv"
    cat "$FILE-d-split2.csv" | sed 1d > "$FILE-d-split-deriv.csv"
    rm $FILE-d-split[0-2].csv
    echo "Cleaning up split files"

  elif [ ! -f "$FILE-d.csv" ]
  then
    echo "$FILE-d.csv not found!"
  fi
```

```
91
92    # Layout to table
93  read -r -p "Do you wish to generate a table of the 96 well plate layout? [Y/n]
      " response
94  case "$response" in
95      [nN][oO]|[nN])
96          echo "Continuing without generating table"
97          ;;
98      *)
99    if [ -f "$FILE-l.txt" ]
100    then
101      Rscript layout-to-table.R "$FILE" &> /dev/null
102      echo "Writing to $FILE-plate-layout.csv"
103       # Generate Latex
104       read -r -p "Do you wish to automatically generate a Latex table file? [Y
      /n]" response
105       case "$response" in
106           [nN][oO]|[nN])
107               echo "Generating Latex Table"
108               ;;
109           *)
110         if [ ! -d processed/Latex/ ]; then
111           mkdir -p processed/Latex/;
112         fi
113
114         python generate-latex.py "$FILE" &> /dev/null
115         mv "$FILE-plate-layout.tex" processed/Latex/
116         ;;
117       esac
118
119    elif [ ! -f "$FILE-l.txt" ]
120    then
121      echo "$FILE-l.txt not found!"
122    fi
123   ;;
124 esac
125
126 echo "Files cleaned"
127         ;;
128 esac
129
130 # Run StdCurve script
131 read -r -p "Generate standard curve? [y/N] " response
132 case "$response" in
133     [yY][eE][sS]|[yY])
134   if [ -f "$FILE.csv" ]
135    then
136     if [ ! -d processed/StdCurves/ ]; then
137       mkdir -p processed/StdCurves/;
138     fi
139
140     Rscript StdCurve.R "$FILE" &> /dev/null
141      cat "$FILE-l.txt" | sed 1,6d > "temp-l-cat.csv"
142      echo "Cleaning $FILE-l.txt"
```

```
143     cat "temp-l-cat.csv" | sed 's/\t/,/g' > "$FILE-l-cat.csv"
144     echo "Converting to csv"
145     rm "temp-l-cat.csv"
146 
147   elif [ ! -f "$FILE.csv" ]
148   then
149     echo "$FILE.csv not found!"
150   fi
151         ;;
152    *)
153         echo "Continuing without generating standard curve"
154   ;;
155 esac
156 
157 # Run DisCurve script
158 read -r -p "Generate dissociation Curve? [Y/n] " response
159 case "$response" in
160    [nN][oO]|[nN])
161   echo "Continuing without generating dissociation curve"
162   ;;
163    *)
164   if [ -f "$FILE-d.csv" ]
165    then
166     if [ ! -d processed/DisCurves/ ]; then
167       mkdir -p processed/DisCurves/;
168     fi
169 
170   echo "Make different files per primer pair or compile all data to one figure
       ? "
171   choise=("Make seperate figure per primer pair" "Compile all data to one
       figure")
172   select c in "${choise[@]}"; do
173       case $c in
174         "Make seperate figure per primer pair")
175           Rscript DisCurve-byPrimer.R "$FILE" &> /dev/null
176           break
177           ;;
178         "Compile all data to one figure")
179           Rscript DisCurve.R "$FILE" &> /dev/null
180           break
181           ;;
182       esac
183     done
184       echo "Generated dissociaion curve"
185 
186   elif [ ! -f "$FILE-d.csv" ]
187    then
188     echo "$FILE-d.csv not found!"
189   fi
190   ;;
191 esac
192 
193 # Final Cleanup
194 if [ ! -d processed/cleaned_files/ ] && [ $CLEAN==1 ]; then
```

```
  mkdir -p processed/cleaned_files/
  rm $FILES
  ls | grep $FILES | while read f; do mv "$f" processed/cleaned_files/; done;

elif [ $CLEAN==1 ]; then
  rm $FILES
  ls | grep $FILES | while read f; do mv "$f" processed/cleaned_files/; done;
fi

echo "Finished!"
```

Listing A.1: Bash script for data handling

```
# Sverre Branders
#
# Format "Sample Name" in following format:
# primerpairname-sampledilution/concentration-templatename.replicatenumber
# export dissociation curve raw data to csv and append file name with -d
# export Sample Layout to txt and append file name with -l

# Get user input from Bash Script ----
flnm <- commandArgs(trailingOnly = TRUE)
# Debugging
# flnm <- "test"

# Create file paths and import as dataframes
d1path <- paste(flnm,"-d-split-temp.csv",sep="") # Complete file path
d2path <- paste(flnm,"-d-split-deriv.csv",sep="") # Complete file path
dnmpath <- paste(flnm,"-l-cat.csv",sep="") # Complete file path

d1 <- read.csv(d1path, header = FALSE)  # Import file into dataframe
d2 <- read.csv(d2path, header = FALSE)  # Import file into dataframe
dnm <- read.csv(dnmpath)  # Import file into dataframe

# Renaming columns
colnames(d1)[1] <- "Well"
colnames(d2)[1] <- "Well"
colnames(dnm)[2] <- "snm"

# Removing unnecessary columns
d1[c(2, 3)] <- NULL
d2[c(2, 3)] <- NULL
dnm[c(3:5)] <- NULL

# Add sample based on well number and get columns
library("tidyr")
library("dplyr")
d1 <- merge(dnm, d1, by = "Well")
d1[1] <- NULL
cn <- colnames(d1)
# cn <- sub("[V]*","T",cn)
colnames(d1) <- cn
colnames(d1)[1] <- "snm"
d1cn <- colnames(d1)[-1]
```

```
d2cn <- colnames(d2)[-1]

d2 <- merge(dnm, d2, by = "Well")
d2[1] <- NULL

# PIVOT FIRST THEN COMBINE THE TABLES
d1piv <- pivot_longer(d1,
                      cols = d1cn,
                      names_to = "v",
                      values_to = "temp")
d2piv <- pivot_longer(d2,
                      cols = d2cn,
                      names_to = "v",
                      values_to = "deriv")
d12longer <- merge(d1piv, d2piv, by = "snm")
d12longer_f <- d12longer[d12longer$v.x==d12longer$v.y,] %>% # Remove unrelated
    values
  mutate(snm2 = snm) %>% # Duplicate snm col
  separate(snm, c("primer", "conc", "templ"), sep = "-") %>% # Separate primer
    concentration and tempelate
  separate(templ, c("templ", "repl"), sep = "\\Q.\\E", extra = "merge") #
   Separate replicate number

# ggplot
library("ggplot2")
library("ggpmisc")

d12longer_f <- d12longer_f %>%
  group_by(snm2)
lapply(sort(unique(d12longer_f$primer)), function(i) {
  ggplot(d12longer_f[d12longer_f$primer==i,], aes(x = temp, y = deriv, colour
    = repl)) +
  scale_x_continuous(breaks = seq(0, 100, by = 5)) +
  geom_line(size = 0.25) +
  ggtitle(i) +
  labs(x = "Temperature (\u00B0C)", y = "Derivative")+
  theme_minimal() +
  theme(plot.title = element_text(size=15)) +
  facet_wrap(conc~templ)
  # Save plot
  svgname <- paste("processed/DisCurves/",flnm,"-DisCurve-",i,".svg",sep="")
  ggsave(file = as.character(svgname), width = 15, height = 11.5, dpi = 300)
})
```

Listing A.2: R script for generating dissociation curves

```
# Sverre Branders
#
# Format "Sample Name" in following format:
# primerpairname-sampledilution/concentration-templatename.replicatenumber
# export dissociation curve raw data to csv and append file name with -d
# export Sample Layout to txt and append file name with -l

# Get user input from Bash Script ----
```

```
flnm <- commandArgs(trailingOnly = TRUE)
# Debugging
# flnm <- "test"

# Create file paths and import as dataframes
d1path <- paste(flnm,"-d-split-temp.csv",sep="") # Complete file path
d2path <- paste(flnm,"-d-split-deriv.csv",sep="") # Complete file path
dnmpath <- paste(flnm,"-l-cat.csv",sep="") # Complete file path

d1 <- read.csv(d1path, header = FALSE)  # Import file into dataframe
d2 <- read.csv(d2path, header = FALSE)  # Import file into dataframe
dnm <- read.csv(dnmpath)  # Import file into dataframe

# Renaming columns
colnames(d1)[1] <- "Well"
colnames(d2)[1] <- "Well"
colnames(dnm)[2] <- "snm"

# Removing unnecessary columns
d1[c(2, 3)] <- NULL
d2[c(2, 3)] <- NULL
dnm[c(3:5)] <- NULL

# Add sample based on well number and get columns
library("tidyr")
library("dplyr")
d1 <- merge(dnm, d1, by = "Well")
d1[1] <- NULL
cn <- colnames(d1)
# cn <- sub("[V]*","T",cn)
colnames(d1) <- cn
colnames(d1)[1] <- "snm"
d1cn <- colnames(d1)[-1]
d2cn <- colnames(d2)[-1]

d2 <- merge(dnm, d2, by = "Well")
d2[1] <- NULL

# PIVOT FIRST THEN COMBINE THE TABLES
d1piv <- pivot_longer(d1,
                      cols = d1cn,
                      names_to = "v",
                      values_to = "temp")
d2piv <- pivot_longer(d2,
                      cols = d2cn,
                      names_to = "v",
                      values_to = "deriv")
d12longer <- merge(d1piv, d2piv, by = "snm")
d12longer_f <- d12longer[d12longer$v.x==d12longer$v.y,] %>% # Remove unrelated
    values
  mutate(snm2 = snm) %>% # Duplicate snm col
  separate(snm, c("primer", "conc", "templ"), sep = "-") %>% # Separate primer
    concentration and tempelate
  separate(templ, c("templ", "repl"), sep = "\\Q.\\E", extra = "merge") #
```

```
    Separate replicate number

# ggplot
library("ggplot2")
library("ggpmisc")

d12longer_f %>%
  group_by(snm2) %>%
  ggplot(aes(x = temp, y = deriv, colour = repl)) +
  scale_x_continuous(breaks = seq(0, 100, by = 5)) +
  geom_line(size = 0.25) +
  labs(x = "Temperature (\u00B0C)", y = "Derivative")+
  theme_minimal() +
  facet_wrap(primer~templ~conc)


# Save plot
svgname <- paste("processed/DisCurves/",flnm,"-DisCurve.svg",sep="")
ggsave(file = as.character(svgname), width = 15, height = 11.5, dpi = 300)
```

Listing A.3: R script for generating dissociation curves

```
# Sverre Branders
#
# Format "Sample Name" in following format:
# primerpairname-sampledilution/concentration-templatename.replicatenumber
#

# Get user input from Bash Script ----
flnm <- commandArgs(trailingOnly = TRUE)

# Create file path and import as dataframe
flpath <- paste(flnm,"-cat.csv",sep="") # Complete file path
print(flpath)
data_in <- read.csv(flpath)  # Import file into dataframe

# DEBUGGING
# data_in <- read.csv("testfile.csv")
# flnm <- "test"

# Rename columns
names(data_in) <- c("well", "Sname", "det", "task", "Ct", "stdev", "qty", "
    qtym", "qtysd", "filtered","Tm", "usr", "usr2", "usr3")

# Split "Sample Name" into different columns
library("tidyr")
dt <- data_in %>%
  separate(Sname, c("primer", "conc", "templ"), sep = "-")
dt <- dt %>%
  separate(templ, c("templ", "repl"), sep = "\\Q.\\E", extra = "merge") #
    Separate replicate number


# Group by primer, conc & templ and calculate average and sdv
```

```r
library("dplyr")
grps <- c("primer", "conc", "templ")
dtgr <- dt %>%
  group_by(across(all_of(grps))) %>%
    dplyr::summarise(
        avg_Ct = mean(as.numeric(Ct)),
        sdv = sd(as.numeric(Ct))
    )

# Calculate log(conc)
dtgr$log_conc <- log(as.numeric(dtgr$conc), base = exp(1))

# Calculate ymax and ymin from stdev
dtgr$ymax <- dtgr$avg_Ct + dtgr$sdv
dtgr$ymin <- dtgr$avg_Ct - dtgr$sdv

# Ommit NTC
dtgr <- dtgr %>% filter(conc != "")

# SVG name
svgname <- paste("processed/StdCurves/",flnm,"-StdCurve.svg",sep="")
print(svgname)

# Plot avg_Ct to log_conc
library("ggplot2")
library("ggpmisc")

ggplot(dtgr, aes(log_conc, avg_Ct))+
  geom_point(size = 2)+
  geom_errorbar(aes(ymin = ymax, ymax = ymin), size = 0.15)+
  theme_minimal(base_size = 7)+
  labs(x = "Log(Dilution)", y = "Ct")+
  geom_smooth(method = lm, se = F, color = "black", size = 0.25,)+
  stat_poly_eq(
               aes(label = paste(..eq.label.., ..rr.label.., sep = "~~~")),
               rr.digits = 3,
               parse = TRUE,
         size = 2)+
  facet_wrap(~primer*templ)

ggsave(file = as.character(svgname), dpi = 300)
```

Listing A.4: R script for generating standard curves

```r
# INPUT
flnm <- commandArgs(trailingOnly = TRUE)
dnmpath <- paste(flnm,"-1-cat.csv",sep="") # Complete file path

plate <- data.frame(c(1:96))
colnames(plate)[1] <- "Well"

l <- read.csv(dnmpath)
l[c(3:5)] <- NULL

```

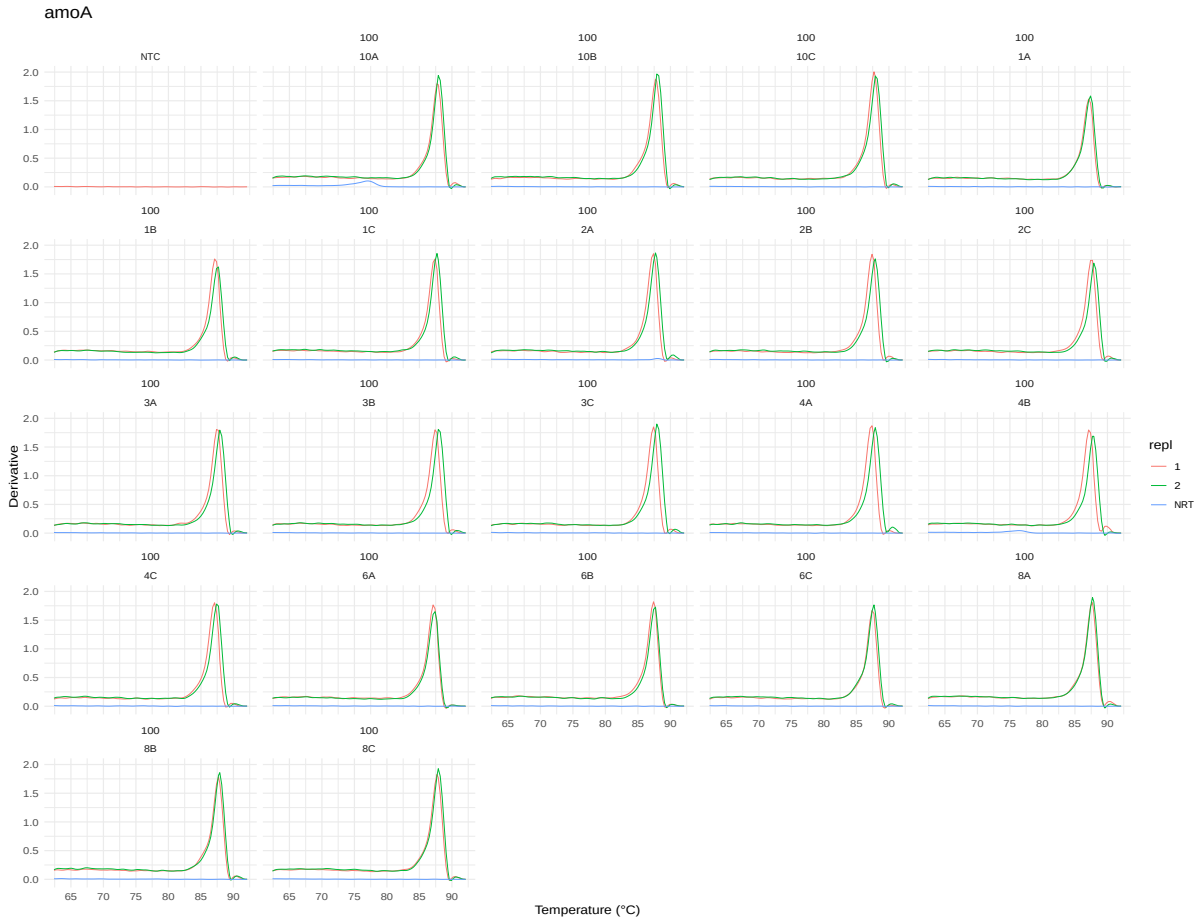
```r
# Add empty rows
i <-  96 - nrow(l)
i2 <- i
while (i > 0) {
  l[nrow(l) + 1,] = c(0 ,NA)
  i <- i-1
}

# Merge
plate_l <- merge(l, plate, by = "Well", all.y = TRUE)
# Transform to 8 x 12
row1 <- plate_l[c(1:12),2]
row2 <- plate_l[c(13:24),2]
row3 <- plate_l[c(25:36),2]
row4 <- plate_l[c(37:48),2]
row5 <- plate_l[c(49:60),2]
row6 <- plate_l[c(61:72),2]
row7 <- plate_l[c(73:84),2]
row8 <- plate_l[c(85:96),2]

plate_l <- rbind(row1, row2, row3, row4, row5, row6, row7, row8)

# OUTPUT
ofl <- paste(flnm,"-plate-layout.csv",sep="") # Complete file path
write.table(plate_l, ofl, sep=",", row.names = FALSE, col.names = FALSE)
```

Listing A.5: R script to make a table of plate layout

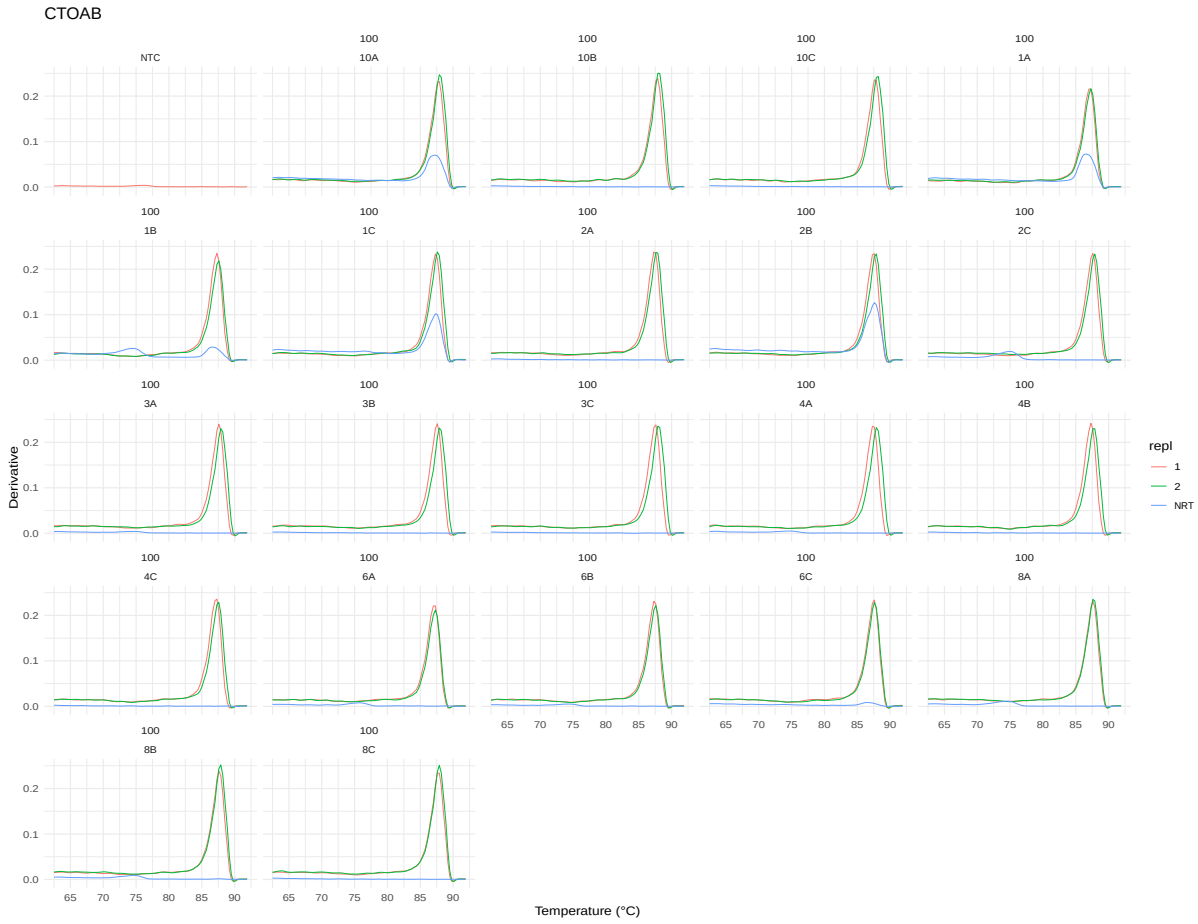
```python
import pandas as pd
import sys
from os.path import exists as file_exists

flnm = sys.argv[1]

# Plate layout to table
table = str(flnm + "-plate-layout")
if file_exists(table + ".csv"):
    df = pd.read_csv(table + ".csv")
    df.to_latex(table + ".tex", index=False, caption="<++>", label=table)
    print("Table of layout saved to " + table + ".tex")
else:
    print(table + ".csv not found!")
```

Listing A.6: python script to convert csv files into LaTeXtables

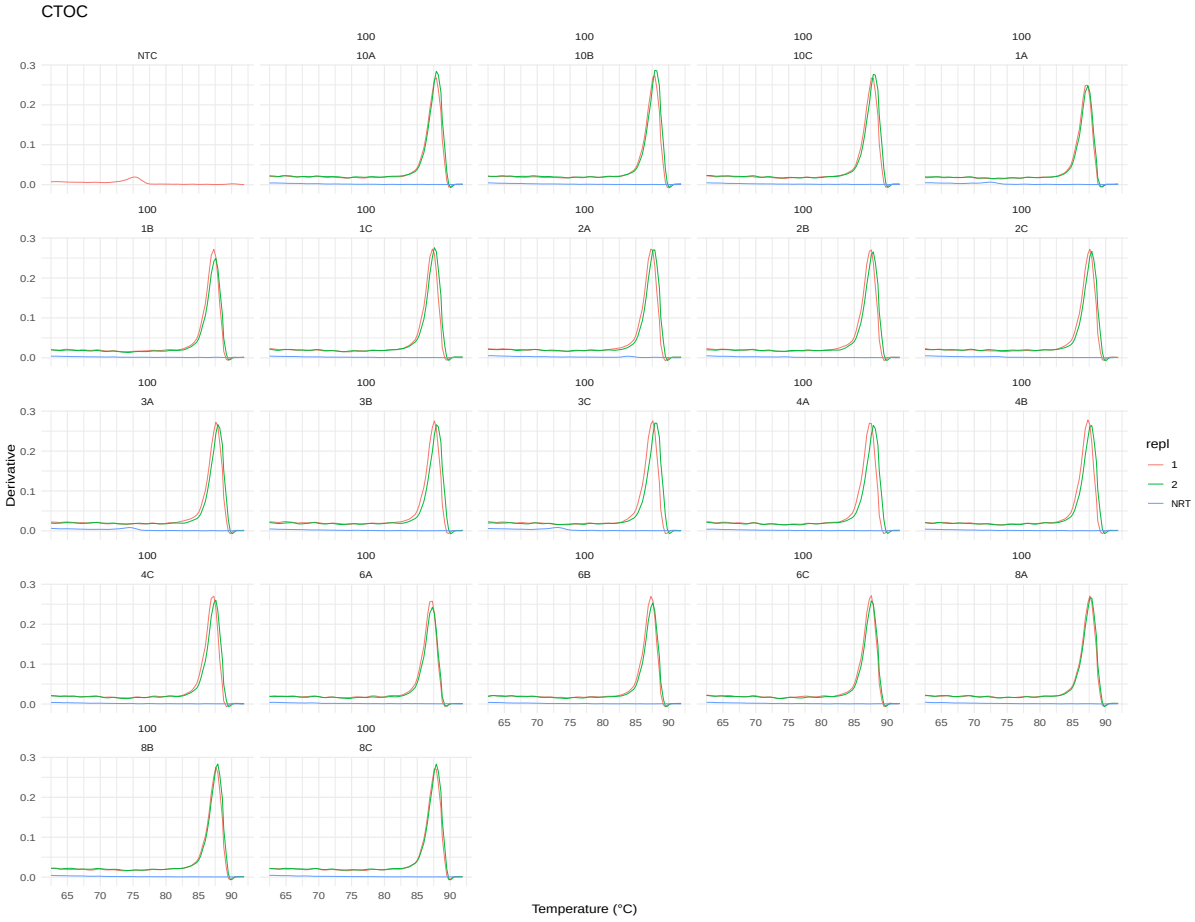
A.5 Post dissociation analysis of final qPCR reactions



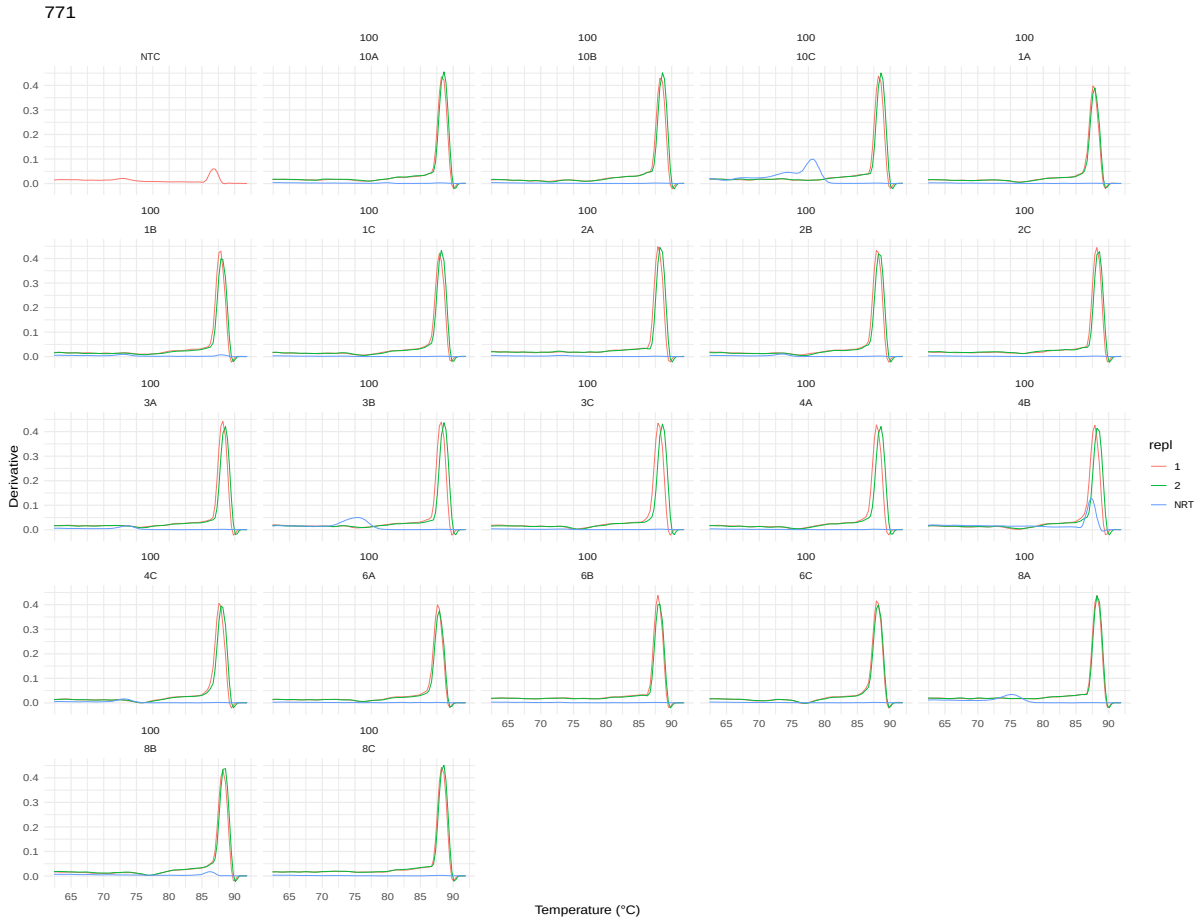
(a) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: amoA, Primers: amoA-1F & amoA-2R



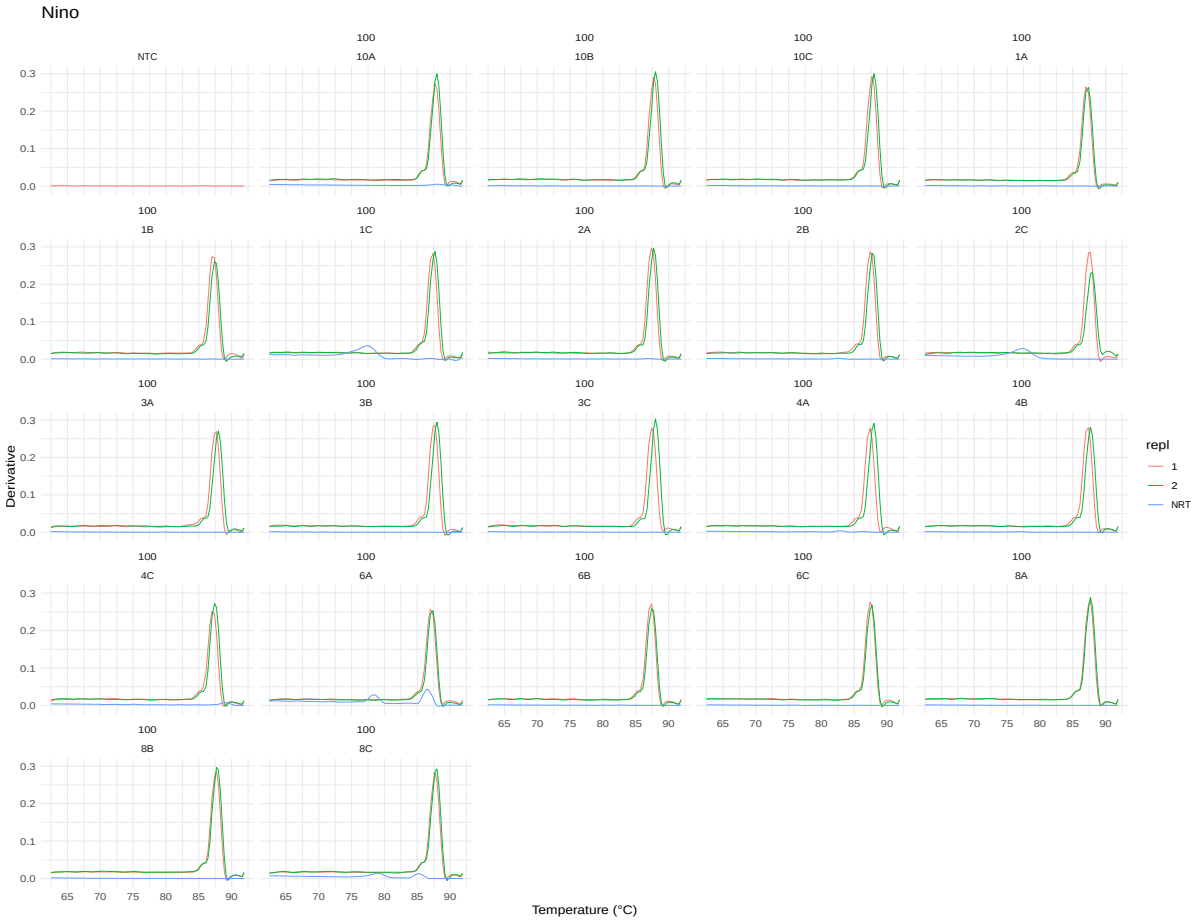
(b) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: 16s rRNA, Primers: CTO189fA/B & CTO189r



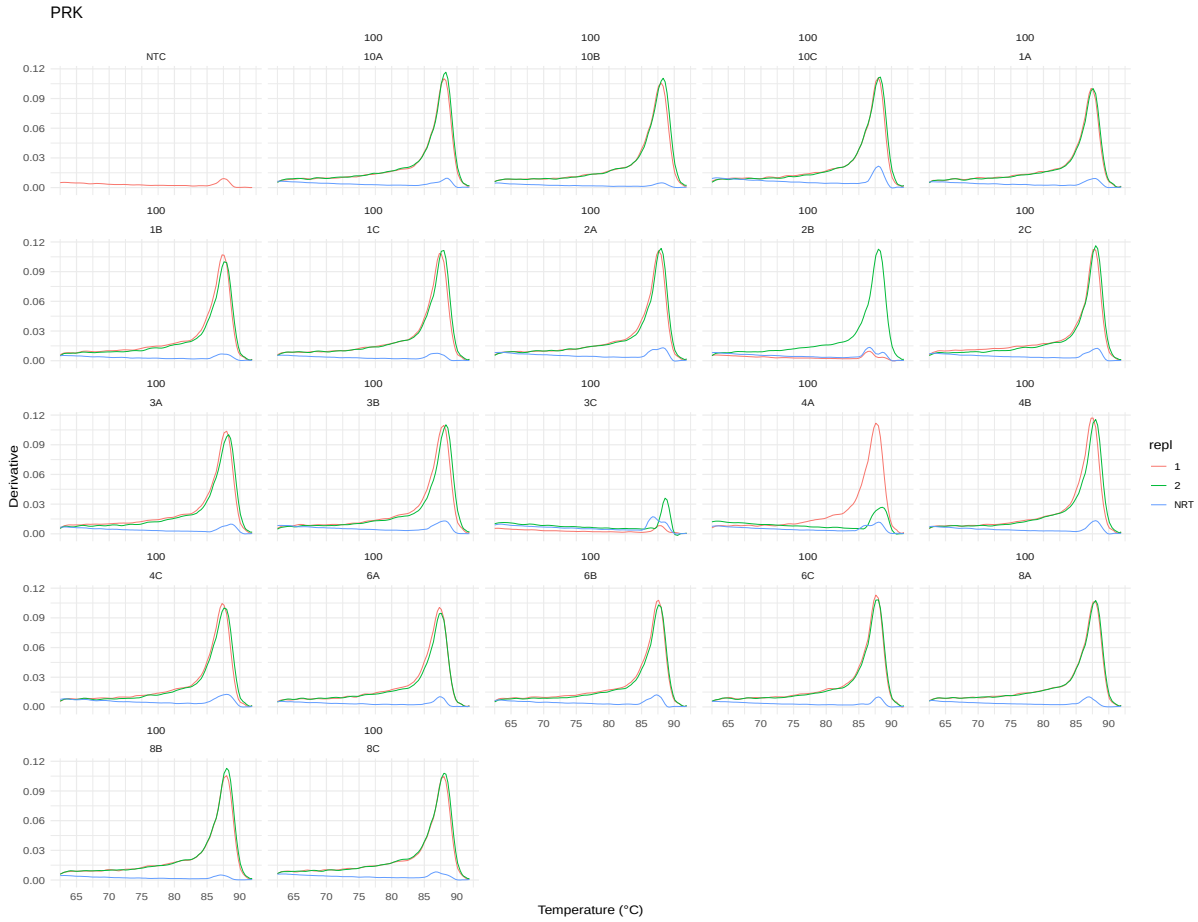
(c) Micro organisms: Ammonium Oxidising Bacteria, Functional Gene: 16s rRNA, Primers: CTO189fC & CTO189r



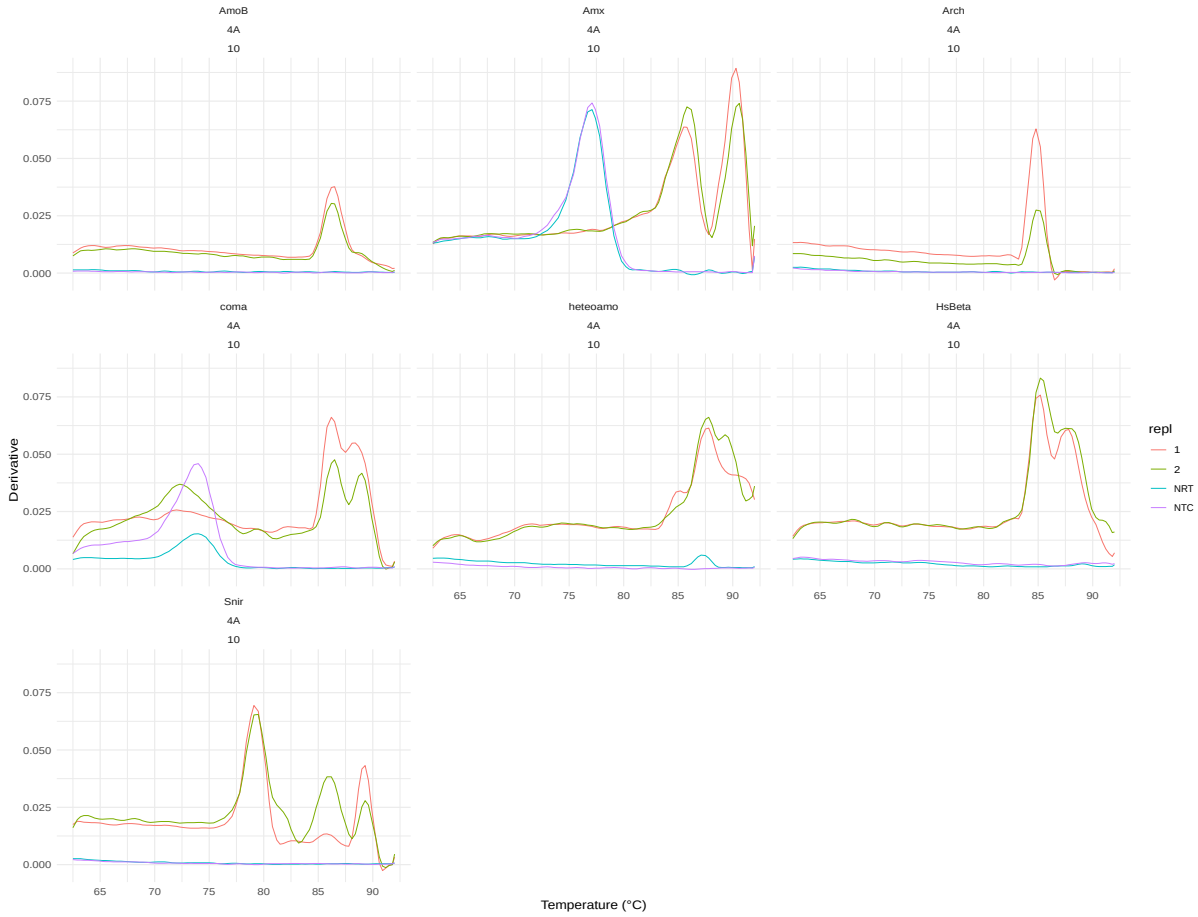
(d) Micro organisms: Ammonium Oxidising Archaea, Functional Gene: 16s rRNA, Primers: 771F & 957R



(e) Micro organisms: Complete Ammonium Oxidising Bacteria, Functional Gene: amoA, Primers: Nino_amoA_19F & Nino_amoA_252R



(f) Micro organisms: Prokaryotes, Functional Gene: 16S rRNA, Primers: PRK341F & PRK806R



(g) Primers resulting in no substantial amplification

Figure A.5: Dissociation curves of qPCR reactions preformed with different primer pairs on cDNA made from RNA extracted from biofilm from waste water treatment reactor.

A.6 Statistics

A.6.1 Table of data used for Statistical analysis

Table A.8: Cycle threshold and chemical data used for statistical analysis

Sample Name	Zone	NH4	amoA	CTO AB	CTO C	771	Nino
1A.1	1	73.0	30.7345	18.3538	16.9818	25.2194	28.5547
1B.1	1	73.0	30.1709	18.6667	17.0994	25.0306	28.6874
1C.1	1	73.0	30.1254	18.1982	16.7013	24.6792	28.4866
2A.1	2	74.0	29.1388	17.8530	16.4141	25.1883	28.6085
2B.1	2	74.0	29.7081	18.3338	16.9444	24.6928	28.5099
2C.1	2	74.0	30.0553	18.4548	17.0331	24.7515	28.1103
3A.1	3	74.5	29.0970	17.7427	16.1776	24.7176	28.4903
3B.1	3	74.5	29.2326	17.4608	16.2190	24.1340	28.1328
3C.1	3	74.5	29.0596	17.2695	16.1705	24.0843	28.1740
4A.1	4	74.0	28.7632	17.1673	16.1028	24.3505	28.1431
4B.1	4	74.0	29.7565	17.5157	16.3473	24.5407	28.2757
4C.1	4	74.0	29.4379	17.6090	16.4782	24.2279	29.1373
6A.1	6	70.0	30.0206	17.4637	16.1633	25.1663	28.2908
6B.1	6	70.0	29.7523	17.7716	16.4393	25.2537	28.4493
6C.1	6	70.0	30.3945	17.5943	16.3394	24.6227	28.1182
8A.1	8	64.0	29.3127	17.3919	16.3135	25.0224	28.0886
8B.1	8	64.0	30.0525	17.7333	16.4858	24.6643	28.3162
8C.1	8	64.0	29.2746	17.7134	16.3389	25.0169	28.2801
10A.1	10	57.5	29.7166	17.6820	16.4331	24.4676	28.2079
10B.1	10	57.5	28.7208	17.1441	15.9664	24.8636	28.0594
10C.1	10	57.5	29.1437	17.0247	15.9527	24.8557	27.7635
1A.2	1	73.0	30.4225	17.7624	16.3918	24.9288	28.4623
1B.2	1	73.0	30.4814	18.4878	16.9737	25.0722	28.5901
1C.2	1	73.0	30.0564	17.9945	16.6447	25.0082	28.4686
2A.2	2	74.0	29.3133	17.4302	16.2762	25.2160	28.5717
2B.2	2	74.0	30.1051	17.6359	16.3071	24.6319	28.7331
2C.2	2	74.0	30.7261	17.3333	16.2196	26.0528	29.8858
3A.2	3	74.5	29.3407	16.5757	15.5538	25.3027	29.7185
3B.2	3	74.5	29.6288	17.3419	16.1507	24.1929	28.1564
3C.2	3	74.5	28.9299	17.1149	16.0667	24.0442	28.0585
4A.2	4	74.0	29.0830	17.0448	16.0017	24.2741	28.0341
4B.2	4	74.0	30.3283	17.2799	16.3264	24.5863	28.0767
4C.2	4	74.0	29.7350	17.4664	16.3370	24.2319	28.0467
6A.2	6	70.0	29.9087	17.4758	16.2403	25.3436	28.3027
6B.2	6	70.0	30.0926	17.7380	16.4381	25.2553	28.4573
6C.2	6	70.0	30.0210	17.6017	16.3140	24.6984	27.9535
8A.2	8	64.0	29.0775	17.3673	16.2101	24.9859	27.8873
8B.2	8	64.0	30.2021	17.5697	16.3249	24.8192	28.1569
8C.2	8	64.0	29.3986	17.3602	16.1468	24.9822	28.1440
10A.2	10	57.5	29.8720	17.5796	16.3173	24.6370	28.0635
10B.2	10	57.5	29.2964	17.0055	15.9160	25.1420	28.3884
10C.2	10	57.5	29.0721	16.9787	16.0342	25.0350	28.1598

A.6.2 Script used for statistical analysis

Normality of Ct data was analysed by plotting a histogram and quantile-quantile (QQ) plot Statistical significance of variation in obtained Ct value across reactor zones was determined using analysis of variance (anova) grouped by primer pair and mean and standard deviation of those Ct values were plotted by reactor zone for each primer pair. Linear regression was used to evaluate correlation between Ct value and ammonium concentration in each reactor zone and results were plotted.

```
# Sverre Branders

library('tidyr')
library('dplyr')
library('ggplot2')
library('ggrepel')
library('ggtext')
library('ggpmisc')
library('stargazer')

# Data Import
dt <- read.csv('../qPCR_Stat_2.csv')

colnames(dt) <- c('sn','Zone','NH4','amoA','CTO AB','CTO C','771','Nino')
dtt <- dt %>%
  pivot_longer(cols = c('amoA','CTO AB','CTO C','771','Nino'), names_to = '
    PrimerPair', values_to = 'Ct')

# Check Normality
# Histogram
fig <- dtt %>%
  ggplot(aes(x = Ct))+
  geom_histogram(bins = 7, color = 'black', fill = NA)+
  theme_classic()+
  facet_wrap(~PrimerPair, scales = "free")+
  theme(strip.background = element_blank())

ggsave('Histogram.svg', plot = fig, width = 10, height = 7)

# QQ-plot
fig <- dtt %>%
  ggplot(aes(sample = Ct))+
  geom_qq(size = 0.5)+
  geom_qq_line()+
  ylab(label = "Sample Quantiles")+
  xlab(label = "Normal Quantiles")+
  theme_classic()+
  facet_wrap(~PrimerPair, scales = "free")+
  theme(strip.background = element_blank())

ggsave('Normal_Dist.svg', plot = fig, width = 10, height = 7)

#===============
# Stat + Plot
```

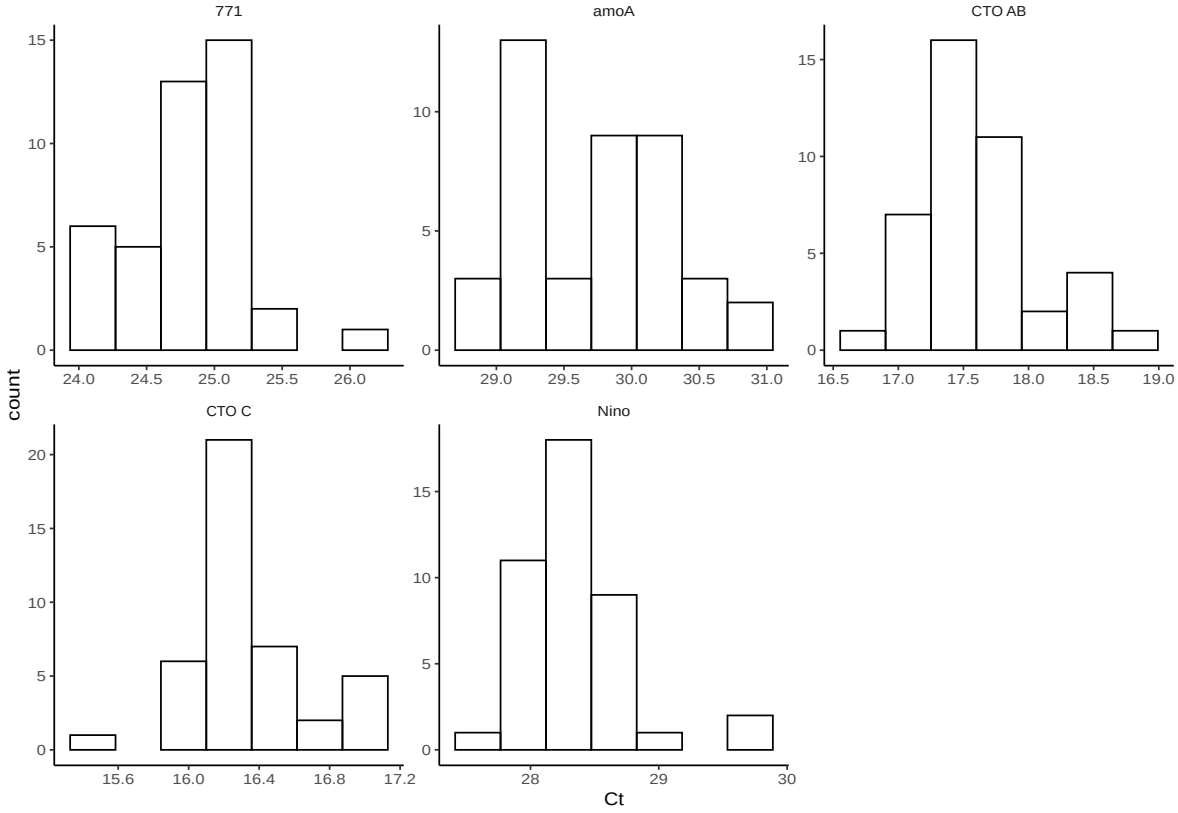
```
#================

# generate stat data
dtti <- dtt %>% group_by(PrimerPair) %>%
  summarise(Ct, Zone, PrimerPair, NH4) %>%
  mutate(id = paste(PrimerPair, Zone, sep = '_')) %>%
  group_by(id) %>%
  summarise(Zone,
            PrimerPair,
            Ct,
            NH4,
            id,
            count = n(),
            mean = mean(Ct, na.rm = TRUE),
            sd = sd(Ct, na.rm = TRUE))
# remove duplicate data
dtti <- dtti[!duplicated(dtti),]

# generate figures + ANOVA + LM
astat <- data.frame(ANOVA = c('Zone',' Residuals'))
# Loop through primer pairs
for (i in dtti$PrimerPair){
fign <- paste(i,'Ct.svg',sep = '_')

fdt <- dtti %>%
    filter(PrimerPair == i)
#PLOT
fig <- fdt %>%
    ggplot(aes(x = Zone, y = mean, ymax = mean+sd, ymin = mean-sd))+
    geom_line(size = 0.2)+
    geom_point(size = 0.5)+
    geom_errorbar(size = 0.1)+
    ylab(label = "Mean Ct")+
    labs(title = i)+
    coord_cartesian(ylim = c(33, 16))+
    scale_x_continuous(breaks = c(1,2,3,4,6,8,10))+
    scale_y_reverse(breaks = c(15:35))+
    theme_classic()

    ggsave(fign, plot = fig)

#ANOVA
astat[,i] <- print(summary(aov(mean ~ Zone, data = fdt)))

#Regression AGAINST NH4
fign <- paste(i,'correlation.svg',sep = '_')
fig <- fdt %>%
    ggplot(aes(y = NH4, x = Ct))+
    geom_point()+
    geom_smooth(method = 'lm', se = F, color = 'black', size = 0.5)+
    stat_poly_eq(
      aes(label = paste(..eq.label.., ..rr.label.., sep = "~~~")),
      rr.digits = 3,
      parse = TRUE,
```

```
      label.y = 0.5,
      size = 6)+
    labs(title = i, y = 'NH<sub>4</sub>')+
    theme_classic()+
    theme(axis.title.y = element_markdown(angle = 0, vjust = 0.5))

    ggsave(fign, plot = fig)

txtn <- paste(i,'LM.txt',sep = '_')
write(stargazer(lm(Ct ~ NH4, data = fdt), type = 'text'), txtn)

}
write.csv(astat, "Anova.csv")
```

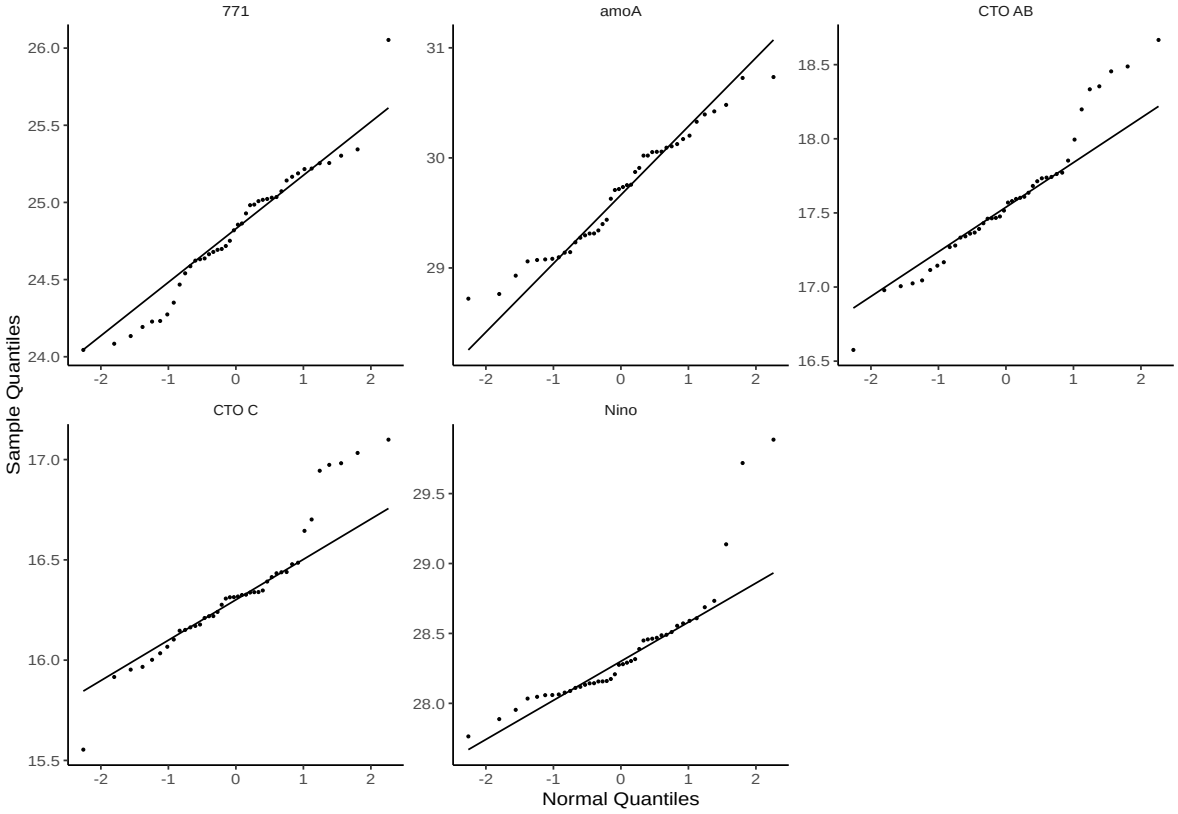
Listing A.7: R script for statistical analysis of variance in Ct value across different reaction zones as well as correlation between ammonium concentration and Ct value

A.6.3 Results of statistical analysis

Histogram and qq-plots show the data does not clearly follow normal distribution (A.6). As a result, statistical tests such as anova must be evaluated with this in mind.



(a) Histogram of Ct values for evaluated primer pairs



(b) Quantile-quantile plot of Ct values for evaluated primer pairs

Figure A.6: plots related to normality of Ct values obtained through qPCR, analysing expression of genes involved in nitrogen removal in waste water.

A.7 Chemical parameters of the waste water

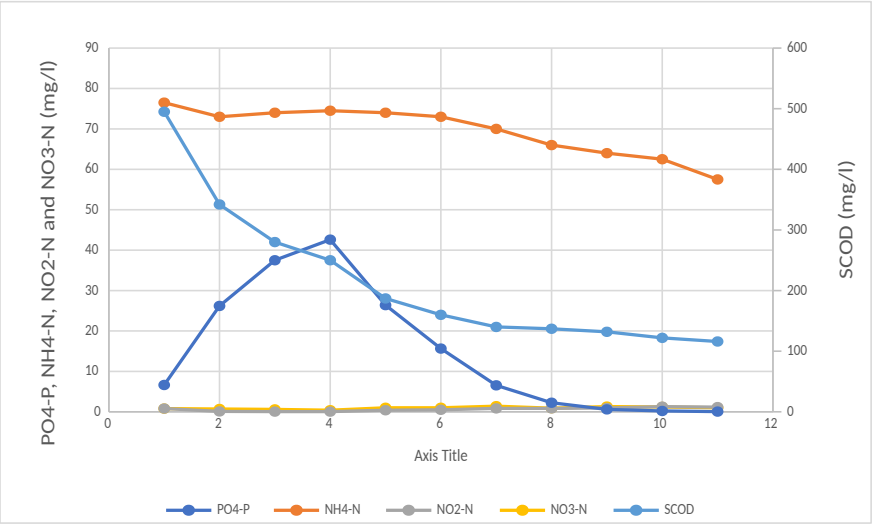


Figure A.7: Plot of chemical parameters of waste water in Hias line 6, measured in the different reactor zones.

Table A.9: Chemical parameters of waste water in HIAS line 6 at flow rate of 15 l/s, measured on 27th of januari 2022 at water temperature = 9 °C in mg/l

	PO4-P	NH4-N	NO2-N	NO3-N	SCOD
Inn Bio	6,64	76,5	0,81	0,8	495
F1	26,2	73	0,1	0,7	342
F2	37,5	74	0,05	0,6	280
F3	42,6	74,5	0,05	0,4	250
F4	26,4	74	0,39	1	187
F5	15,65	73	0,5	1	160
F6	6,55	70	0,85	1,4	140
F7	2,25	66	0,84	0,9	137
F8	0,6	64	0,94	1,3	132
F9	0,21	62,5	1,28	1,2	122
F10	0,055	57,5	1,16	1	116

A.8 Optimisation of gDNA extraction designed for the RNA Later extraction method

Though the RNA later method for RNA extraction was not used, a DNA extraction method to extract DNA in parallel with RNA from Hias biofilm was developed, the optimisation done in this regard may serve as future reference and is thus listed here.

The standard protocol for DNeasy Powersoil Pro Kit (Qiagen, Germany) resulted in acceptable DNA yield and quality, however not all of the biofilm material was dissolved during lysis, thus potentially leaving out gDNA from organisms only present in the innermost layers of the biofilm. Therefore different extraction conditions were tested to ensure full release of biofilm material from the carrier material. Hias biofilm carriers were collected from the Hias reactor and T-shaped sections were cut out from the carrier along the middle of each connecting plastic edge of the outer rings of the biofilm carrier and the sections were frozen to -20 °C for storage. T-shaped carrier sections were thawed and transferred to standard PowerBead Tubes included in the DNeasy Powersoil Pro kit and in tubes containing 0.2 g of 0.1 mm and 0.2 g of 0.5 mm silica beads. Powersoil Pro CD1 Buffer was added (500 μl) and the samples were vortexed at full speed for 30, 40 and 50 minutes using the Vortex Adapter for 24 (1.5–2.0 ml) tubes (Qiagen, Germany) as recommended in the Powersoil Pro Protocol. In addition, Powerbead tubes containing three T-shaped biofilm carrier sections and 500 μl of Powesoil CD1 Buffer were processed in a Mini Beadbeater 96 (Biospec, USA) two times for 40 s with 1 min rest on ice between processing to avoid the sample overheating. The carrier sections were imaged using a microscope to assess the dissociation of the biofilm from the carrier sections.

A.8.1 Results of optimisation of gDNA extraction

The DNeasy Powersoil Pro Kit (Qiagen, Germany) uses a lysis buffer in combination with mechanical force to lyse cells in the sample for extraction of DNA. Increasing the time used for mechanical lysis to 30, 40 and 50 minutes did not result in sufficient release of biofilm material from the carrier (Figure A.8). Furthermore, there was no significant improvement observed between biofilm carriers vortexed for 30 and 50 minutes. Though using the custom bead mix did appear to improve release minimally, there was still much biofilm left on the carrier after lysis. Increasing the intensity of the mechanical lysis however, did result in complete release of all biofilm material on the carrier (Figure A.9). Extraction of DNA using this method still resulted in DNA integrity expected for samples of biofilm from waste water (Figure A.10).

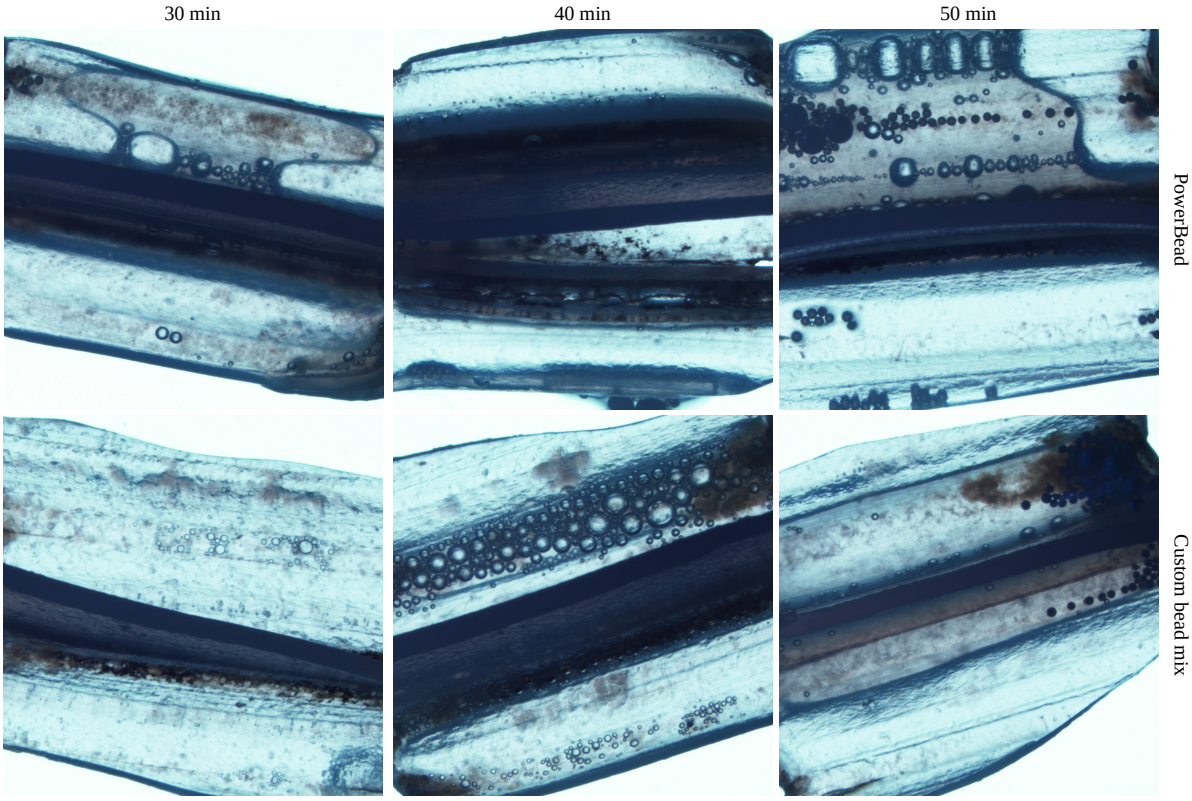


Figure A.8: Sections of biofilm carrier, vortexed using the Vortex Adapter for 24 (1.5–2.0 ml) tubes (Qiagen, Germany) for 30, 40 or 50 minutes in PowerBead tubes (included in DNeasy Powersoil Pro kit, Qiagen, Germany) or using a Custom bead mix (0.2 g of 0.1 mm and 0.2 g of 0.5 mm silica beads) with 500 μl of Powersoil Pro CD1 Buffer (Qiagen, Germany)

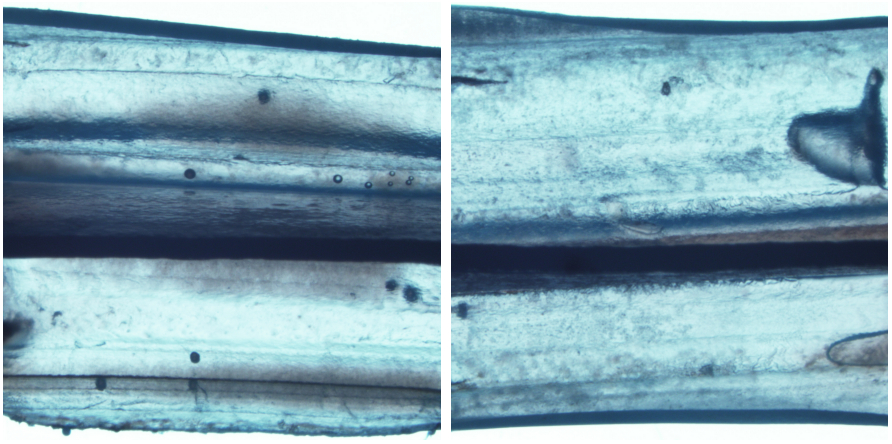


Figure A.9: Sections of biofilm carrier, processed in the Mini Beadbeater 96 (Biospec, USA) two times for 40 s with 1 min rest on ice between processing in PowerBead tubes with 500 μl of CD1 Buffer (DNeasy Powersoil kit, Qiagen, Germany)

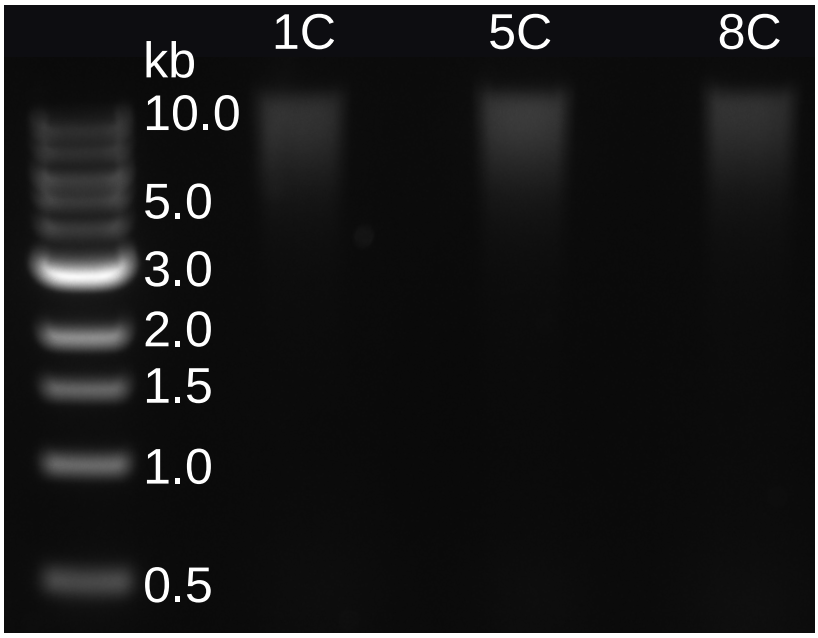


Figure A.10: genomic DNA extracted from biofilm carriers found in different reactor zones (1, 5 and 8) in a waste water treatment plant using the DNeasy Powersoil kit (Qiagen, Germany) with beadbeating as mechanical lysis, separated using 1% agarose gel electrophoresis

B. Appendix: 16S rDNA sequencing

B.1 Blast of ref_seq AOA against obtained OTUs

Ref_seq ammonium oxidising archaea 16S rDNA sequences (accession NR_159208.1, NR_159207.1, NR_159206.1, NR_102913.1 and NR_134097.1) were blasted against the OTUs obtained from the Hias reactor using blastn. No hits were found. Blast output:
BLASTN 2.13.0+

Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", J Comput Biol 2000; 7(1-2):203-14.

Database: Fasta/otus.fa 1,701 sequences; 697,760 total letters

Query= NR_159208.1 Nitrosopumilus ureiphilus strain PS0 16S ribosomal RNA, complete sequence

Length=1471

***** No hits found *****

Lambda K H 1.33 0.621 1.12

Gapped Lambda K H 1.28 0.460 0.850

Effective search space used: 959956550

Query= NR_159207.1 Nitrosopumilus oxyclinae strain HCE1 16S ribosomal RNA, complete sequence

Length=1471

***** No hits found *****

Lambda K H 1.33 0.621 1.12

Gapped Lambda K H 1.28 0.460 0.850

Effective search space used: 959956550

Query= NR_159206.1 Nitrosopumilus cobalaminigenes strain HCA1 16S ribosomal RNA, complete sequence

Length=1471

***** No hits found *****

Lambda K H 1.33 0.621 1.12

Gapped Lambda K H 1.28 0.460 0.850

Effective search space used: 959956550

Query= NR_102913.1 Nitrosopumilus maritimus SCM1 16S ribosomal RNA, complete sequence

Length=1469

***** No hits found *****

Lambda K H 1.33 0.621 1.12

Gapped Lambda K H 1.28 0.460 0.850

Effective search space used: 958632472

Query= NR_134097.1 Nitrososphaera viennensis strain EN76 16S ribosomal RNA, complete sequence

Length=1470
***** No hits found *****
Lambda K H 1.33 0.621 1.12
Gapped Lambda K H 1.28 0.460 0.850
Effective search space used: 959294511
Database: Fasta/otus.fa Posted date: Apr 27, 2022 1:36 PM Number of letters in database: 697,760 Number of sequences in database: 1,701
Matrix: blastn matrix 1 -2 Gap Penalties: Existence: 0, Extension: 2.5

B.2 Data Handling

```
# Author: Sverre Branders
library('tidyr')
library('dplyr')
library('ggplot2')
library('ggrepel')
library('ggtext')
library('stringr')

# Define colours
ramp <- colorRamp(c('#8fd44a','#304b69','#593677'))
ramp.list <- rgb( ramp(seq(0, 1, length = 7)), max = 255)
ramp.otu <- rgb( ramp(seq(0, 1, length = 36)), max = 255)
ramp.otu[37] <- '#6E7B8B'

#=====================================================
# Data Import and Formatting
#=====================================================
otu <- read.csv('./blast-out/top_hits_table.csv')
tax <- read.table(file = './raw/MiDAS_4-8-1_tax.txt', sep = ';')
meta <- read.table(file = './raw/MiDAS_Metadata.csv', sep = ';', header = T)
tax <- tax %>%
  separate(V1, c('Seq','V1'), sep = '\t')
tax <- tax %>%
  mutate(Kingdom = str_replace(V1, "k__" , "")) %>%
  mutate(Phylum = str_replace(V2, "p__" , "")) %>%
  mutate(Class = str_replace(V3, "c__" , "")) %>%
  mutate(Order = str_replace(V4, "o__" , "")) %>%
  mutate(Family = str_replace(V5, "f__" , "")) %>%
  mutate(Group = str_replace(V6, "g__" , "")) %>%
  mutate(Species = str_replace(V7, "s__" , "")) %>%
  select(-c(2:8)) %>%
  mutate(Kingdom = str_replace(Kingdom, " ", "")) %>%
  mutate(Phylum = str_replace(Phylum, " ", "")) %>%
  mutate(Class = str_replace(Class, " ", "")) %>%
  mutate(Order = str_replace(Order, " ", "")) %>%
  mutate(Family = str_replace(Family, " ", "")) %>%
  mutate(Group = str_replace(Group, " ", "")) %>%
  mutate(Species = str_replace(Species, " ", ""))
colnames(meta)[1] <- 'Group'
```

```r
colnames(meta)[2] <- 'Species'

count <- read.csv('./raw/rarefied_25000_OTU_table_with_taxonomy.csv')
count <- count[,c(1:26)]
count <- count %>%
  pivot_longer(cols = colnames(count)[2:26], names_to = 'Sample', values_to = 'Count') %>%
  mutate(Sample = str_replace(Sample, '[a-zA-Z]' , '')) %>%
  group_by(Sample) %>%
  mutate(Abundance = 100*Count/sum(Count)) %>%
  mutate(Zone = str_replace(Sample, '[a-zA-Z]$' , '')) %>%
  mutate(Zone = factor(Zone, levels = c(1:10))) %>%
  rename(OTU = OTU.ID)

# Merging Metadata to OTUs
otu_tax <- merge(otu, tax, by = 'Seq')
otu_meta <- left_join(otu_tax, meta, by = c('Kingdom','Phylum','Class','Order','Family','Group'))

# Extract groups metadata
groups <- otu_meta %>%
  filter(AOB.In.situ == 'pos' | AOB.Other == 'pos' | NOB.In.situ == 'pos' | NOB.Other == 'pos'
         | Anammox.In.situ == 'pos' | Anammox.Other == 'pos' | PAO.In.situ == 'pos' |
           PAO.Other == 'pos' | GAO.In.situ == 'pos' | GAO.Other == 'pos')
groups <- groups[,c(2,15:20,23:26)]
groups <- groups[!duplicated(groups),]%>%
  mutate(Meta = case_when(AOB.In.situ == 'pos' | AOB.Other == 'pos' ~ 'AOB',
                   NOB.In.situ == 'pos' | NOB.Other == 'pos' ~ 'NOB',
                   Anammox.In.situ == 'pos' | Anammox.Other == 'pos' ~ 'Anammox',
                   PAO.In.situ == 'pos' | PAO.Other == 'pos' ~ 'PAO',
                   GAO.In.situ == 'pos' | GAO.Other == 'pos' ~ 'GAO'))
groups <- groups[,c(1,12)]

# Funct
group_count <- merge(groups, count, by = "OTU", all.x = T)
group_count_zone <- group_count %>%
  filter(Sample != '4A') %>%
  group_by(across(all_of(c('Zone', 'Meta')))) %>%
  mutate(zone_abund = (sum(Abundance)/n_distinct(Sample))) %>%
  select(Zone, zone_abund, Meta)
group_count_zone <- group_count_zone[!duplicated(group_count_zone),] %>%
  mutate(Zone = factor(Zone, levels = c(1:10)))

# Heterotroph
hetero <- otu_meta %>%
  filter(Aerobic.heterotroph.In.situ == 'pos' | Aerobic.heterotroph.Other == 'pos')
hetero <- hetero[!duplicated(hetero),2]
hetero_count <- merge(as_tibble(hetero), count, by.x = "value", by.y = "OTU", all.x = T)
hetero_count <- hetero_count[!duplicated(hetero_count),] %>%
```

```
  filter(Sample != '4A') %>%
  group_by(Zone) %>%
  mutate(hetero_abund = (sum(Abundance)/n_distinct(Sample))) %>%
  select(Zone, hetero_abund)
hetero_count <- hetero_count[!duplicated(hetero_count),] %>%
  mutate(Zone = factor(Zone, levels = c(1:10))) %>%
  mutate(Name = 'Heterotroph')

# Abund
phyl_count <- merge(otu, tax, by = 'Seq', all.x = T) %>%
  mutate(Phylum = str_replace(Phylum, '^(\\S*)$', '*\\1*')) %>%
  mutate(Phylum = ifelse(is.na(Phylum), "Unidentified", Phylum)) %>%
  select(OTU, Phylum) %>%
  merge(count, by = 'OTU') %>%
  group_by(OTU) %>%
  filter(Sample != '4A')
phyl_count <- phyl_count[!duplicated(phyl_count),] %>%
  group_by(across(all_of(c('Phylum', 'Sample')))) %>%
  mutate(Abundance = sum(Abundance)) %>%
  select(-OTU, -Count)
phyl_count <- phyl_count[!duplicated(phyl_count),] %>%
  group_by(across(all_of(c('Phylum', 'Zone')))) %>%
  mutate(Abundance = sum(Abundance)/n_distinct(Sample)) %>%
  select(-Sample)
phyl_count <- phyl_count[!duplicated(phyl_count),]

# Output otu_meta
write.csv(otu_meta, file = './OTU_Midas_Metadata.csv')

#=====================================================
# Plot
#=====================================================

# Funct
group_zone_plot <- ggplot(group_count_zone, aes(x = Zone, y = zone_abund, fill
    = Zone)) +
  geom_col(position = 'dodge', width = 0.97, show.legend = F)+
  scale_fill_manual(values = ramp.list)+
  xlab('Zone')+
  ylab('Abundance (%)')+
  theme_classic()+
  facet_wrap(~Meta, ncol = 4, strip.position='bottom')+
  theme(strip.background = element_blank(), strip.text = element_text(size =
    12))
ggsave('./out/MiDas_funct.svg', plot = group_zone_plot)

# Heterotroph
hetero_plot <- ggplot(hetero_count, aes(x = Zone, y = hetero_abund, fill =
    Zone)) +
  geom_col(position = 'dodge', width = 0.97, show.legend = F)+
  scale_fill_manual(values = ramp.list)+
  xlab('Zone')+
  ylab('Abundance (%)')+
  theme_classic()+
```

```
  facet_wrap(~Name, labeller = label_parsed,  strip.position='bottom')+
  theme(strip.background = element_blank(), strip.text = element_text(size =
    12))
ggsave('./out/MiDas_heterotro.svg', plot = hetero_plot ,width = 3, height = 5)

# Abund
phyl_plot <- phyl_count %>%
  group_by(Phylum) %>%
  filter(sum(Abundance) > 4.2) %>%
  ungroup() %>%
  ggplot(aes(x = Abundance, y = reorder(Phylum, -Abundance), fill = Zone ))+
  geom_col(position = 'dodge', width = 0.97)+
  scale_fill_manual(values = ramp.list)+
  theme_classic()+
  xlab('Abuncance (%)')+
  ylab('Phylum')+
  theme(axis.text.y = element_markdown(size = 12),
        legend.position = c(0.9, 0.5))
ggsave('./out/MiDas_top_abund.svg', plot = phyl_plot ,width = 7, height = 7)

phyl_zone_plot <- phyl_count %>%
  mutate(zone = factor(Zone, levels = c(1:10))) %>%
  ggplot(aes(x = Zone, y = Abundance, fill = Phylum))+
  geom_col(position = 'stack')+
  scale_fill_manual(values = ramp.otu)+
  scale_x_discrete()+
  labs(x = 'Zone', y = 'Relative Abundance (%)')+
  theme_classic()+
  theme(legend.text = element_markdown())
ggsave('./out/MiDas_abund.svg', plot = phyl_zone_plot)

#=====================================================
# Detecting groups
#=====================================================

# Find AOA (phylum Thaumarchaeota/ Nitrososphaerota)
arch_otus <- otu_meta %>%
  filter(Kingdom == 'Archaea')
arch_otus <-unique(arch_otus$Phylum)
print(arch_otus)
# Find Comammox (in genus Nitrospira)
coma_otus <- otu_meta %>%
  filter(str_detect(Group, 'Nitro*'))
coma_otus <- unique(coma_otus$Group)
print(coma_otus)
```

Listing B.1: R script for generating figures related to OTU taxonomy using the MiDas database

```
# Author: Sverre Branders
library('tidyr')
library('tibble')
library('dplyr')
library('stringr')
library('ggplot2')
```

```
library('ggrepel')
library('ggtext')
library('ggpmisc')
library(vegan)

# Define colours
ramp <- colorRamp(c('#8fd44a','#304b69','#593677'))
ramp.list <- rgb( ramp(seq(0, 1, length = 7)), max = 255)
ramp.8 <- append(ramp.list,'Black')

#============================================================
# ALPHA DIVERSITY
#============================================================
# Define functions
sample_order <- c("1A","1B","1C","1D","1E","2A","2B","2C","3A","3B","3C","4B",
    "4C","4D","4E","6A","6B","6C","8A","8B","8C","10A","10B","10C")
clean_a_dt <- function(b){
  b %>%
    filter(X != 'Pos') %>%
    mutate(zone = str_replace(X, '[a-zA-Z]$' , '')) %>%
    rename(sample = X) %>%
    arrange(factor(sample, levels = sample_order)) %>%
    mutate(sample = as.character(sample)) %>%
    mutate(zone = factor(zone, levels = c(1:10))) %>%
    mutate(id = c(1:(nrow(b)-1))) %>%
    ungroup()
}
my_a_plot <- function(j){
  j %>%
    ggplot(aes(x = id, y = value))+
    geom_smooth(aes(x = id), method = 'lm', se = F, color = 'grey', size =
    0.5)+
    geom_line(color = '#304b69')+
    geom_point(aes(x = sample, color = zone), size = 3)+
    stat_poly_eq(
      aes(label = paste0("atop(", ..eq.label.., ",", ..rr.label.., ")")),
      rr.digits = 3,
      parse = TRUE,
      label.y = 'top',
      label.x = 'right',
      hstep = 2,
      size = 10)+
    xlab('Sample')+
    scale_x_discrete(limits = sample_order)+
    scale_color_manual(values = ramp.list)+
    theme_classic(base_size = 25)+
    theme(axis.text.x = element_text(size = 18, angle = -90, hjust = 0, vjust
    = 0.5))

}

alpha_zone <- function(i){
  b <-   summary(aov(value ~ zone, data = i))[[1]][["Pr(>F)"]][1]
  b <- paste('P-value: ' , round(b, digits = 5))
```

```
  c <-   summary(aov(value ~ zone, data = i))[[1]][["F value"]][1]
  c <- paste('F-value: ' , round(c, digits = 5))
  d <- paste(c,"\n",b,sep = "")
  a <- i %>%
    group_by(zone) %>%
    summarise(avr = mean(value),
              std = sd(value),
    )
  a[,'id'] <- c(1,2,3,4,6,8,10)
  text_y <- max(a$avr) + (0.1* (max(a$avr) - min(a$avr)))

  a %>%
    ggplot(aes(x = id, y = avr))+
    geom_line(color = '#304b69', show.legend = F)+
    geom_point(aes(x = id, color = zone), size = 3, show.legend = F)+
    geom_errorbar(aes(ymin = (avr-std), ymax = (avr+std), color = zone), show.
    legend = F)+
    geom_text(aes(y = text_y, x = 10), label = d,
                  color = 'Black', size = 10, hjust = 1, check_overlap = T)+
    xlab('Zone')+
    ylab('Average value')+
    scale_color_manual(values = ramp.list)+
    scale_x_continuous(breaks = c(1,2,3,4,6,8,10))+
    theme_classic(base_size = 25)
}

# Data
shannon_entropy <- read.table("./QIIME/Diversity_metrics/shannon_vector/data/
    alpha-diversity.tsv", sep = '\t', header = T)
shannon_entropy <- clean_a_dt(shannon_entropy) %>%
  rename(value = shannon_entropy)
shannon_zone <- alpha_zone(shannon_entropy)
ggsave('./out/QIIME/OTU/shannon_entropy_zone.svg', plot = shannon_zone)
shannon_entropy_plot <- my_a_plot(shannon_entropy)
ggsave('./out/QIIME/OTU/shannon_entropy.svg', plot = shannon_entropy_plot)

pielou_evenness <- read.table("./QIIME/Diversity_metrics/evenness_vector/data/
    alpha-diversity.tsv", sep = '\t', header = T)
pielou_evenness <- clean_a_dt(pielou_evenness) %>%
  rename(value = pielou_evenness)
alpha_anova(pielou_evenness)
pielou_zone <- alpha_zone(pielou_evenness)
ggsave('./out/QIIME/OTU/pielou_evenness_zone.svg', plot = pielou_zone)
pielou_evenness <- my_a_plot(pielou_evenness)
ggsave('./out/QIIME/OTU/pielou_evenness.svg', plot = pielou_evenness)

faith_pd <- read.table("./QIIME/Diversity_metrics/faith_pd_vector/data/alpha-
    diversity.tsv", sep = '\t', header = T)
faith_pd <- clean_a_dt(faith_pd) %>%
  rename(value = faith_pd)
faith_zone <- alpha_zone(faith_pd)
ggsave('./out/QIIME/OTU/faith_pd_zone.svg', plot = faith_zone)
faith_pd <- my_a_plot(faith_pd)
ggsave('./out/QIIME/OTU/faith_pd.svg', plot = faith_pd)
```

```
108
109 observed_features <- read.table("./QIIME/Diversity_metrics/observed_features_
    vector/data/alpha-diversity.tsv", sep = '\t', header = T)
110 observed_features <- clean_a_dt(observed_features) %>%
111   rename(value = observed_features)
112 observed_zone <- alpha_zone(observed_features)
113 ggsave('./out/QIIME/OTU/observed_features_zone.svg', plot = observed_zone)
114 observed_features <- my_a_plot(observed_features)
115 ggsave('./out/QIIME/OTU/observed_features.svg', plot = observed_features)
116
117
118 #=========================================================
119 # BETA DIVERSITY
120 #=========================================================
121 # Define functions
122 my_b_plot <- function(i){
123   a <- i[-25,-25]
124   a <- metaMDS(a)
125   a <- as.data.frame(a$points) %>%
126     rownames_to_column('sample') %>%
127     arrange(factor(sample, levels = sample_order)) %>%
128     mutate(zone = str_replace(sample, '[a-zA-Z]$' , '')) %>%
129     mutate(zone = factor(zone, levels = c(1:10)))
130   b_stat <- adonis2(i[-25,-25]~zone, data = a, permutations = perm)
131   c <- d <- paste(paste('P-value: ' , round(b_stat$`Pr(>F)`[1], digits = 5)),
132                   "\n",
133                   paste('R²: ' , round(b_stat$R2[1], digits = 5)),sep = "")
134
135 text_x <- max(a$MDS1) + (0.2* (max(a$MDS1) - min(a$MDS1)))
136 text_y <- max(a$MDS2) + (0.3* (max(a$MDS2) - min(a$MDS2)))
137
138   ggplot(a, aes(x = MDS1, y = MDS2, color = zone))+
139     geom_text(aes(x = text_x, y = text_y),label = c, color = 'Black', size =
    10, hjust = 1, vjust = 1, check_overlap = T)+
140     geom_point(size = 3)+
141     geom_text_repel(aes(label = sample), show.legend = F, size = 6)+
142     scale_color_manual(values = ramp.list)+
143     scale_fill_manual(values = ramp.list)+
144     scale_linetype_manual(element_blank())+
145     expand_limits(y = text_y, x = text_x)+
146     theme_classic(base_size = 25)
147 }
148
149
150 # Seed
151 set.seed(79836528)
152 perm <- 100000
153 # Data
154 BC <- read.table("./QIIME/Diversity_metrics/bray_curtis_distance_matrix/data/
    distance-matrix.tsv", sep = '\t', header = T, row.names = 1)
155 BC_plot <- my_b_plot(BC)
156 ggsave('./out/QIIME/OTU/BC_NMDS.svg', plot = BC_plot)
157
158 jaccard <- read.table("./QIIME/Diversity_metrics/jaccard_distance_matrix/data/
```

```
    distance-matrix.tsv", sep = '\t', header = T, row.names = 1)
j_plot <- my_b_plot(jaccard)
ggsave('./out/QIIME/OTU/Jaccard_NMDS.svg', plot = j_plot)

uw_unifrac <- read.table("./QIIME/Diversity_metrics/unweighted_unifrac_
    distance_matrix/data/distance-matrix.tsv", sep = '\t', header = T, row.
    names = 1)
uwu_plot <- my_b_plot(uw_unifrac)
ggsave('./out/QIIME/OTU/Unweighted_Unifrac_NMDS.svg', plot = uwu_plot)

w_unifrac <- read.table("./QIIME/Diversity_metrics/weighted_unifrac_distance_
    matrix/data/distance-matrix.tsv", sep = '\t', header = T, row.names = 1)
wu_plot <- my_b_plot(w_unifrac)
ggsave('./out/QIIME/OTU/Weighted_Unifrac_NMDS.svg', plot = wu_plot)
```

Listing B.2: R script for generating figures related to alpha and beta metrics generated using QIIME