Optimising and applying RNA based approaches to identify active nitrifiers in coastal sediments

Fabien Cholet MSc

Submitted in fulfilment of the requirements for the Degree of
Doctor of Philosophy (PhD) by Research

James Watt School of Engineering,

University of Glasgow

Table of content

Declaration of Originality

I, Fabien Cholet, certify that this thesis is my own work, other than where I clearly indicated the use of previous data or the contribution of others. I have:

Read and understood the guidance on plagiarism in the Student Handbook, including the University of Glasgow Statement on Plagiarism.

Clearly referenced, in both the text and the bibliography or references, all sources used in this thesis.

Fully referenced and used inverted commas for all text quoted from books, journals etc.
Provided the sources for all tables, figures, data etc. that are not my own work.

Not made use of the work of any other student(s) past or present without acknowledgement. This includes any of my own work, which has been previously, or concurrently, submitted for assessment, either at this or any other educational institution, including school.

Not sought or used the services of any professional agencies to produce this work.

In addition, I understand that any false claim in respect of this work will result in disciplinary action in accordance with University regulations.

Acknowledgement

First and foremost I would like to thank my supervisor Pr Cindy Smith. Starting from my Master's internship in 2015 in Galway and following with this PhD thesis, Cindy has given me an amazing opportunity to develop a scientific career and I am truly grateful for that. Throughout the last 4 years of my PhD thesis, she has always provided support, advice and encouragement. It is an understatement to say I could not have achieved this much without her help.

I am very grateful to Dr Umer Ijaz. I have learnt a great deal by spending time with him and I thank him for his time and patience. His expertise in bioinformatics and statistics has hugely helped me.

A huge thanks you to Julie Russel and Anne McGarrity. Their work, dedication and kindness made a big difference and provided the perfect environment in the lab for me and many other students.

I would like to thank the head of the department William Sloan and all and other members of the Infrastructure and Environment Dr Marta Vignola, Dr Rebeca Gonzalez-Cabaliero, Dr Eri Tsagkari, Dr Ciara Keating, Dr Stephanie Connelly and Dr Bachar Cheiab for their help and friendly attitude. A special thank you to Dr Aoife Duff for all the help and guidance at the beginning of my project.

Thank you to all the members of Cindy's group Dominic Quin, Valentine Okonkwo, Claire Thom, Kelly Steward and Dr Anastasiia Kostrytsia for the help, advice and laughter during the Friday meetings.

Thank you to everyone who helped me at the University of La Rochelle Hélène Agogué for the help, support and making the exchange possible, Philippe Pineau and Nicolas Lachaussée for the help with the sampling and Martine Bréret for the help in the lab.
I would also like to thanks the friends, PhD students and others I met during my time in Glasgow for all the good times, laughs, discussions and fun. I am grateful for having met such a nice group of people.

I want to thank my family, for their constant love, support and encouragement and especially my parents for always supporting me in my decisions and for their financial and moral backing whenever I needed it. Finally, a special thank you to Mary-Julie for her love and her kindness and to her family for making me feel welcome in Glasgow.

Abstract

Nitrogen is an essential element for all forms of life on earth. Since the beginning of the industrial revolution, nitrogen has become a major pollutant of marine and coastal ecosystems due to the huge rise in the use of inorganic fertiliser. Like many other nutrients, the transformations of the nitrogen cycle are mainly controlled by the activity of microorganisms. Understanding the factors influencing the activity of microbes involved in the biochemical transformation of nitrogen in the environment is therefore crucial.

The aim of this thesis is to establish a robust workflow for the study of microbial activity in coastal sediment using transcriptomics. In particular, this work focuses on nitrification, the aerobic chemo-litho-autotrophic oxidation of ammonia to nitrite carried out by ammonia oxidizing bacteria and archaea (AOB and AOA respectively).

The First part of the thesis (Chapter I) will consist of a review of the literature on the nitrogen cycle, with a particular focus on nitrification. A review of the techniques used to measure microbial activity in natural environment will also be presented and the knowledge gap that exist in transcriptomic workflow in environmental microbiology identified alongside current understanding of active nitrifiers in coastal sediments.

The second part of the thesis (Chapters II and III) will present the first experimental work package which consists of the optimisation of reverse-transcription (RT)- based protocol for the study of microbial activity via transcriptomics. First, a new technique to evaluate RNA integrity, extracted from environmental samples, based mRNA will be developed and tested in a controlled-RNA degradation experiment. We show that this technique can provide a useful complement to the commercial approaches that evaluate RNA integrity mainly through the 16S/23S rRNA ratio. Then, the effect of the RT protocol itself on RT-Q-PCR and RT-PCR-sequencing results will be evaluated by testing a combination of four different RT enzymes and two priming strategies. We show that the choice of the correct protocol can greatly improve accuracy and precision of RT-based results.

The third part (Chapters IV) will present the application of the optimised protocol to study the effects of sedimentary structures (ridge/runnel) on microbial nitrification activity measured via reverse-transcriptase quantitative PCR (RT-Q-PCR) and reverse-transcriptase PCR-sequencing. Here, the work developed in part two to ensure RNA integrity and optimal RT-

PCR protocols will be applied to ensure robust and reliable measure of nitrifier mRNA from within coastal sediments to inform ecological understanding of the active organisms and controls of nitrification. The study site chosen was the Montportail-Brouage intertidal mudflat, located along the French Atlantic coast. This site has been shown to display interesting characteristics in term of microbial dynamics, with the sedimentary structures (ridges/runnels) significantly influencing microbial nitrification rates. The hypothesis proposed previously to explain the differences in nitrification rates is that AOB are more abundant in the runnels, where the higher nitrification rates had been measured. Here, we will show that these differences are explained by the presence of low abundance but highly active AOB groups that drive ammonia oxidation. Furthermore, we show the inadequacy of DNA studies as stand alone methods to explore nitrification activity, with a negative correlation between abundance of AOB *amoA* genes and nitrification rates, due to the presence of a highly abundant but inactive AOB cluster.

List of Figures

List of Tables

List of Abbreviations

Biochemical processes

ANAMMOX: Anaerobic ammonia oxidation
ANR: Assimilatory Nitrate Reduction
CBB: Calvin-Benson-Bassham cycle
COMAMMOX: Complete ammonia oxidation
SULFAMMOX: Anaerobic ammonia oxidation using sulphate
FEAMMOX: Anaerobic ammonia oxidation using iron
DNRA: Dissimilatory Nitrate Reduction to Ammonia
HN: Heterotrophic Nitrification
IAMO: Intra-Aerobic Methane Oxidation
ND: Nitrifier Denitrification
TCA: tricarboxylic Acid Cycle
ARMN: anaerobic oxidation of methane via reverse methanogenesis using nitrate
DSR: dissimilatory reduction of sulphate to sulphide

Molecular terms and techniques

ATP: Adenosine Triphosphate
(c)DNA: (complementary) Deoxyribonucleic Acid
Mo-bis-MGD: Molybdenum *bis* molybdopterin guanine dinucleotide
NADH: Nicotinamide Adenosine Dinucleotide
RT-(Q)-PCR: Reverse Transcriptase-(Quantitative)-polymerase Chain Reaction
(m/r) RNA: (messenger/ribosomal) Ribonucleic Acid
Transcriptase)
HGT: Horizontal Gene Transfer
SIP: Stable Isotope Probing
RH: Random hexamer
GS: Gene Specific
SSIII/IV: Superscript III/IV
Omni: Omniscript
Sensi: Sensiscript
R_{amp}: Ratio Amplicon
Δamp: Differential Amplicon
RIN: RNA Integrity Number
RQS: RNA Quality Score
RT-(Q)-PCR: Reverse-Transcriptase-(Quantitative)-Polymerase Chain Reaction
Ct: Cycle Threshold
rRNA: ribosomal RNA
mRNA: messenger RNA
cDNA: complementary DNA
bp: base pair
OTU: Operational Taxonomic Unit
ASV: Amplicon Sequence Variant
RACE: Rapid Amplification of cDNA Ends
R_{amp}: Ratio Amplicon

Microorganisms

AOA: Ammonia Oxidizing Archaea
AOB: Ammonia Oxidizing Bacteria
AOM: Ammonia Oxidizing Microbes
MOB: Methane Oxidizing Bacteria
(p)NOB: (phototrophic) Nitrite Oxidizing Bacteria
NS: Nitrospira-like NOB
NB: Nitrobacter-like NOB
SOM: Sulphur Oxidizing Microbes

Enzymes/proteins

Amo: Ammonia Mono Oxygenase
CCP: Cytochrome *c* peroxidase
Ccm552: Membrane cytochrome
CP_{460}: Cytochrome P 460
Cu aa3: Cytochrome *c* oxidase
Cu Nir: Copper containing nitrite reductase
GAPDH: Glyceraldehyde 3-Phosphate Dehydrogenase
GFP: Green Fluorescent Protein
Hao: Hydroxylamine Oxidoreductase
Hzo: Hydrazine Oxidoreductase
MCC: Multi heme Cytochrome C
pMMO: particulate Methane Mono Oxygenase
Nap: Periplasmic dissimilatory nitrate reductase
Nar: Membrane-bound dissimilatory nitrate reductase
Nas: Prokaryotic Assimilatory nitrate reductase
Nif: Nitrogenase
Nir: Nitrite Reductase
Nod: nitric oxide dismutase
Nos: Nitrous Oxide Reductase
Nor: Nitric Oxide Reductase
NR: Eukaryotic assimilatory nitrate Reductase

Nrf: Nitrite reductase
Nxr: Nitrite oxido-Reductase
Q: Ubiquinone
GFP: Green Fluorescent Protein

Genes
amoA: Ammonia monooxygenase subunit A
glnA: Glutamine synthetase subunit A
nxrB: Nitrite oxydoreductase subunit B
nirS: Nitrite reductase subunit S
nirK: Nitrite reductase subunit K
nrfA: (ammonia forming) Nitrite reductase subunit A
16S rRNA: 16S ribosomal RNA gene
18S rRNA: 18S ribosomal RNA gene
pmoA: particulate methane monooxygenase

General terms
EOCs: Emerging trace Organic contaminants
EPS: Exo-Polymeric Substances
OMZ: Oxygen minimum zone
GWP: Global Warming Power
PMF: Proton Motive Force
QS: Quorum Sensing
TEX_{86}: TetraEther indeX of tetraethers consisting of 86 carbon atoms
RIN: RNA Integrity Number
WWTP: Waste Water Treatment Plant
(PERM)ANOVA
TukeyHSD: Tukey Honestly Significant Difference
PEG: Poly Ethylene Glycol
CTAB: CetylTrimethylAmmonium **Bromide**
DTT: DiThioThreitol
PNR: Potential Nitrification Rate
ChllA: Chlorophyll A
Pheop: Pheophytin

Physio-chemical terms
TOC: Total Organic Carbon
DOC: Dissolved Organic Carbon
TN: Total Nitrogen
TDN: Total Dissolved Nitrogen
SGS: Sediment Grain Size

Statistical analyses
PCA: Principal Component Analysis
PCoA: Principal Component Analysis
NMDS: Non-metric Multi Dimensional Scaling
(PERM)ANOVA: (PERMutationnal) ANalysis Of Variance
CCA: Canonical Correspondence Analysis

Chapter I General Introduction

1.1 Coastal ecosystems

1.1.1 Definitions

Coastal ecosystems are transition zones between land and marine ecosystems where the spatial and temporal alternation of contrasting conditions creates zones of remarkable biological diversity and activities (McClain et al., 2003; Foster et al., 2013; Yang, 2014). They encompass different types of ecosystems such as wetlands, mangroves, estuaries, marshes, lagoons, rocky or muddy intertidal areas, beaches and dunes, coral reef systems, and nearshore coastal waters of the continental shelves. A general definition of coastal systems is: "*the line where land-based influences dominate up to a maximum of 100 kilometres from the coastline or 50-meter elevation and with the outward extent as the 50-meter depth contour*" (Agardy et al., 2005).

Tidal flats are defined as coastal ecosystems that undergo regular tidal inundations. These coastal wetlands, often found in sheltered estuaries with high sediment deposition, are strongly affected by tidal dynamics. It is estimated that tidal flats cover at least 127,921 km^2 of the earth's surface (Murray et al., 2019). Three types can be distinguished depending on the latitude: 1) low-latitudes mudflats (arid, wet tropical, and subtropical), 2) mid-latitudes mudflats of temperate regions and 3) high-latitudes mudflats, influenced by ice (Dyer et al., 2000). Depending on the climate, tidal level, substrate, hydrology, and salinity, mudflats can be inhabited by mangroves, sea-grass, or algal/microbial mats (Semeniuk, 2019). Intertidal mudflat's sediments are generally composed of mud and sand with a high mud content which makes them cohesive. They are also often characterized by high contents of organic matter, which degradation by deposit feeders and microbes stimulates primary production in the water columns (Dissanayake et al., 2018).

1.1.2 Importance

Coastal ecosystems are recognised as important areas, both from an ecological and economical point of view: these ecosystems are crucial, for example, for flood control and protection, prevention of coastline erosion, nutrient recycling, and food production. It is estimated that the goods and services they provide are higher per unit area than terrestrial habitats (Costanza et al., 2014). These transition zones between marine and terrestrial systems are among the most productive habitats on the planet and provide important services to wildlife. For example,

mudflats and saltmarshes are the living space for numerous invertebrates and provide nursery and feeding ground for fishes, therefore supporting global fisheries. They also provide feeding, nesting, and roosting areas for migrating birds and wildfowl (Foster et al., 2013; Seitz et al., 2014). Finally, they protect the coastline from erosion due to wave energy, protect from flooding, and process nutrients coming from the land (Foster et al., 2013). It is therefore important to better understand and protect these ecosystems, especially since it is estimated that up to 86% of the European coast is under moderate to high risk due to human activities (Seitz et al., 2014). In a global survey, it has been estimated that 16.02% of tidal flats have been lost since 1984 due to the combined negative effect of increased anthropogenic coastal development, reduced sedimentation, erosion, and sea-level rise (Murray et al., 2019).

1.1.3 Anthropogenic impacts

Coastal ecosystems are experiencing rapid changes due to the increase in atmospheric CO_2 concentration caused by human activities. This results in general acidification of marine waters and a general increase in temperatures. Probably the most known example of the effect of ocean acidification on coastal ecosystems is on coral reefs where acidification reduces the concentration of carbonate ions in water, making it harder for corals to construct their skeletons. This in turns affects fish populations that use coral reefs as nursery and feeding grounds (Hall-Spencer and Harvey, 2019). This effect is also felt in temperate coasts with an overall deterioration in conditions for coralline algae, which are considered habitat-forming species, and the favouring of non-calcified algae. This in turn will likely have a strong negative effect on the capacity of coastal ecosystems to mitigate erosion and reduce habitat for many crustaceans and fishes. Furthermore, some algae that proliferate in more acidic waters are toxic to fish, which could pose a serious threat to fish production (Godbold and Calosi, 2013; Hall-Spencer and Harvey, 2019).

Coastal ecosystems are also affected by terrestrial pollutants transported by freshwater to the coast. It is estimated that nearly 80% of pollutants reaching the oceans come from terrestrial sources (Agardy et al., 2005). A pollutant is a generic term that refers to objects, for example, plastics (Harse, 2011), chemicals, for example, mercury, (Gworek et al., 2016) nutrients, and living organisms (Bianchi, 2007). Nitrogen is naturally present in coastal environments, and together with phosphorous supports the growth of algae, aquatic plants, and phytoplankton, which are at the basis of the food chain. However, it can be considered a pollutant when too concentrated in the system. Urban wastewater and fertilizers used in agriculture are the two

main sources of the increased amount of nitrogen entering marine systems via rivers and estuaries (Seitzinger et al., 2002). Another important source is nitrogen from atmospheric depositions. Combustion of hydrocarbons (industry, transports) is a major source of nitrogen-containing gas emission (Paerl and Whitall 1999). These sources have drastically increased since the industrial revolution (5-10 times) (Paerl and Whitall 1999) and have major consequences on the functioning of coastal systems (Bianchi, 2007). Nitrogen, together with phosphorous is the major nutrient responsible for algal blooms, leading to eutrophication and subsequently to a reduction of oxygen concentration in the water column (Bianchi 2007; Howarth and Marino 2006). Because of their high levels of biological activity, estuaries behave as active channels that strongly interact with this excess nitrogen. These changes in nitrogen redox states (*i.e.* the nitrogen cycle) are mainly controlled by microbial activities which can either result in a mitigation of the pollution via removal or storage of excess nitrogen or indeed an aggravation of the problem by the creation of more harmful nitrogen species such as nitrous oxide (N_2O), a major ozone-depleting gas. A recent review found that overall, coastal environments are sources of nitrous oxide to the atmosphere (0.15–0.91 Tg $N_2O.yr^{-1}$) (Murray et al., 2015). This figure is relatively small compared to emissions from agricultural soils (4.5–6 Tg $N_2O.yr^{-1}$) (Shakoor et al., 2020), however, it is predicted that coastal ecosystems will strongly respond to increase nitrogen loadings, with an estimated increase in N_2O production of 190% if nitrate concentration doubles in the next decades (Murray et al., 2015).

Coastal systems and wetlands are among the most valued ecosystems on earth based on the services and benefits they provide (Costanza et al., 2014). As discussed previously, they are also one of the most threatened by human activities. Understanding nitrogen cycling in coastal and wetland systems is therefore imperative for setting up efficient monitoring and remediation strategies.

1.2 Nitrogen cycle

1.2.1 General background

Nitrogen is an essential element for all forms of life. It is a constituent of proteins, nucleic acids and many other organic and inorganic compounds found in living cells. Like most elements, nitrogen is transformed from one redox state to another via biological activities and chemical reactions (Figure 1.1), numerous steps being catalysed by microorganisms.

Nearly 80% of the atmosphere is composed of nitrogen (dinitrogen; N_2) which is unavailable to most living organisms including animals and plants. A wide range of prokaryotes that possess the nitrogenase enzyme can fix dinitrogen, by reduction the strong triple bond of the dinitrogen molecule. These prokaryotes often form symbioses with higher organisms such as plants (Biswas and Gresshoff 2014) and insects (Douglas 1998; Kneip et al., 2007). Although represented occurring in the oxic part (Figure 1.1), nitrogen fixation is O_2 sensitive and is performed by organisms living in anoxic environments or in oxic environments but capable of creating anoxic sub-environments (Bianchi, 2007).

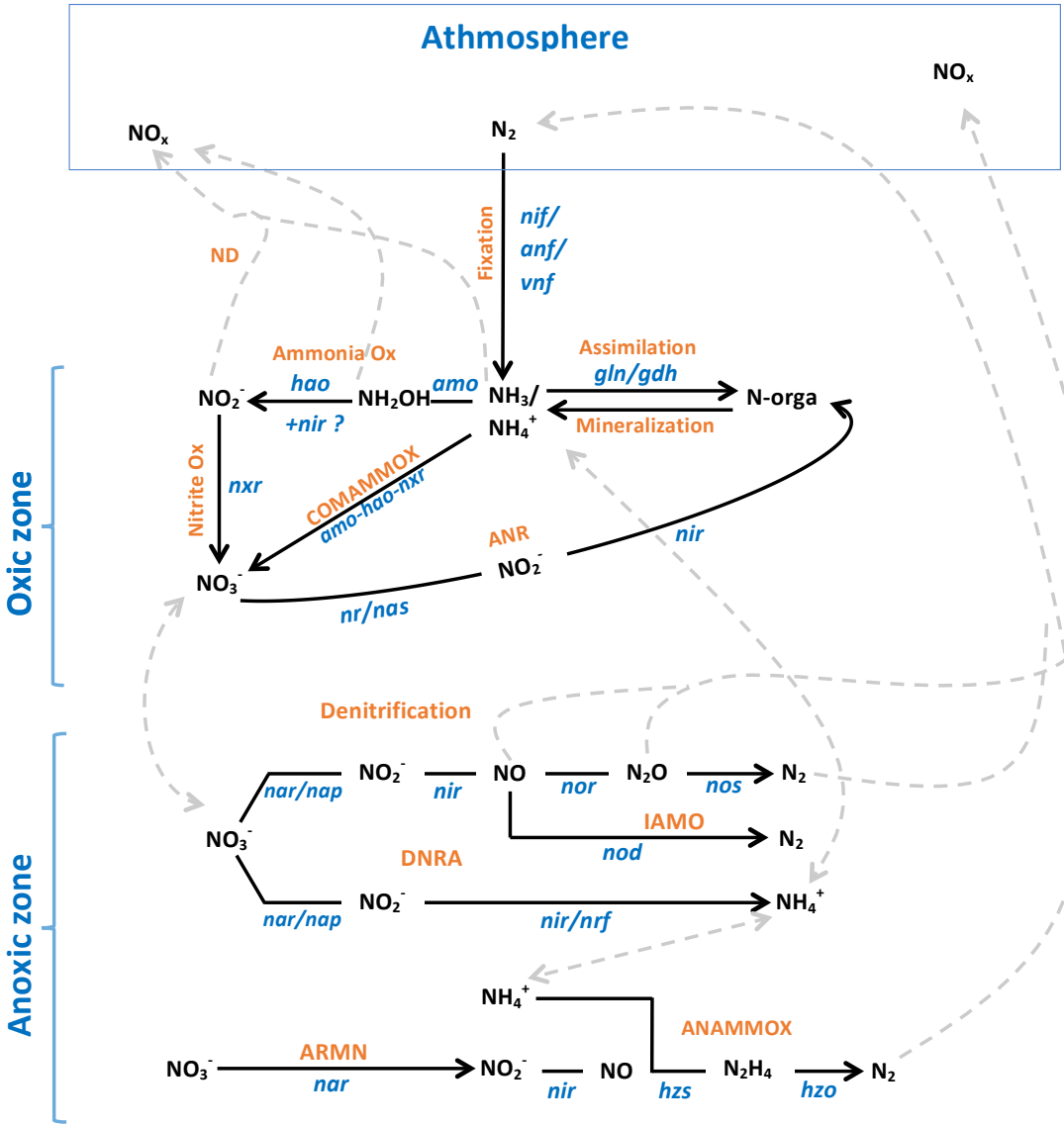


Figure 1.1 Schematic representation of the major processes of the nitrogen cycle. Microbial processes are written in orange and genes involved in blue. Slash bars (/) indicate alternative genes. Black arrows represent biochemical transformations of nitrogen species. Interrupted grey arrows represent transport of nitrogen species. *nas*: prokaryotic assimilatory nitrate reductase; *nr*: eukaryotic assimilatory nitrate reductase; *gln*: glutamine synthetase; *gdh*: glutamate dehydrogenase; *nar*/*nap*: dissimilatory nitrate reductases; *nir*: nitrite reductase; *nrf*: nitrite reductase (associated with *napA*) *nor*: nitric oxide reductase; *nos*: nitrous oxide reductase; *nif*: nitrogenase; *nxr*: nitrite oxidoreductase; *amo*: ammonia monooxygenase *hao*: hydroxylamine oxidoreductase; *hh*: hydrazine hydrolase; *hzo*: hydrazine oxidoreductase; *nod*: nitric oxide dismutase. Ammonia Ox: ammonia oxidation; Nitrite Ox: nitrite oxidation; COMAMMOX: complete ammonia oxidation; ANR: assimilatory nitrate reduction; ND: nitrifier denitrification; IAMO: Intra-Aerobic Methane Oxidation; DNRA: dissimilatory nitrate reduction to ammonia; ARMN: anaerobic oxidation of methane via reverse methanogenesis using nitrate; ANAMMOX: anaerobic ammonium oxidation. N-orga: nitrogen in organic matter. (Adapted from : Arrigo 2005; Canfield *et al.*, 2010; Gruber and Galloway 2008; Karlsson *et al.*, 2009; Kuenen 2008; Kuypers *et al.*, 2018; Moreno-vivián *et al.*, 1999; Simon and Klotz 2013).

After being incorporated into organic matter, nitrogen can be released into the environment by mineralization, mainly as ammonia (Bolle et al., 1982; Bothe et al., 2007). This reduced form of nitrogen can be re-incorporated into organic matter by assimilation. Two pathways can be used by prokaryotes for ammonia assimilation: a high-affinity pathway, prevalent in oligotrophic environments, whereby ammonia is combined with glutamate to form glutamine; this pathway is catalysed by the glutamine synthetase enzyme (encoded by *gln* genes). In the low-affinity pathway ammonia is combined with α-ketoglutarate to produce glutamate. This pathway is catalysed by the glutamate dehydrogenase enzyme (encoded by *ghd* genes) (Damashek and Francis, 2018). Ammonia can also be oxidized by a one or two-step process called nitrification. In nitrification, ammonia is first oxidized to nitrite (ammonia oxidation or nitritation) and then to nitrate (nitrite oxidation or nitratation) by two distinct groups of microorganisms, or directly converted from ammonia to nitrite and then nitrate by a single microorganism, the complete ammonia oxidizers COMAMMOX. These reactions require oxygen and are catalysed by specialised groups of microbes: ammonia-oxidation is carried out by ammonia-oxidizing bacteria (AOB) and archaea (AOA) while nitrite-oxidation is carried out by nitrite-oxidizing bacteria (NOB). COMAMMOX is carried out by a specific group of bacteria closely related to NOB. Nitrite and nitrate both have a negative charge and are therefore less bound to soil/sediment particles than ammonia. They dissolve easily in the water phase and can reach the anoxic zone where they are used by anaerobes that are capable of replacing oxygen with nitrate or nitrite as terminal electron acceptors. Many are only facultative anaerobes and are capable of respiring oxygen in oxic conditions (Damashek and Francis, 2018). They are several pathways by which oxidized forms of nitrogen are used as electron acceptors for respiration: the most simple one is the reduction of nitrate to nitrite, carried out, for example, by the ubiquitous marine clade SAR11 (Kuypers et al., 2018). Other reactions involving nitrate reduction result in the production of more reduced forms of nitrogen, such as dinitrogen, and are called denitrification. Denitrification is carried out by organo-heterotrophs that couple organic matter oxidation to nitrate/nitrite reduction and by litho-autotrophs that couple the oxidation of inorganic compounds (*e.g.* sulfur or iron) with nitrate reduction. NB: in Figure 1.1 the denitrification pathway is shown to occur in the anoxic zone but aerobic denitrification can also happen, although it generally results in the production of nitrous oxide rather than dinitrogen (Stein and Klotz 2011; Zumft 1997). It should also be noted that, although in Figure 1.1 nitrate is reduced all the way to N_2 via nitrite, nitric oxide and nitrous oxide, many denitrifiers possess only a partial pathway with some genes missing and therefore reduce nitrate only partially. Another pathway is the anaerobic ammonia oxidation

(ANAMMOX) carried out by autotrophs, where ammonia is oxidized using nitrite to produce energy and reductive power; both denitrification and ANAMMOX produce dinitrogen. Other electrons acceptors such as Iron (Fe^{3+}) (FEAMMOX (Li et al., 2015; Yao et al., 2019)) and possibly sulphate (SULFAMMOX (Liu et al., 2021)) or even manganese-oxide (Mn-ANAMMOX (Mogollón et al., 2016)) can replace nitrite as the electron acceptor. Organisms carrying out denitrification and ANAMMOX are in direct competition for nitrite and the power struggle between the two is dependent on organic matter availability. Indeed, denitrifiers are generally heterotrophs, requiring an external carbon source whereas ANAMMOX are autotrophs. Consequently, denitrifiers are favoured in organic matter-rich environments where they can outcompete ANAMMOX. Furthermore, the two processes require different nitrogen substrates, with ANAMMOX also using ammonia, whereas denitrification only requires nitrate/nitrite. Percent ANAMMOX should therefore be higher in environments with higher ammonia availability (Damashek and Francis, 2018).

Another fate for nitrate is dissimilatory nitrate reduction to ammonia (DNRA), a nitrate reduction pathway similar to denitrification except that nitrate is reduced to ammonia instead of dinitrogen. Like denitrification, DNRA can be either organotrophic (coupled to organic matter oxidation) or lithotrophic (coupled to inorganic compounds oxidation (Eisenmann et al., 1995; Otte et al., 1999)). DNRA can be favoured over denitrification in anoxic environments with both excess electron donors and limiting nitrate as DNRA can generate a greater number of electrons per molecules of nitrate. It has also been observed that DNRA is enhanced in sulphide-rich environments, either because it can be used as an electron donor or because DNRA organisms have a higher sulphide tolerance compared to denitrifiers (Damashek and Francis, 2018).

Recently, a pathway called Intra-Aerobic Methane Oxidation (coupled with denitrification) (IAMO) was discovered (Ettwig et al., 2010). In this pathway, the anaerobic methane-oxidizing Bacteria *M. oxyfera* couples anaerobic oxidation of methane with the reduction of nitrite to dinitrogen. To do so, the bacterium uses a dismutase to convert two molecules of nitric oxide to dinitrogen and oxygen, the latter being used to oxidize methane under anoxic conditions. Similarly, an anaerobic methanotrophic archaeon (*Candidatus 'Methanoperedens nitroreducens'*) is capable of anaerobic oxidation of methane via reverse methanogenesis using nitrate (ARMN) as the terminal electron acceptor to produce nitrite. An ANAMMOX syntrophic partner then reduces this nitrite to dinitrogen. Comparative genomics shows that the

gene for nitrate reduction (*nar*) in *M.nitroreducens* was acquired via lateral transfer from a bacterial donor (Haroon et al., 2013).

Finally, nitrate can be assimilated into organic matter via Assimilatory Nitrate Reduction (ANR). Understanding the factors controlling the biogeochemical nitrogen cycle is of primary importance as several steps of the nitrogen cycle produce nitrogen oxides, which are greenhouse gases. These steps include nitrification, denitrification and nitrifier denitrification (ND) (Figure 1.1).

In Figure 1.1, a clear separation between the oxic and anoxic zone is represented. In reality, environments typically display a gradual decrease in oxygen, creating sub-oxic zones. Anaerobic processes such as denitrification have been shown to be at the highest in these transition zones in tidal sediments an in fact might not be entirely inhibited even in the oxygen rich zones (Schutte et al., 2018). Similarly, in the ocean, sub-oxic zones are found in between the upper oxygenated water column and the lower oxygen minimum zone (OMZ). These zones of low oxygen concentration can still support ammonia oxidation by aerobic marine archaeal nitrifier, providing nitrite that fuels ANAMMOX (Wakeham, 2020). There is therefore not a clear delimitation between processes occurring in the oxic and anoxic zones. A recent study showed that denitrifiers (*i.e.* "anaerobes") were abundant throughout the oxic, suboxic and anoxic oceanic water column whereas autotrophic ammonia and nitrite oxidizers (*i.e.* "aerobes") were transcriptionally active even at suboxic and anoxic depths, suggesting a tight coupling between nitrogen oxidizing and reducing metabolisms at low-oxygen depths (Suter et al., 2020).

1.2.2 Nitrogen cycle in estuaries

All biogeochemical transformations of nitrogen species presented in Figure 1.1 are found in coastal sediments (Bianchi, 2007). Indeed, as environments displaying both oxic and anoxic compartments, coastal systems can support both oxidation (mainly oxic zone) and reduction (mainly anoxic zones) of nitrogen species. The limit between the oxic and anoxic zone depends on the type of sediment: sandy sediments generally display higher oxygen penetration that muddy ones (Bianchi, 2007). Nitrogen-related microbial processes occurring in coastal systems are of primary importance as they modify the chemical state of nitrogen (organic/inorganic; oxidized/reduced; dissolved in the water phase/bound to sediment/in the gaseous phase) subsequently impacting nitrogen chemistry and bioavailability. These

transformations have an important impact on a local scale for organisms living there but also on a larger scale as they control the amount of nutrient that reach marine waters and control the production of nitrogen gases that affect climate change (Crutzen, 1970; Yung et al., 1976).

In natural environments, the major biogeochemical cycles are linked together via microbial metabolism. Different type of coupling between element cycles can be distinguished: 1) direct assimilatory coupling: elements such as nitrogen, carbon, phosphate etc. are all incorporated into biomass and subsequently released together through mineralisation. 2) Direct dissimilatory coupling: bacteria and archaea possess the most diverse range of metabolism compared to other kingdoms of life. Their energy-releasing metabolisms often involve the change in the oxidation state of different elements (*e.g.* denitrifiers use organic carbon as electron source and nitrate as electron acceptors). Finally, indirectly coupling occurs via the reaction of the products of microbial metabolisms, which can affect other biogeochemical processes (*e.g.* hydrogen sulphide produced by sulphate respiration can inhibit nitrification) (Burgin et al., 2011). As discussed previously, estuaries are complex environments where all major elements are found and are therefore places where the nitrogen cycle is tightly coupled with other major cycles.

1.2.2.a Links with the carbon cycle

The nitrogen cycle is closely coupled to the carbon cycle. Autotrophic nitrification and ANAMMOX are coupled to CO_2 fixation, transforming inorganic carbon into an organic form. Nitrate and nitrite respiration are carried out by heterotrophs. These organisms oxidize organic carbon into CO_2 to generate energy. Nitrate/nitrite are then used as terminal electrons acceptors of the respiratory chain. Methanogenesis, the conversion of simple organic compounds and CO_2 to methane is an important process in coastal sediments that has received great attention, as methane is a potent greenhouse gas (Ferry and Lessner, 2008; Jameson et al., 2019). Methanotrophic organisms can use methane as carbon and energy sources and some use nitrogen species to carry out their metabolisms. As discussed in section 1.2.1, NO is used in the IAMO metabolism to generate oxygen in anoxic zones to oxidize methane. Furthermore, *M.nitroreducen* is capable of direct anaerobic oxidation of methane using nitrate. Nitrogen fixators in symbiosis with plants are also tightly linked to the carbon cycle, as their rate of fixation influences the growth rate of the plant and therefore the CO_2 fixation rate of the autotrophic partner (Schindler and Bayley 1993). The same applies to CO_2 fixation by phytoplankton in coastal water and sediments being limited by nitrogen fixation (Gruber 2004).

Thus, the nitrogen cycle is strongly linked to both fixation and mineralization of carbon (Bolle et al., 1982; Bothe et al., 2007).

Carbon and nitrogen cycles are also indirectly linked as some nitrogen related metabolisms are pH-sensitive, for example, nitrification. An increase in atmospheric CO_2 concentrations leads to an increase in CO_2 concentrations in oceans due to the diffusion of gas in waters. This acidification can result in a lower autotrophic nitrification rate and therefore to a lower CO_2 fixation by nitrifiers.

1.2.2.b Links with the sulphur cycle

Sulphur is an important element in natural ecosystems. Like nitrogen, it is an essential component of living cells as part of some amino acids. Sulphur also changes its redox states, due to abiotic and biotic reactions, many of which are catalysed by microorganisms for their energy yielding metabolisms. The sulphur cycle in marine sediment is dominated by the dissimilatory reduction of sulphate to sulphide (DSR) by anaerobes. The majority of this sulphide is then re-oxidized to sulphate via different biochemical reactions including microbial processes that use oxidized nitrogen species as oxidants (Jørgensen et al., 2019). These Sulphur Oxidizing Microbes (SOM) can perform anaerobic lithotrophic denitrification coupled with the oxidation of sulphide (Baker et al., 2015). One striking example of SOM includes members of the genus *Thioploca* that have been shown to display high motility enabling them to move within the sediment column and even escape into the water column to capture and store nitrate for subsequent use in deeper sediment layers to oxidise sulphide (Schulz and Jørgensen, 2001; Zopfi et al., 2001).

Another example of direct coupling between the nitrogen and sulphur cycles is the SULFAMMOX process, an analogue of the ANAMMOX where nitrite is replaced by sulphate to oxidize ammonia anaerobically. The exact mechanism of SULFAMMOX is not clear and three possible pathways have been proposed: 1) SO_4^{2-} is reduced to S and NH_4^+ is oxidized to N_2; 2) NH_4^+ is oxidized to NO_2^- by SO_4^{2-} and part of this NO_2^- is reduced to N_2 by S^{2-} and finally NO_2^- and NH_4^+ are converted to N_2, like in the conventional ANAMMOX process; 3) NH_4^+ is partially oxidized to NO_2^- by SO_4^{2-} then NO_2^- and NH_4^+ are converted to N_2 (Liu et al., 2021). In coastal sediment, it has been proposed that the sulphate dependant denitrification and SULFAMOX/ ANAMMOX processes are linked: in the case where nitrate is only partially reduced to nitrite, it is used for ANAMMOX. However, when nitrate is completely reduced to

dinitrogen gas, the sulphates originating from sulphide oxidation replaces nitrite inducing a switch from ANAMMOX to SULFAMOX (Rios-Del Toro and Cervantes, 2016; Liu et al., 2021).

1.2.2.c Links with the iron cycle

Iron rapidly changes between its Fe^{2+} and Fe^{3+} oxidation states in the environment. This fluctuation between redox states has been named the "FeII-FeIII redox wheel" and has been shown to interact with the major nutrient cycles including carbon, nitrogen, phosphorus and sulphur in numerous environments including coastal sediments (Burgin et al., 2011; Li et al., 2012). As previously described for the sulphur cycle, lithotrophic denitrifying microorganisms can directly link the nitrogen and iron cycles by using Fe^{2+} as an electron source to reduce nitrate, generating nitrite, nitrous oxide or dinitrogen and Fe^{3+}(Burgin et al., 2011; Li et al., 2012). Similarly, Fe^{2+} can be used as an electron source for lithotrophic DNRA. Incubations of estuarine sediments with added Fe^{2+} has been shown to enhance DNRA and reduce denitrification, indicating that Fe^{2+} availability may be a significant factor controlling the fate of nitrogen: either loss via denitrification or retention via DNRA. Another important Fe^{3+} generating mechanism is the aerobic iron oxidation carried out in upper sediment layers by autotrophic microorganisms (Burgin et al., 2011).

The second part of the FeII-FeIII redox wheel is the reduction Fe^{3+} to Fe^{2+}. In marine sediment, heterotrophic anaerobes can replace oxygen by Fe^{3+} as the terminal electron acceptor to oxidize organic carbon, generating Fe^{2+} (Burgin et al., 2011; Li et al., 2012). Another iron-reducing pathway is an analogue to ANAMMOX, but replaces nitrite with Fe^{3+} as an oxidant. This process, called FEAMMOX, has been shown to play an important role in nitrogen loss in intertidal wetlands by oxidation of the soluble ammonia to dinitrogen gas (Li et al., 2015).

1.2.2.d links with the manganese cycle

Manganese oxide (MnO_2) is another potential electron acceptor that is present in marine sediments and the anaerobic oxidation of ammonia by MnO_2, leading to either nitrate or dinitrogen, is thermodynamically favourable (Hulth et al., 2005). Incubation experiments and strain isolation have revealed that anaerobic oxidation of ammonia to nitrate with MnO_2 is a relevant process in marine sediments (Hulth et al., 2005; Bartlett et al., 2008; Javanaud et al., 2011; Lin and Taillefert, 2014). This process, however, might be only significant in Mn-rich sediments and when sulphate reduction is absent (Lin and Taillefert, 2014) which may apply

to only a minority of coastal marine ecosystems and be more relevant to deep-sea sediments (Mogollón et al., 2016).

A schematic representation of the major interactions between the nitrogen cycle and other major biochemical cycles is presented in Figure 1.2. This figure represents a simplified situation with a gradient of concentration of the major electrons acceptors: because of the decrease in energy yield obtained by respiration from O_2 to NO_3^- to Fe^{3+} and then SO_4^{2-}, these oxidants are successively depleted in the sediment column from top to bottom. Picks of NO_3^- and SO_4^{2-} are due to microbial nitrification and sulfur oxidation, respectively.

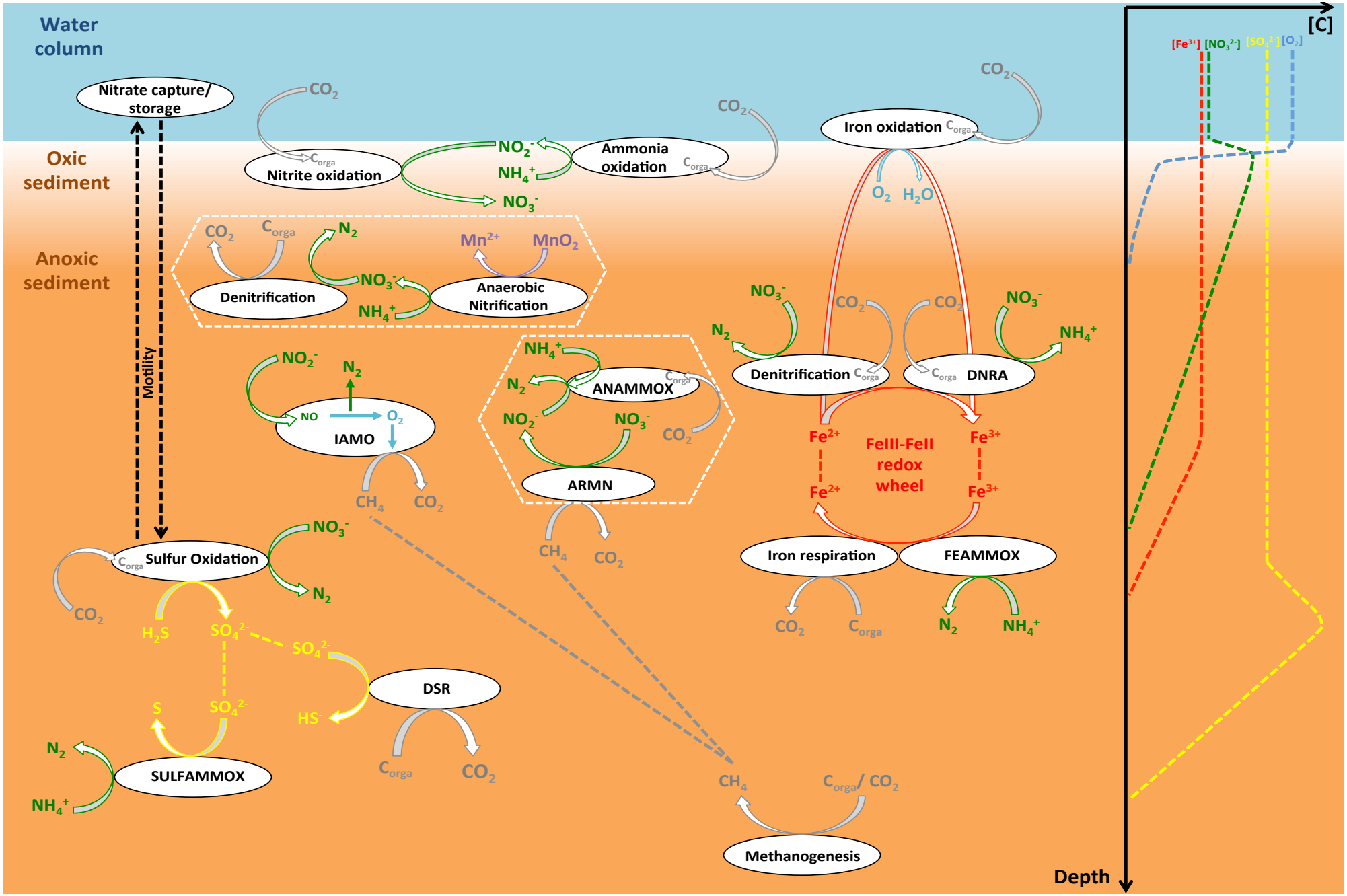

Water column
Oxic sediment
Anoxic sediment
Motility
Nitrate capture/ storage
Nitrite oxidation
Ammonia oxidation
Iron oxidation
Denitrification
Anaerobic Nitrification
IAMO
ANAMMOX
ARMN
Denitrification
DNRA
FeIII-FeII redox wheel
Iron respiration
FEAMMOX
Sulfur Oxidation
DSR
SULFAMMOX
Methanogenesis
[C]
Depth

Figure 1.2 Schematic representation of the major interactions between the nitrogen cycle and the carbon, iron, sulfur and manganese cycles in marine sediments. Microbial processes are indicated in white bubbles showing how they couple the nitrogen cycle (green arrows), carbon cycle (grey arrows), iron cycle (red arrows), sulfur cycle (yellow arrows) and manganese cycle (purple arrows). Dashed lines represent transport of elements. Microbial processes drawn in a white dashed hexagon are actively coupled *i.e.* the product of one metabolism feeds the other. Concentration profiles of the major electron acceptors along the sediment depth are represented by dashed line-plots on the right of the figure. C_{orga}: carbon stored in organic matter. DSR: dissimilatory sulphate reduction. IAMO/ARMN: see 1.2.1.

1.2.3 Production of nitrous oxide by nitrogen cycle-related organisms in coastal sediments

Nitrous oxide (N_2O) is a major atmospheric pollutant and contributes significantly to global climate change (Crutzen, 1970; Yung et al., 1976). Nitrous oxide has a global warming potential (GWP) approximately 300 times higher than carbon dioxide (Table. 1.1). GWP is a measure of a greenhouse gas effect on climate and depends on a) its atmospheric concentration, b) how strongly and where in the infrared spectrum it absorbs energy and c) its lifetime in the atmosphere (CO_2 as a reference with GWP=1) (Elrod, 1999). Emissions from soils, sediments and waters are due to both biotic and abiotic processes. The major biotic processes are driven by microorganisms and among them, the most prominent are nitrification, denitrification and nitrifier-denitrification (Hu et al., 2015).

Denitrification-derived N_2O is generally recognised as the major biotic source in marine sediment. The importance of this N_2O production is dependent on the amount of dissolved inorganic nitrogen (DIN) and oxygen concentration (Murray et al., 2015). The expected increase in nitrogen loadings and the resulting eutrophication of coastal environments could result in an overall increase in N_2O production in marine sediments in future (Naqvi et al., 2010). As discussed in 1.2.1, nitrification controls the amount of nitrite and nitrate produced, which then feeds denitrification. Nitrification is therefore important for its direct and indirect link to N_2O production in coastal sediments.

Table 1.1 Global Warming Potentials (GWP) and atmospheric lifetimes (years) of some of the major greenhouse gases. Source: U.S. Environmental Protection Agency 2002.

Gas	Atmospheric lifetime	GWP (over 100 year)
Carbon dioxide	50-200	1
Methane	12±3	21
Nitrous oxide	120	310
Hydrofluorocarbons	1.5-209	140-11,700
Sulfur hexafluoride	3,200	16,300

1.2.3.a Nitrification

During the nitrification process, chemical reactions can occur with intermediates between NH_3 and NO_2^- (and NO_2^- itself) leading to the production of N_2O. Incomplete oxidation of hydroxylamine is also a source of N_2O in AOB. AOA also produce N_2O but the mechanism is unclear. Indeed, the intermediate redox forms between NH_3 and NO_2^- during AOA ammonia oxidation are uncertain. A better understanding of AOA's metabolism is therefore crucial as AOA are believed to be largely responsible for oceanic N_2O production (Hu et al., 2015). In agricultural soils, the relative contribution of AOA and AOB to N_2O production is linked to the source of nitrogen: under unamended conditions, where ammonia mainly originated from mineralisation, N_2O production is lower and is mainly due to AOA; on the other hand, when high amounts of inorganic ammonia are supplied, N_2O production is higher and is mainly due to AOB activity (Hink et al., 2017).

1.2.3.b Denitrification

Denitrification is a chemo-heterotrophic or chemo-autotrophic anaerobic respiratory metabolism in which nitrate is used as a terminal electron acceptor. Nitrogen oxides are formed during the sequential reduction of nitrate to dinitrogen and can be released into the atmosphere (Pilegaard, 2013; Wrage et al., 2001):

$$NO_3^- \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O \rightarrow N_2$$

Another pathway called nitrification-coupled denitrification (a combination of autotrophic nitrification and heterotrophic denitrification) is responsible for the production of nitrogen oxides; it takes place when denitrifiers directly reduce nitrite and nitrate produced by

autotrophic nitrifiers. The mechanisms responsible for N_2O production are the same as the ones involved in nitrification and denitrification (Pilegaard, 2013; Wrage et al., 2001).

1.2.3.c Chemodenitrification

The abiotic reactions of nitrite in soils and sediments can also be a source of nitrogen oxides (NO_x). Chemical decomposition of nitrite can be important in environments with low pH and include the reactions with protons (a, b), ferrous iron (c) and hydroxylamine (d) (Pilegaard, 2013; Wrage et al., 2001).

(a) $3NO_2^- + 2H^+ \rightarrow 2NO + NO_3^- + H_2O$

(b) $NO_2^- + H^+ \rightarrow HNO_2 \rightarrow HONO$

(c) $NO_2^- + Fe^{2+} + 2H^+ \rightarrow NO + Fe^{3+} + H_2O$

(d) $NH_2OH + NO_2^- + H^+ \rightarrow N_2O + 2H_2O$

1.2.3.d Nitrifier denitrification

Nitrifier denitrification (ND) is a pathway carried out by autotrophic ammonia oxidizers in which the oxidation of ammonia is followed by the reduction of nitrite to nitrous oxide and eventually dinitrogen (Wrage et al., 2001). This pathway is not coupled to the oxidation of organic compounds and thus differs from classical heterotrophic denitrification (Stein, 2011). Using pure cultures (and cell-free extracts), of *Nitrosomonas europaea* Ritchie and Nicholas, (1972) found that N_2O was produced by reduction of nitrite with hydroxylamine as an electron donor under both oxic and anoxic conditions. Poth and Focht, (1985) also found that *Nitrosomonas europaea* was able to produce N_2O via reduction of nitrite but only under oxygen-limited conditions. ND is now believed to be a common trait in ammonia-oxidizing β-proteobacteria (Shaw et al., 2006) and a major contributor to N_2O production in soils and sediments (Baggs, 2008; Firestone and Davidson, 1989; Kool et al., 2011; Muller et al., 2014; Poth and Focht, 1985; Venterea, 2007; Webster and Hopkins, 1996; Wrage et al., 2005; Wrage et al., 2004; Wrage et al., 2001; Zhu et al., 2013). Factors controlling N_2O production by ND are still unclear, but it seems that soil moisture/O_2 availability is an important one (the two are related: when moisture increases, O_2 availability generally decreases). Using a combined ^{18}O and ^{15}N tracing method (developed by Wrage et al. (2005)) to differentiate the sources of N_2O production from the soil, Zhu et al. (2013) and Kool et al. (2011) found that ND was the major source of N_2O at low O_2 concentration (0.5% and 3%) and high moisture (50 and 70% WFPS (Water Fill Pore Space)); heterotrophic denitrification being the major source at lower O_2

concentration (0%) and higher WFPS (90%). Despite differences in measurement, both studies showed that ND could be a major contributor to N_2O production from the soil where O_2 availability is low. WFPS also influences nutrients diffusion and transport. Its influence on N_2O production may involve more complex mechanisms than only O_2 availability (Hu et al., 2015). Other soil parameters such as carbon and nitrogen content, pH, temperature (Wrage et al., 2001) and trance metals availability (Glass and Orphan 2012) are key factors regulating ND and N_2O production.

Also unclear is the reason for autotrophic nitrifiers to carried out ND. Oxidation of ammonia by nitrite could be used to conserve oxygen from the initial step and be linked to energy production, as the energy gained when oxidizing ammonia with dioxygen or nitrite as an electron acceptor is similar (Kool et al., 2011; Poth and Focht, 1985; Wrage et al., 2001). ND could also be used to eliminate nitrite, which is toxic for cells (Beaumont et al., 2004; Kool et al., 2011; Poth and Focht, 1985; Stein and Arp, 1998; Wrage et al., 2001). Other possible explanations for the occurrence of ND include the possible decreasing competition for oxygen by consuming the substrate of nitrite oxidizers (Poth and Focht 1985) and the production of N_2 as a mean of transport (upward motion) for nitrifiers out of the oxygen-depleted sediment to reach the water column (Philips et al., 2002). Finally, at least in *N. europaea*, ND could serve as an electron sink to speed the oxidation of hydroxylamine during aerobic metabolism (Stein, 2011).

In AOB, at least two enzymes seem to be involved in nitrifier denitrification: a nitrite reductase (Nir) and a nitric oxide reductase (Nor). The first one is believed to reduce nitrite to nitric oxide and the second one to reduce nitric oxide to nitrous oxide (Garbeva et al., 2007; Kozlowski et al., 2014). Most AOB possesses genes encoding *nirK* and *nor* (B and/or Y) in their genomes. It has been shown that NorB but not NirK is required for the production of N_2O by *N.europaea*, suggesting that there could be an alternate (unidentified) nitrite reductase responsible for this activity (Kozlowski et al., 2016; Kozlowski et al., 2014). In the new NH_3 oxidation pathway proposed by Caranto and Lancaster (2017) NO is an obligate intermediate produced by the oxidation of NH_2OH by the HAO enzyme. The authors proposed that at low O_2 concentrations, the rate of NO oxidation to NO_2^- is slower than its rate of production. The resulting accumulation of NO would lead to NO emission or its conversion to N_2O. The production of N_2O from AOB in oxygen limiting conditions could therefore have a double origin: reduction of NO_2^- and incomplete oxidation of NH_2OH.

Studies of metagenomes (Bartossek et al., 2010; Treusch et al., 2005), metatranscriptomes (Hollibaugh et al., 2011) and sequencing of genomes from pure cultures (Blainey et al., 2011; Walker et al., 2010) have revealed the presence of copper-containing nitrite reductases genes in AOA. However, none of the other genes required to carry out the nitrifier denitrification pathway (*e.g. nor* genes) has been characterized (Lund et al., 2012). Consequently, the ability of AOA to carry out ND is still controversial (Jung et al. 2014 vs. Stieglmeier et al. 2014). Regardless of the biochemical pathway used, pure and enrichment cultures of AOA do produce nitrogen oxides (Jung et al., 2011; Jung et al., 2014; Stieglmeier et al., 2014). And AOA are now designated as the main organisms responsible for N_2O production in the ocean (Löscher et al., 2012; Santoro et al., 2011). In terrestrial environments, AOA could also be responsible for N_2O production, especially in low-ammonia conditions, such as unfertilized or acidic soils. In agricultural land, however, where high amounts of inorganic nitrogen fertilizers are applied, AOB remains the main N_2O producer (Hink et al., 2016).

1.3 Nitrification

The step that transforms ammonia, the most reduced form of nitrogen to nitrate, the most oxidized form, is called nitrification. In natural environments, nitrification is mainly a biological process (Stein and Klotz 2011), catalysed by three functionally distinct groups of organisms: the ammonia oxidizers that oxidize ammonia to nitrite, the nitrite oxidizers that oxidize nitrite to nitrate and the complete ammonia oxidizers that can oxidize both ammonia to nitrite and nitrite nitrate. Although nitrification does not directly change the total amount of nitrogen in a given ecosystem, the oxidation of ammonia to nitrate often results in the loss of total nitrogen. Indeed, ammonia generally remains bound to soil particles whereas nitrate is more easily dissolved into ground and surface water and can be washed away. Nitrification can therefore be the cause of important input of nitrate into fresh and seawater causing eutrophication (Bianchi 2007; Howarth and Marino 2006). Plants generally depend on soil nitrogen, as they are unable to directly use atmospheric N_2. Some are better able to incorporate ammonia than nitrate (and vice versa). Nitrification, therefore, has a major impact on plants and is of central importance in agriculture. Nitrification has another important impact on soils: acidification. The production of H^+ by this process can cause a decrease in pH in non-buffered soils resulting for example in the mobilisation of toxic metals. Nitrification is also a key step in some wastewater treatments to eliminate excess nitrogen via feeding of denitrifiers. Finally,

nitrification has been shown as an important source of the greenhouse gas N_2O contributing to global climate change (Stein and Klotz 2011).

1.3.1 Ammonia oxidation

1.3.1.a Autotrophic ammonia oxidizers

i) Diversity of Ammonia Oxidizing Microorganisms (AOM)

Autotrophic ammonia oxidizers are represented by both Bacteria (AOB) and Archaea (AOA). Bacterial ammonia oxidation was discovered long before Archaeal ammonia oxidation. Biological oxidation of ammonia was first discovered by Houzeau (1872) and Muller (1873) and further investigated by Warington (1878 and 1891). The first isolation of an AOB was achieved by Frankland and Frankland (1890). Initially, AOB were classified based on morphological criteria (arrangement of internal membranes, cell shape and size, flagella, etc.) (Monteiro et al., 2014) but subsequently, analyses based on *16S rRNA* and *amoA* gene sequences have been used to resolve the phylogeny of this group: there are two monophyletic lineages among proteobacteria: one affiliated to the subdivision β-proteobacteria (β-AOB) and includes genera *Nitrosospira*, *Nitrosomonas* and one representative of *Nitrosococcus* (*N. mobilis*). The other is placed within the γ-proteobacteria subdivision (γ-AOB) and includes representatives of the *Nitrosococcus* genera (*Nitrosococcus oceani*, *Nitrosococcus halophilus* and *Nitrosococcus watsonii*) (Monteiro et al., 2014; Prosser, 2005; Robertson and Groffman 2015; Campbell et al., 2011b) and the newly discovered *Nitrosoglobus* (Hayatsu et al., 2017).

The discovery of AOA is more recent. The first clues of their existence came from metagenomic studies where genes related to ammonia oxidation were found in DNA libraries from soil (Treusch et al., 2005) and seawater (Treusch et al., 2005; Venter et al., 2004). AOA were first isolated by Könneke et al. (2005) from a seawater aquarium. The isolation of AOA allowed whole genome sequencing which provided further insights into the nitrogen metabolism of AOA (Blainey et al., 2011; Walker et al., 2010). These discoveries lead to a rearrangement of archaeal phylogenetic classification with the creation of a third phylum containing all AOA, the Thaumarchaeota (Brochier-Armanet et al., 2008). All AOA are now included in the class *Nitrososphaeria* of this phylum, with a proposed division into 4 (or 5) order-level lineage, based on the phylogeny of the *amoA* gene: *Ca. Nitrosocaldales*, *Nitrososphaerales*, *Ca. Nitrosotaleales*, *Nitrosopumilales*, the fifth group *Nitrosotaleales* or *Nitrosopumilales-Incertae sedis* could not be assigned to *Nitrosotaleales* or *Nitrosopumilales* (Alves et al., 2018).

ii) Biochemistry of autotrophic ammonia oxidation

AOB and AOA are chemolithoautotrophic microorganisms that oxidize ammonia using dioxygen to produce energy and reducing power to fix CO_2 (Ward, 2013).

In AOB, nitritation starts with the exergonic oxidation of ammonia with oxygen, producing hydroxylamine (1). This is catalysed by the ammonia monooxygenase (Amo), a membrane-bound enzyme composed of three subunits (Amo A, B and C) (Arp et al., 2002; González-Cabaleiro et al., 2019):

$$NH_3 + O_2 + 2H^+ + 2e^- \rightarrow NH_2OH + H_2O; \qquad \Delta G^{01} = -170.49 \text{ kJ/mol } NH_3 \qquad (1)$$

The 2 electrons necessary for the reaction are thought to be provided by ubiquinol, however, this hypothesis is still to be verified (González-Cabaleiro et al., 2019). While one O from O_2 is used to form NH_2OH, the other one is reduced to H_2O (Arp et al., 2002) in an exergonic reaction (2). The final step is the endergonic oxidation of hydroxylamine to nitrite using water (3) (González-Cabaleiro et al., 2019). This reaction is catalysed in AOB by the hydroxylamine oxidoreductase (HAO), a homotrimer periplasmic enzyme (each subunits containing eight c-types hemes) (Arp et al., 2002; Robertson and Groffman, 2015). Overall, the oxidation of ammonia to nitrite should result in a highly exergonic reaction (4).

$$0.5O_2 + 2H^+ + 2e^- \rightarrow H_2O; \qquad \Delta G^{01} = -165.46 \text{ kJ/mol } NH_3 \qquad (2)$$

$$NH_2OH + H_2O \rightarrow NO_2^- + 5H^+ + 4e^-; \qquad \Delta G^{01} = 28.60 \text{ kJ/mol } NO_2^- \qquad (3)$$

$$NH_3 + 1.5O_2 \rightarrow NO_2^- + H_2O + H^+; \qquad \Delta G^{01} = -307.35 \text{ kJ/mol } NH_3 \qquad (4)$$

Recently, Caranto and Lancaster (2017) proposed a revised pathway in which ammonia is oxidized in a three-steps process to nitrite, with NO as an obligate intermediate. In this model, HAO oxidizes NH_2OH to NO, which is then further oxidized to NO_2^- by a NO-oxidoreductase (possibly NirK). The latest findings and remaining questions regarding NH_3 oxidation by AOM were recently reviewed (Lancaster et al., 2018).

Regardless of the number of steps required, oxidation of NH_3 to NO_2^- results in an electron flow which is then channelled through cytochrome c554 to cytochrome cm552 (Arp et al.,

2002; González-Cabaleiro et al., 2019): of the four electrons produced by hydroxylamine oxidation (3) 2 return to Amo via ubiquinol (1) (Prosser, 2005; González-Cabaleiro et al., 2019). On average, 1.65 of the 2 remaining electrons pass through the terminal oxidase, using ½ O_2 as terminal electron acceptor (2). The other 0.35 electrons are used to reduce NAD^+ to NADH via a reverse electron transfer catalysed by an NADH dehydrogenase (Figure 1.3). Electrons for reaction (2) and the reduction of NAD^+ to NADH are also provided by ubiquinol. The reducing power (NADH) is used for autotrophic fixation of HCO_3^- into organic matter (González-Cabaleiro et al., 2019). The energy for ATP synthesis comes from the translocation of protons from the cytoplasm to the periplasm by the terminal oxidase: The complex III (ubiquinol-cytochrome c reductase) and complex IV (cytochrome c oxidase) likely translocate four and two protons, respectively, into the periplasm for every two electrons initially provided by ubiquinol. On the other hand, the protons generated by hydroxylamine oxidation in the periplasm might not contribute to the proton gradient as they are consumed for the reduction of ubiquinone to ubiquinol. Interestingly, only reaction (2) has been proven to be coupled with proton translocation and the highly exergonic energy of the oxidation of ammonia to hydroxylamine (1) could be fully dissipated as heat (González-Cabaleiro et al., 2019).

It has also been demonstrated that Amo can oxidize different types of organic substrates, transforming C-H bound to alcohol, C=C bound to epoxides and sulphides to sulfoxides (Arp et al., 2002; Sayavedra-Soto et al., 2010). This secondary activity is attributed to the low selectivity of Amo and, according to the current biochemical understanding of AOB, might not be linked to any energy harvesting system (González-Cabaleiro et al., 2019).

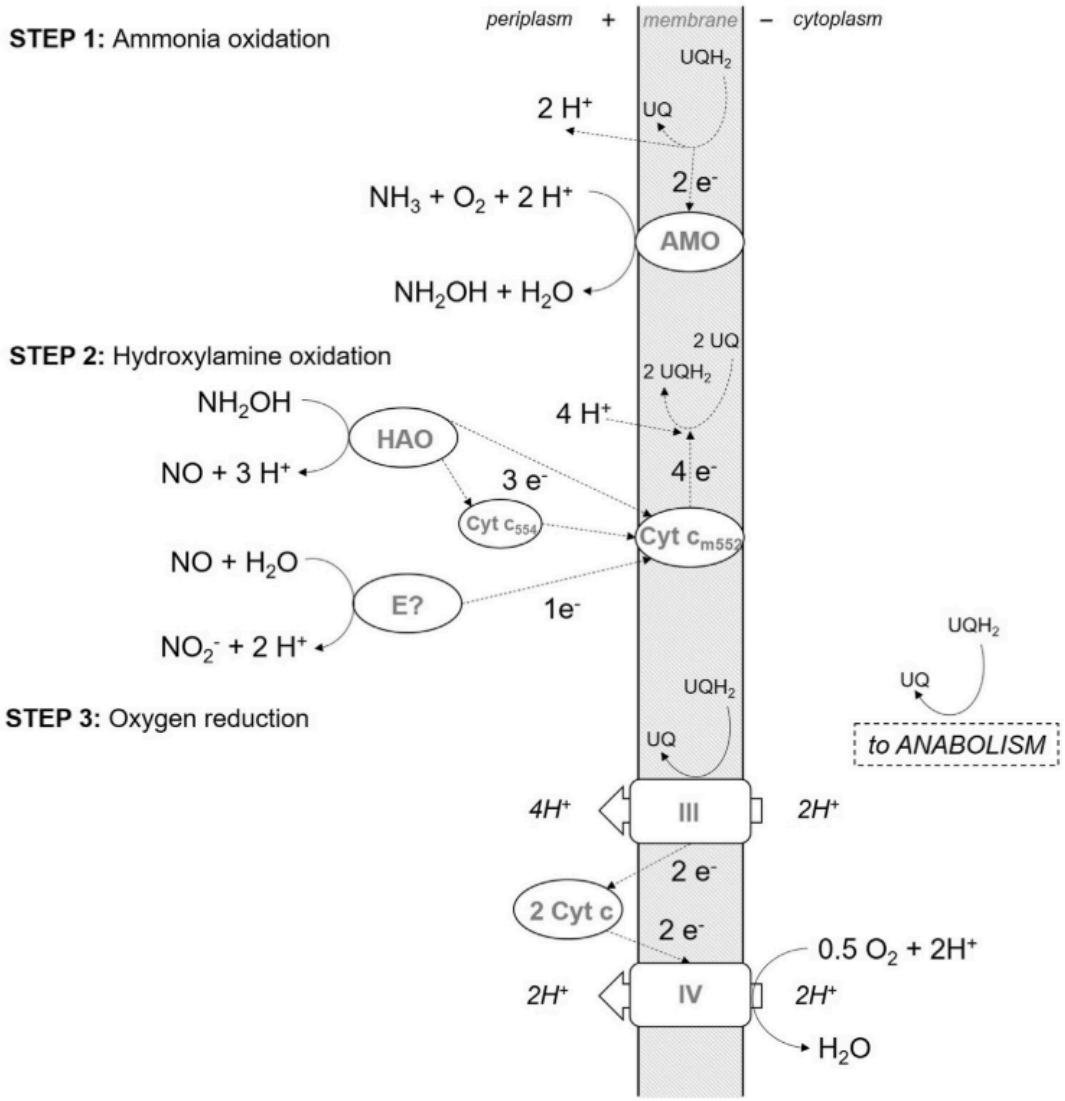


Figure 1.3 Catabolic process of AOB. Biochemical reactions and electron flow are represented by curved and straight arrows, respectively. AMO: Ammonia monooxygenase; HAO: Hydroxylamine oxidoreductase; E?: putative NO-oxidoreductase; Cyt c: Cytochrome c; Cyt c m552 = membrane cytochrome 552. NB: ATP synthase is not represented. Cyt c554: cytochrome c554. III: complex III (ubiquinol-cytochrome c reductase); IV: complex IV (cytochrome c oxidase). Copied from (González-Cabaleiro *et al.*, 2019).

An alternative pathway for ammonia oxidation involves nitrogen dioxide instead of dioxygen as an oxidant of ammonia under both oxic and anoxic conditions (Figure 1.4) (Kampschreur et al., 2006; Schmidt et al., 2001). This theory is based on the study of the capacity of *Nitrosomonas* to oxidize ammonia under different culture conditions. Authors suggest that in both oxic and anoxic conditions, *Nitrosomonas*-like organisms oxidise ammonia using nitrogen tetraoxide (dimer of nitrogen dioxide). This reaction is catalysed by Amo and produces hydroxylamine and NO. Then, hydroxylamine is further oxidised to nitrite by Hao.

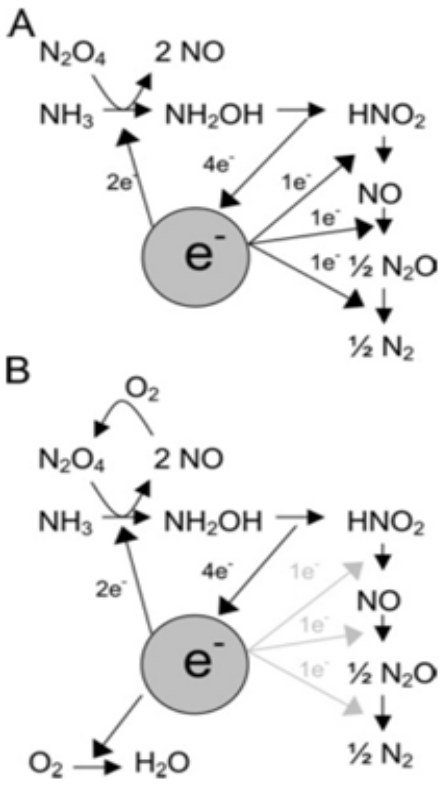


Figure 1.4 Alternative pathways for ammonia oxidation in *Nitrosomonas.* Under oxic conditions (B) O_2 is used to re-oxidise 2NO into N_2O_4, the oxidizing agent and as terminal electron acceptor. Under anoxic conditions (A), nitrite is partially used as electron acceptor leading to the production of N_2. Under oxic conditions with low NO_x concentrations, nitrite is also used as electron acceptor for ammonia oxidation leading to the production of N_2 and N_2O (traces). Copied from Kampschreur *et al*. (2006).

The authors cited the following observations to support this model:

1) *Nitrosomonas*-like organisms are inhibited when gaseous NO was removed from the cultures by intensive aeration; nitrification started again when NO was added to the gas.
2) The lag phase during the recovery of ammonia oxidation in starved cells can be reduced when NO_x as added.
3) Cells produce NO via denitrification when external NO_x is not available.
4) NO_x addition to pure cultures of *N. eutropha* resulted in an increase in nitrification rate, cell density and aerobic denitrification capacity.
5) Different species of ammonia oxidizers (e.g., *Nitrosomonas europaea*, *Nitrosolobus multiformis*) were able to oxidize ammonia under anoxic conditions when NO_2 was present, and the aerobic ammonia oxidation activity was increased in the presence of NO and NO_2.

Interestingly, Schmidt et al. (2001) showed that acetylene, which blocks Amo activity by covalent binding, didn't have any effect on anaerobic NO_2-dependant ammonia oxidation by *N. europaea*. Furthermore, they showed that both the inhibition and the labelling of AMO by acetylene is O_2 dependant in *N. europaea*.

In AOA, there is no evidence of the presence of HAO, which implies that either another enzyme is responsible for hydroxylamine oxidation or that another pathway is used. Walker et al., (2010) proposed a pathway that involves nitroxyl hydride (HNO) but Vajrala et al., (2013) showed the production and consumption of hydroxylamine during the oxidation of ammonia to nitrite by *N. maritimus*. Later, the observation that NO is essential to NH_3 oxidation in AOA lead to a proposed mechanism where both NH_2OH and NO are oxidized by a putative "Cu-HAO" to form 2 molecules of NO_2^- (Figure 1.5) (Kozlowski et al., 2016). One possible candidate for the production of NO in AOA is the copper-containing nitrite reductase NirK, which is widely distributed in marine and soil AOA. However, no *nirK* homolog was found in genomes of the thermophilic *Ca. Nitrosocaldales* despite the observation that NO is essential for its ammonia-oxidizing metabolism. The role of NirK in the nitrification pathway of AOA therefore remains unclear (Qin et al., 2020). Alternatively, this dependency upon NO could indicate that, as for the revised pathway in AOB, AOA oxidize NH_3 in a three steps process, with NO as an obligate intermediate (Lancaster et al., 2018).

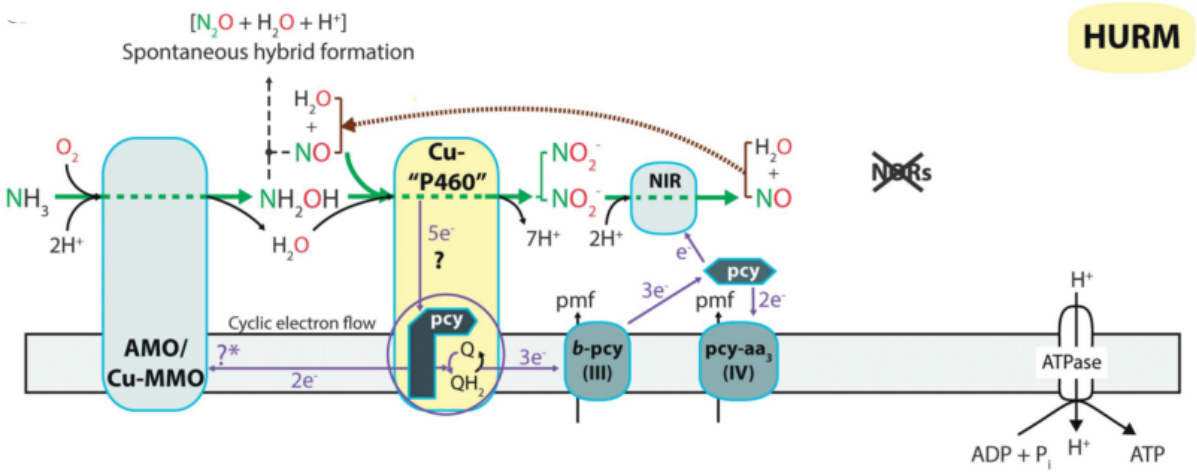


Figure 1.5 Proposed pathway for NH_3 oxidation in AOA. This revised pathway highlights the role of NO as a co-substrate for NH_2OH oxidation and its contribution to N_2O production. AMO/Cu-MMO, ammonia monooxygenase; c552, cytochrome c redox carrier; CytS: cytochrome c'-beta; HAO, hydroxylamine dehydrogenase; HCO, heme-copper oxidase; HURM, hydroxylamine:ubiquinone redox module; NirK, Cu-containing NO-forming nitrite

reductase; NOR, nitric oxide reductase; P460, tetraheme cytochrome c protein P460 (CytL); pcy, plastocyanin; pmf, proton-motive force; Q/QH2, quinone/quinol pool. Copied from Kozlowski *et al*. 2016.

Finally, in both AOA and AOB, NADH and ATP generated by ammonia oxidation are used to fix CO_2 into organic matter. In AOB, this is carried out via the Calvin-Benson-Bassham cycle (CBB) (Ward, 2013) whereas AOA use a modified version of the hydroxypropionate/hydroxybutyrate (HP/HB) cycle of the Crenarchaeota, which is more efficient than any other metabolism for aerobic carbon fixation (Könneke et al., 2014).

iii) Kinetic considerations of autotrophic ammonia oxidation

The high abundance of AOA in the ocean can be surprising considering the low ammonia concentration in seawater (≈10 nM). The key feature of the AOA metabolism that enables them to survive in such oligotrophic environments is the high affinity of their Amo enzyme for ammonia. Martens-Habbena et al. (2009) showed that the half-saturation constant of the mesophilic AOA *'Candidatus Nitrosopumilus maritimus'* strain SCM1 is the lowest of all AOM studied (K_m=133nM total ammonium) which closely matches the kinetics of in situ nitrification in the ocean. Because the first cultivated AOA had a much higher affinity for ammonia than AOB (lower K_m), it was first assumed that AOA would always outcompete AOB in low ammonia environments whereas AOB would be more competitive when ammonia is not limiting. However, this depends on the specific growth rate (μ_{max}: the amount of biomass produced per unit of time) of AOA and AOB: as explained in Prosser and Nicol (2012), AOB (with low ammonia affinity/high K_m) will outcompete AOA in high ammonia environments only if they display a higher specific growth rate. Otherwise, AOA will be more competitive, regardless of the ammonia concentration. Since the study of AOA isolated from soils show that they possess a higher μ_{max} than AOB, the dominance of AOB in most eutrophic environments cannot be explained by substrate affinity alone (Prosser and Nicol 2012). Furthermore, recent studies indicate similar K_m for some *Nitrosomonas* AOB and some soil AOA (Kits et al., 2017; Hink et al., 2017).

iv) Genetic basis of autotrophic ammonia oxidation amo genes

***amoA* genes**: The Amo protein is composed of at least three polypeptides (Amo A, B and C), encoded by three contiguous genes (amoA, amoB and amoC respectively). The three *amo* genes are regrouped into a single operon in the *amoCAB* order in AOB and the number of

operons varies from 1 to 3 copies (Norton et al., 2002; Lehtovirta-Morley, 2018) (Figure 1.6). Two copies of *amoA* are found in the genome of *N. europaea* (99% identity between each other). The genome of *N. europaea* also contains a third copy of *amoC* (≈60% similarity with the other two). Interestingly, it was found by Alzerreca et al., (1999) that the amino acid sequence of the Amo protein from *Nitrosococcus oceani* (a γ-AOB) is more similar to the pMMO found in γ-proteobacterial methanotrophs than to the Amo of β-AOB. This observation suggests that the *amo* and *pmmo* genes may have evolved from an ancestral gene cluster before the divergence of α proteobacteria from the γ and β proteobacteria. Recently, new *amo* genes were discovered: the *amoD*, which encodes a potential membrane protein and a likely duplicated orthologue *amoE*. Their function(s) is still unknown but, interestingly, they differ both in number and regulation between γ-AOB and β-AOB (Simon and Klotz 2013). In *N. europaea*, the *amoC*, *A* and *B* genes are co-transcribed into a single mRNA *amoCAB*, but *amoC* and *amoAB* mRNA can also be detected possibly resulting from the processing of the *amoCAB* mRNA or differential transcription start positions (Sayavedra-Soto et al., 1998). The transcription of *amoAB* and *amoCAB* is triggered by ammonia, even in the presence of Amo inhibitor (Sayavedra-Soto et al., 1996).

In most AOA, a single copy of *amoA*, *amoB*, *amoC* and *amoX* is generally found, grouped together in the genome in the *amoAXCB* arrangement but in some cases (e.g. *Nitrososphaera viennensis*) multiple open reading frames are interjected between the *amo* genes (Walker et al., 2010; Zhalnina et al., 2014; Berg et al., 2015; Lehtovirta-Morley, 2018). Interestingly, the N-terminus of amoC and C-terminus of *amoB* are truncated in archaea compared to bacteria. The *amoX* reading frame has been found in all genomes of sequenced AOA, adjacent to *amoA*, possibly encoding an additional subunit of the Amo enzyme. Other AOA have multiple copies of isolated *amoA* or *amoC* in their genomes and some *Nitrososphaerales* species possess extra copies of *amoC* homologs, possibly acting as chaperons to maintain the structural integrity of the Amo enzyme in low-nutrient environments (Qin et al., 2020) (Figure 1.6). Herbold et al. (2017) showed that the second *amoA* copy in *Candidatus Nitrosotalea okcheonensis* genome, located >400 kb upstream from the *amoAXCB* gene cluster, has an unclear role in the AOA's metabolism as it is not transcribed during growth in batch culture. *Candidatus Nitrosopumilus piranensis DC3* possess a second *amoB* (*amoB2*) copy, which is highly divergent from the *amoB* copy in the *amoAXCB* cluster (48% amino-acid identity) and might have an additional catalytic function, potentially enabling further metabolic versatility to the marine AOA (Bayer et al., 2016).

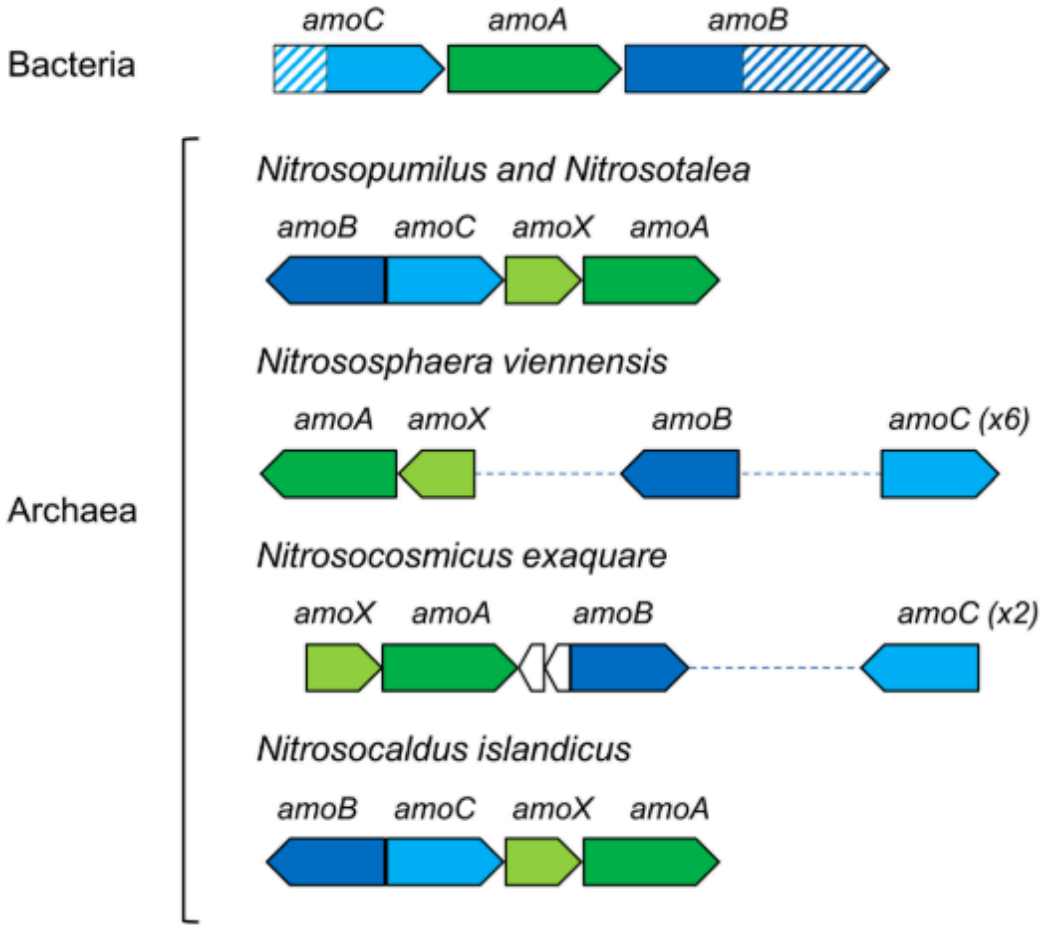


Figure 1.6 Organisation of the *amo* genes in AOB and AOA. The N- terminus of *amoC* and C-terminus of *amoB* are truncated in archaea (indicated by stripes in bacteria). Copied from Lehtovirta-Morley (2018).

***hao* genes**: Hao is encoded by a 1,710 bp gene in *N. europaea*, and present in three copies widely separated in the genome. They are identical to each other except for one nucleotide difference in one gene copy (Arp et al., 2002). As for *amo*, the number of copies varies among AOB, for example, only one copy is found in *Nitrosococcus oceani* (Arp et al., 2002). The study of the upstream regions of this gene in *N. europaea* revealed that they are nearly identical for two copies whereas the sequence of the third diverged shortly upstream of the start codon (Hommes et al., 2001), suggesting potential differences in the transcription regulation systems. Hao is a homologue of the Hzo enzyme of ANAMMOX organisms, both belonging to the multiheme cytochrome c (MCC) superfamily, which also includes NrfA (Simon and Klotz 2013). *hao* mRNA is monocistronic (Arp et al., 2002) and expressed upon ammonia addition, however to a lower extent than *amo* (Sayavedra-Soto et al., 1996).

Carbon assimilation genes: Utåker et al. (2002) compared the rubisCO amino acid sequences of 14 AOB strains. They found that the rubisco from the β subgroup AOB showed a high level of identity to the one found in *R. eutropha* indicating that these bacteria have red-like type rubisco. The only exception was *Nitrosomonas europaea* that apparently possess a green-like *rubisCO*. Interestingly, the sequence of the *rubisCO* in *Nitrosomonas europaea* is closer to the one in *Nitrosococcus halophilus* (γ-proteobacteria also containing a green-like *rubisCO*) than the sequences in other β-proteobacteria.

As mentioned before, AOA use a modified version of the HP/HB cycle that allows for a very efficient CO_2 assimilation. One of the characteristics that make it different is the fact that two key enzymes of the cycle (3-hydroxypropionate synthase and 4-hydroxybutyrate synthase) are ADP forming enzymes whereas in crenarchaeota they are AMP forming. One "loop" of the thaumarchaeal HP/HB cycle, therefore, requires two fewer energy bonds compared to the crenarchaeal one. Furthermore, some of the enzymes are able to catalyse multiple steps of the cycle, therefore reducing the cost of biosynthesis (Könneke et al., 2014).

Other genes: Cytochromes involved in electron transfer from HAO (cytochrome c554 and cm552) are encoded by the *hcy* or *cyc* genes and are present in multiple copies and are probably co-transcribed (Arp et al., 2002).

Some AOM also possess urease genes that allow them to produce ammonia and carbon dioxide from urea. This capacity may be the curtail for nitrification in acidic environments where ammonia availability is low (Burton and Prosser 2001; Lu et al., 2012). Furthermore, at least one AOM, the archaea *Nitrososphaera gargensis*, possess a cyanase gene and uses it to produce ammonia from cyanate. As all other sequenced AOM genomes do not contain cyanase genes, it was proposed that *N. gargensis* acquired this gene through lateral transfer from a *Nitrospira* (NOB) strain (Palatinszky et al., 2015).

1.3.1.b Heterotrophic nitrification

Heterotrophic nitrification (HN) is carried out by a wide range of heterotrophic bacteria and fungi. HN is not related to ATP production and, in fungi, is linked to lignin degradation (Prosser, 2005). There are two pathways for HN, the first one is similar to autotrophic nitrification (involves ammonia- and hydroxylamine-oxidizing enzymes). The second one carries out the oxidation of amines or amides of organic molecules (Robertson and Groffman

2015). The physiological role of HN is not clear but it could play a significant role in acidic soils (Hayatsu et al., 2008). In some bacteria, it is linked to aerobic denitrification (Robertson and Kuenen, 1990). The capacity for HN could be an advantage for denitrifying in an environment where O_2 partial pressure fluctuate: HN could be used to generate nitrite or nitrate under oxic conditions which would then be available for denitrification when anoxic conditions are re-established (Castignetti and Hollocher 1984).

1.3.2 Nitrite oxidation

1.3.2.a Organisms involved in nitrite oxidation

The second part of biological nitrification is the oxidation of nitrite to nitrate and is carried out by Nitrite Oxidizing Bacteria (NOB), which are, like AOM, (chemo)-lithoautotrophes. NOB use nitrite as a source of electrons and CO_2 as a source of carbon although some strains possess heterotrophic capabilities (Ward, 2013). Recently, strains belonging to the genera *Thiocapsa* and *Rhodopseudomonas* capable of anoxygenic phototrophic nitrite oxidation to nitrate were discovered (Griffin et al., 2007; Schott et al., 2010). These phototrophic nitrite-oxidizing bacteria (pNOB) couple the photo-oxidation of nitrite to the fixation of CO_2 in organic compounds. The sequencing of *Thiocapsa KS1*'s genome revealed a diverse set of nitrogen utilization genes including *nxr* nitrite oxidoreductases, a complete set of nitrogen fixation *nif* genes and ammonia - and nitrite/nitrate-assimilating genes. The bacterium also possesses the genomic potential for sulfur, thiosulfate and sulphide photo-oxidation, making a direct link between the nitrogen and the sulfur cycles (Hemp et al., 2016). Relatively little is known about pNOBs, the rest of this section will therefore focus on chemo-autotrophic nitrite oxidation.

Nitrite oxidation has received relatively little attention compared to other nitrogen-related microbial processes (*e.g.* ammonia oxidation). However, new discoveries have proven that this process is much more complicated and interesting than previously thought. Representatives of chemo-autotrophic NOB are widely distributed in the bacterial phylogenetic tree, including gram-negative and gram-positive (one known) strains. NOB belong to seven genera shared between four phyla (Daims et al., 2016) :

- Phylum Proteobacteria: genus *Nitrobacter* (α-proteobacteria), genus *Nitrotoga* (β-proteobacteria) and genus *Nitrococcus* (γ-proteobacteria);
- Phylum Chloroflexi: genus *Nitrolancea*
- Phylum Nitrospinae: genus *Nitrospina* and genus "*Candidatus Nitromaritima*"
- Phylum Nitrospirae: genus *Nitrospira*

1.3.2.b Biochemistry of autotrophic nitrite oxidation

NOB oxidize nitrite to nitrate:

$NO_2^- + H_2O \rightarrow NO_3^- + 2H^+ + 2e^-$ $\Delta G^{0'} = -74$ kJ/mol NO_2^-

This reaction is catalysed by nitrite oxidoreductase (Nxr), probably composed of three subunits (NxrA, NxrB and NxrC) with a possible α2β2γ1 stoichiometry (Lücker et al., 2010). It has been proposed that in *Nitrobacter*, the Nxr is composed of four subunits (ABIC) (Simon and Klotz 2013). They are two phylogenetically distinct types of Nxr: periplasmic and cytoplasmic (Figure 1.7).

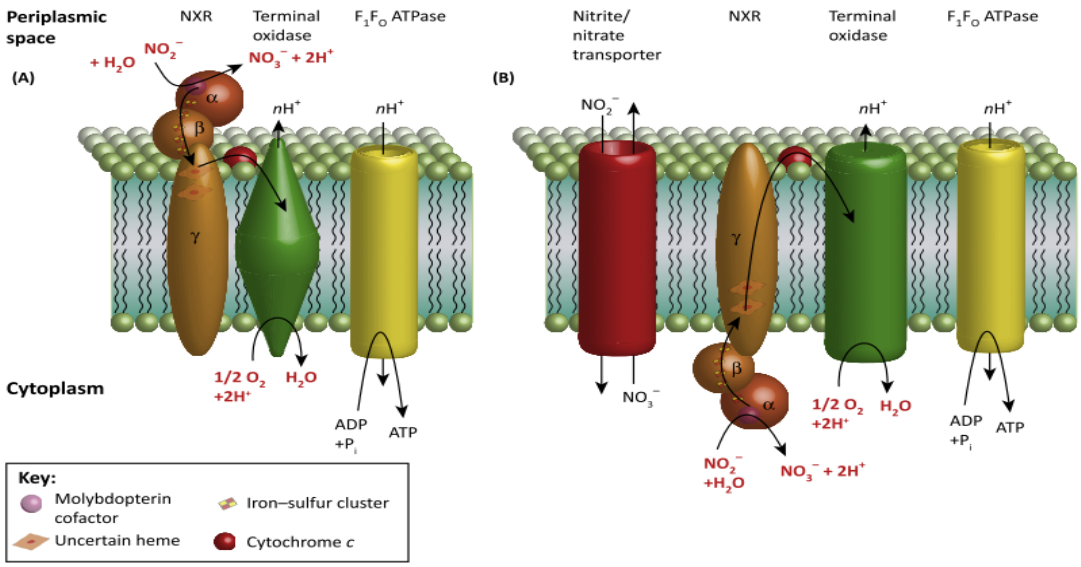


Figure 1.7 Schematic representation of the periplasmic (A) and cytoplasmic (B) Nitrite Oxidoreductase of NOB. Subunits of the Nxr protein are designated with Greek letters (α for NxrA, β for NxrB and γ for NxrC). Processes involving consumption or production of protons are highlighted in red. Electron flow from Nxr is represented with black arrows. Copied from Daims *et al*., (2016).

The α-subunit is the catalytic one and contains the molybdenum bis molybdopterin guanine dinucleotide (Mo-bis-MGD) cofactor and several iron-sulfur (Fe-S) clusters. The β subunit also contains Fe-S clusters and transfers the electrons from the α to the γ subunit. Then the γ subunit channels the electrons to and from the electron-transport chain via heme(s). The γ subunit also serves as a membrane anchor for the enzyme (Lücker et al., 2013).

The location of the Nxr (periplasmic or cytoplasmic) has a crucial impact on the biochemical reaction of nitrite oxidation: with a periplasmic Nxr, protons produced by the reaction are released in the periplasmic space, contributing positively to the proton motive force (PMF) at the origin of ATP production. On the other hand, a cytoplasmic Nxr imply that protons contribute negatively to PMF making nitrite oxidation less energetically attractive. Plus, with cytoplasmic Nxr, nitrite must be brought into the cell via a nitrite/nitrate transporter, which can constitute another limiting step depending on substrate affinity and turnover rate of the transporter (Daims et al., 2016).

Like in AOM, electrons are used for generating the PMF force via a terminal oxidase and to reduce NAD^+ to NADH via a reverse electron transfer. NADH and ATP are used for fixing CO_2 into organic matter via the CBB cycle or via the reductive tricarboxylic acid pathway (Ward, 2013).

1.3.2.c Metabolic versatility of NOB

Although the presence of NOB in the environment has generally been linked to nitrification, it is now known that they are very versatile organisms. For example, *Nitrospira moscoviensis*, besides having a classical aerobic nitrite-oxidizing metabolism, can grow anaerobically using H_2 and nitrate as electron donor and acceptor respectively. This "nitrifier" is therefore also able to consume nitrate (Ehrich et al., 1995). Furthermore, Koch et al. (2014) showed that *Nitrospira moscoviensis* can grow chemotropically using H_2 and O_2 as sole electron source and acceptor, respectively. They also linked this reaction to autotrophy by quantifying $[^{13}C]HCO_3^-$ uptake of *N. moscoviensis* during H_2-dependant growth. Finally, *N. moscoviensis* can perform a chemo-organotrophic metabolism using formate as an electron source with NO_3^-, O_2 or both as oxidants. As no genes encoding a dissimilatory nitrate reductase could be found in the genome of *N. moscoviensis*, it was proposed that the Nxr enzyme can operate in the reverse direction (NO_3^- to NO_2^-) (Koch et al., 2015). Together these studies show a rather large metabolic flexibility of *N. moscoviensis* with potentially four metabolisms 1: "classical" nitrification, 2: Aerobic oxidation of H_2 using O_2, 3: anaerobic oxidation of H_2 using NO_3^- 4: oxidation of formate with NO_3^-, O_2 or both as oxidants). This high versatility could be advantageous for surviving in oxic-anoxic transition zones (biofilms, flocs, sediments and soils). Some *Nitrobacter* and *Nitrospira* members can also have a mixotrophic metabolism, *i.e.* using simple organic carbon for heterotrophic growth together with autotrophic nitrite oxidation. Other strains of *Nitrobacter* can grow completely heterotrophically (Daims et al., 2016). The concept

of NOB as a functional group could therefore be questioned as these organisms have other major activities (beside autotrophic nitrite oxidation) depending on the environmental conditions of their habitats.

1.3.2.d Genetic basis of NOB

Nitrite-oxidation related genes: The α, β and γ subunits of Nxr are encoded by *nxrA*, *nrxB* and *nxrC* respectively. The number of copies and the organisation of these genes depends on the organism. Starkenburg et al. (2006) found two copies of *nxrA* and *nxrB* in the genome of *Nitrobacter Winogradsky Nb-255*. They identified a putative operon composed of one copy of *nxrA* and one of *nxrB*, together with genes associated with electron transport. Lücker et al. (2010) detected two clusters *nxrA1B1* and *nxrA2B2* in the genome of "*Candidatus Nitrospira defluvii*", with protein sequence identity between the two copies of 86.6% for *nxrA* and 100% for *nxrB*. They also found two putative genes encoding a nitrite reductase NirK suggesting that this organism might denitrify, and eventually from nitrate if the Nxr can catalyse the reverse reaction (nitrate to nitrite). However, it hasn't been proved in pure cultures. Lücker et al. (2013) studied the marine NOB *Nitrospina gracilis* and found two putative operons *nxrABC*. The protein sequences of α (NxrA), β (NxrB), and γ (NxrC) subunits have identities of 94.9%, 100%, and 87.5%, respectively. They also found several other putative NxrC encoding genes in the genome, some could be involved in the mechanism of nitrite oxidation. It is interesting to notice that Lücker et al. (2010) and Lücker et al. (2013) found that NrxB is highly conserved whereas NrxA slightly differs within a genome. The importance of the amino acids that differ between copies of NxrA (and NxrC) is however not known. Finally, Sorokin et al. (2012) studied the genome of *Nitrolancea hollandicus* (phylum *Chloroflexi*) and found four highly similar copies of *nxrA*, one of *nxrB* and one of *nxrC*. Three copies of *nxrA* were found to form a cluster also containing a two-component response regulator. The other one is located in a potential functional unit with the *nxrB* gene, a *narJ* homologue and *nxrC*.

Phylogenetic comparison of gene sequences of *nxr* revealed an interesting and complicated evolutionary history of nitrite oxidation. There are apparently two different phylogenetic origins of nitrite oxidation, reflected by the two orientations of the Nxr enzyme (cytoplasmic vs. periplasmic). In *Nitrobacter* and *Nitrococcus*, Nxr is oriented toward the cytoplasm. This type of Nxr resembles the one found in phototrophic purple proteobacteria. It was thus postulated that *Nitrobacter* and *Nitrococcus* evolved from phototrophic ancestors where nitrite was used as an electron donor for anoxygenic photosynthesis (Lücker et al., 2010; Lücker et

al., 2013). Recent phylogenetic analysis of the genome of the phototrophic nitrite oxidizer *Thiocapsa KS1* however didn't confirm the hypothetical phototrophic origin of Nxr in *Nitrobacter* and *Nitrococcus* (Hemp et al., 2016). On the other hand, the periplasmic Nxr found in NOB of the genera *Nitrospira* and *Nitrospina* show a closer relationship with nitrite oxidoreductase of ANAMMOX organisms. ANAMMOX are *Planctomycetes* and are not closely related to *Nitrospira* or *Nitrospina*, suggesting that these genes were horizontally transferred during the evolution of these lineages (Lücker et al., 2013). The Nxr in *Nitrolancea hollandicus* is oriented toward the cytoplasm but seems to be more closely related to the one found in nitrite dependant anaerobic methane oxidizer "*Candidatus Methylo-mirabilis oxyfera*" and, apparently, did not evolve from the same ancestors as the other known NOB (Sorokin et al., 2012).

Autotrophy: NOB are capable of autotrophic growth, using reducing power generated by the oxidation of nitrite to incorporate CO_2 into organic compounds. Genomic studies of representative NOB identified two different pathways for the fixation of CO_2: members of the genera *Nitrobacter*, *Nitrococcus* and *Nitrolancea* possess genes necessary for the CBB cycle (Sorokin et al., 2012; Starkenburg et al., 2006) whereas for members of the genera *Nitrospira* and *Nitrospina*, genes of the reductive Tri Carboxylic Acid (rTCA) cycle were found instead (Lücker et al., 2010; Lücker et al., 2013).

Other genes: NOB generally possess a rather wide metabolic versatility, and some are capable of heterotrophic growth. These observations on pure culture were in accordance with observations resulting from the sequencing of full genomes of NOB: Starkenburg et al. (2006) found genes encoding enzymes necessary for pyruvate, acetate, and glycerol metabolism in the genome of *Nitrobacter winogradskyi*. Lücker et al. (2010) showed that the genome of *Nitrospira* contains genes encoding enzymes for catabolic degradation and the assimilation of acetate, pyruvate, and formate; and candidate genes were found for the degradation of branched amino acids. They also found genes for the complete Embden–Meyerhof–Parnas pathway and putative sugar transporters. Sorokin et al. (2012) showed that *Nitrolancea hollandicus* can use H_2 as an electron donor and to use formate. Indeed, genomic annotation revealed the presence of hydrogenase genes (for the use of H_2), formate transporter and formate dehydrogenase in the genome of *Nitrolancea*. Another interesting feature of NOB genomes is the presence of ureases and cyanase genes, most likely involved in reciprocal feeding with AOM. NOB are able to produce ammonia from urea (Koch et al., 2015) and cyanate (Palatinszky et al., 2015).

A model of reciprocal feeding between ureolytic NOB and urease/cyanase-negative AOM was proposed, where NOB hydrolyses urea/cyanate to produce ammonia, which is used to feed the AOM. Nitrite is then produced by ammonia oxidation and is used by the NOB as an electron donor.

1.3.3 COMAMMOX

So far, nitrification has been presented as a two-step pathway with ammonia and nitrite oxidation as two distinct steps. Recently, another pathway, by which ammonia is successively converted to hydroxylamine, nitrite and nitrate was shown to occur in a single organism from the genus *Nitrospira* (Daims et al., 2015; van Kessel et al., 2015). Before these discoveries, complete oxidation of ammonia to nitrate was only hypothesised as it theoretically provides more energy than the two separate oxidations (Costa et al., 2006).

COMAMMOX *Nitrospira* possess Amo and Hao enzymes for the oxidation of ammonia to hydroxylamine and nitrite respectively. Interestingly, there is a higher similarity level between the Amo and Hao protein sequences between COMAMMOX *Nitrospira* and β-AOB (≈60% amino acid sequence identity for AmoA and ≈66% for HaoA) than between β-AOB and γ-AOB (≈45% for AmoA and ≈53% for HaoA), indicating a possible horizontal gene transfer (HGT) from β-AOB to an ancestor of COMAMMOX *Nitrospira*. They also possess an Nxr enzyme, for the oxidation of nitrite to nitrate, which is highly similar to the one found in nitrite oxidizing *Nitrospira* (Daims et al., 2016). The number of *nxr* gene copies varies from one to two in most COMAMMOX genomes to four (*Ca. N. nitrificans*) and five (*N. moscoviensis*) (Palomo et al., 2018).

Like ammonia and nitrite oxidizers, COMAMMOX are able to grow autotrophically using the reducing power generated by the oxidation of ammonia and nitrite to fix CO_2 into organic matter. Like the NOB *Nitrospira*, COMAMMOX use the reductive TCA cycle (Koch et al., 2019). As reported for NOB, the presence of organic carbon degradation genes in COMAMMOX genomes suggests that they have the potential to grow mixotrophically but the environmental importance of this metabolic capacity is still to be fully understood (Koch et al., 2019).

The recent studies of ammonia oxidation kinetics of *Nitrospira inopinata* revealed that the Comammox bacteria possess a lower K_m (0.84 μM total ammonia) than any other kinetically

characterized non-marine ammonia oxidizer, suggesting an adaptation to slow growth in oligotrophic environments (Kits et al., 2017; Koch et al., 2019). However, generalisation to all members of the group is difficult as other factors such as metabolic versatility and the ability to obtain nitrogen from other sources (*e.g.* urea, cyanate) also influence niche partitioning of nitrifiers (Kits et al., 2017). Furthermore, the presence of COMAMMOX-like genes (*amoA*, *amoB, amoC* and *hao*) in metagenomic datasets from freshwater, wastewater treatment plants (WWTPs) and drinking water could suggest that these organisms are widespread in the environment (Daims et al., 2016).

1.3.4 Nitrification in the environment

1.3.4.a Factors influencing AOM

The abundance and diversity of AOM are determined by environmental physio-chemical parameters. Nitrogen substrate is an important one. NH_3 rather than NH_4^+ is the substrate for AOM. If NH_3 concentration is too high, it can be toxic for cells but if it is too low it can cause substrate limitation (Prosser, 2005; Robertson and Groffman, 2015). NH_3 availability is controlled by different other parameters (*e.g.* salinity, temperature, pH):

$$F=1/(1+10^{pk_s}-pH-S)$$

and

$$pK_s= 2792.92/(TC +273.16) + 0.09018$$

(Where F = free ammonia fraction of NH_x; TC= temperature in °C; S = salinity factor (Bower and Bidwell 1978; Groeneweg *et al*., 1994))

Nitrifiers are in competitions with plants and heterotrophs for ammonia. In environments with low nitrogen mineralisation or fixation, the competition with plants and heterotrophs for the same substrate can be a major limitation, especially given that nitrifiers are relatively poor competitors for ammonia (Robertson and Groffman, 2015).

By controlling the balance NH_4^+/NH_3, pH is of primary importance for nitrifiers' metabolisms. Acidification of oceans due to the increase in atmospheric concentrations of CO_2 could therefore have a major impact on nitrification rates, influencing both the nitrogen and carbon cycle in oceans (Ward, 2013).

As nitrification requires oxygen, AOM are limited to oxic environments, although microaerophilic environments could also support autotrophic ammonia oxidation (Ward, 2013). Other important parameters are salinity, temperature and light (Ward, 2013). Salinity affects ammonia availability: high salinities release ammonia bound to sediments whereas low salinity tends to increase its absorption to sediments particles (Boatman and Murray 1982; Rysgaard et al. 1999; Seitzinger et al., 1991), but also causes osmotic and toxicity stress to cells. The effect of temperature on AOM is not clear as different results are reported in the literature. For example, Horak et al. (2013) examined the effect of temperature on nitrification rates of natural marine water samples and found that these rates were not significantly different in the 8°-20°C range. On the contrary, Berounsky and Nixon (1990), who measured nitrification rates over an annual cycle, found that it was strongly correlated with temperature (higher during summer). Likewise, Urakawa et al. (2008) reported that lower temperatures resulted in a lower diversity of AOA and AOB in aquarium biofilter systems. In soils, the temperature optimum for nitrification is often 30°C (Taylor et al., 2017).

The presence of inhibitors also affects nitrification. A wide range of synthetic compounds (McCarty, 1999; Kim et al., 2008) or naturally occurring molecules such as sulphide can inhibit nitrifiers. It is also possible that natural chemicals produced by plant roots, such as polyphenols, can inhibit nitrifier growth and activity (Hättenschwiler and Vitousek, 2000; Lodhi, 2016). On the other hand, some studies suggest that AOM, and in particular AOA, could benefit from plant root exudations (Chen et al., 2008; Herrmann et al., 2008; Ke et al., 2013; Wei et al., 2011). AOA and AOB do not respond equally to changes in these parameters and, although co-existing in the same environment, present a certain degree of niche differentiation.

1.3.4.b Interactions between AOM and NOB

Nitrite rarely accumulates in natural environments because nitrite oxidizers rapidly consume it. This partnership between AOM and NOB necessitates a microscale-spatial interaction between the two groups in biofilms to facilitate nitrite exchange (Arp and Bottomley 2006; Schramm et al., 1999). These interactions might not be limited to nitrite: NOB could stimulate the biofilm formation by nitric oxide (NO) production, as NO stimulates biofilm formation by *N. europaea*. This NO production by NOB could therefore be a means to recruit NO sensitive AOB to form nitrifying aggregates (Daims et al., 2016). Other cooperation may occur in these aggregates, for example, *N. europaea* doesn't possess siderophore-encoding genes but has siderophore receptors. It is therefore possible that *N. europaea* utilises siderophores produced

by NOB for iron uptake (Daims et al., 2016). Koch et al. (2015) also proposed a model of reciprocal feeding where ureolytic NOB hydrolyses urea to produce ammonia, which is used to feed non-ureolytic AOM. NOB then benefit from nitrite produced by AOM for growth. Finally, both NOB and AOB produce Quorum Sensing (QS) autoinducers, and *N. europaea* produces at least one autoinducer that is also produced by the NOB *N. winogradskyi* (Mellbye et al., 2015). An interspecies communication system might exist between NOB and AOM in order, for example, to synchronise extracellular polymeric substances (EPS) production to form biofilms (Daims et al., 2016; Mellbye et al., 2015).

1.3.4.c Niche differentiation

i) Identification of key environmental parameters

The relatively recent discovery of AOA has been a major breakthrough in the understanding of nitrification in the environment. The high level of difference in AOA and AOB physiologies has fuelled the search for evidence of niche differentiation, *i.e.* the processes that drive the natural selection between competitive species (Prosser et al., 2019). The first clue for niche differentiation is the fact that AOA dominate the AOM community in the open ocean (Francis et al., 2005; Wuchter et al., 2006; Agogué et al., 2008; Schleper and Nicol, 2010; Kitzinger et al., 2020) whereas AOB usually dominates in WWTP (Kumwimba and Meng, 2019). Because the first isolated AOA and AOB displayed high and low affinities (low and high K_m) for ammonia, respectively, the niche differentiation of AOM was first proposed to be driven by environmental ammonia concentration. Although this provides a good explanation for the dominance of the AOM with the lowest measured K_m (the AOA *Nitrosopumilus maritimus*) in the ocean where the concentration of ammonia typically ranges in the nM, it is more questionable for the general dominance of AOA in soils, where the typical ammonia concentration is above the K_m of both AOA and AOB (Hink et al., 2018). Furthermore, the reported K_m for some recently isolated soil AOA is similar to that of *Nitrosomonas* indicating that differences in substrate affinity might not be the explanation for niche differentiation in soils (Kits et al., 2017; Hink et al., 2017).

Further investigation in soils have however pointed to key factors that could explain the patterns of dominance: i) pH: AOA is generally more active in soils with pH < 6 which might be due to the existence of obligate acidophilic AOA (Gubry-Rangin et al., 2010; Shen et al., 2012; Aigle et al., 2019); ii) Ammonia supply source: AOA is preferentially active in soils where ammonia is supplied via mineralisation of organic matter whereas AOB prefers soils

with high inputs of inorganic fertilisers (Di et al., 2010); iii) Temperature: AOA seems to prefer higher temperatures than AOB with optima differences of ≈10°C (Ouyang et al., 2017; Taylor et al., 2017); iv) Competition: it appears that AOA and AOB can be more active than anticipated under unfavourable conditions if the other is inhibited (Hink et al., 2018; Zhao, Bello, et al., 2020). Aigle et al. (2020) experimentally tested these hypotheses by measuring the response (growth and activity) of AOA and AOB following perturbations in pH and temperature in agricultural soil microcosms. The authors showed that both AOA and AOB activity increased with temperature and confirmed AOB optimum at temperature <30°C. Furthermore, they showed that AOA dominated in their experimental conditions where ammonium was mainly supplied via mineralisation of organic matter and AOB was favoured when ammonia accumulated, which tend to confirm the hypothesis of ammonia supply preference between the two groups of AOM. They also showed that, as anticipated, one group is more active when the other one is inhibited. Interestingly, they also showed that AOA adapts faster to increase rather than decrease in pH. On the other hand, they surprisingly found no activity of AOA at low pH (4.5) but indicated this might be due to the insufficient incubation time. AOB, however, showed activity under these acidic conditions, which contradicts the hypothesis that their activity should increase at higher pH.

ii) AOA vs. AOB in estuaries and coastal sediments

Estuaries are interesting ecosystems to study AOM niche differentiation as they typically encompass zones of low and high salinity. Several studies have reported a clear shift in AOA and AOB abundance in sediments along salinity gradients, with a general trend of AOA and AOB more abundant in the low and high salinity parts of the estuary, respectively. In a study of two bays along the Irish coast, Zhang et al. (2018) showed a clear shift in AOA/AOB abundance with AOA dominant in the low salinity soil whereas AOB dominated in the high salinity sediments of the estuaries. The authors also pointed out the potential role of pH (low in soils/ higher in sediments) as an additional driver of this shift. Mosier and Francis (2008) showed a similar trend in the San Francisco Bay estuary with salinity as a major driver of the AOA/AOB dominance whereas ammonia concentration was not. Similarly, Santoro et al. (2008) showed AOA 10 times more abundant at the freshwater site of a subterranean estuary whereas AOB was 30 times more abundant at the marine site, with a possible combined effect of salinity and oxygen availability. These findings are consistent with other studies reporting the overall dominance of AOB in high salinity estuary sediments (Wankel et al., 2011; Zheng, et al., 2013a; Lee et al., 2018; Zhou et al., 2018). In contrast, other studies have reported AOB

always more abundant than AOA regardless of the salinity, with a general trend of AOB abundance increasing from low to high salinity sites (Magalhaes et al., 2009; Li et al., 2015; Liu et al., 2018; Zheng et al., 2020). In a study of temporal and spatial change in AOA and AOB abundance along the Yangtze estuarine Zheng et al. (2013b) showed that, while AOA abundance was generally higher at low salinity, AOB was generally more abundant than AOA regardless of the salinity. Then, in the same estuary, Gao et al. (2018) recorded, in summer, the lowest AOA abundance at the low salinity site and the lowest AOB abundance at the high salinity site, while highest abundance for both were recorded at the intermediate salinity site and overall AOB more abundant than AOA along the salinity gradient. Similarly, Puthiya Veettil et al. (2015) showed that AOB was always more abundant that AOA in the Cochin estuary, including at sites of low salinity. This tends to indicate that other physiochemical parameters than salinity are controlling the AOA/AOB dominance in estuary sediments.

Indeed, conflicting reports have found AOA more abundant than AOB in estuary sediments including marine sites (Caffrey et al., 2007; Marton et al., 2015; Bernhard et al., 2019; He et al., 2018) and along estuaries salinity gradient (Bernhard et al., 2010; Urakawa et al., 2014; Hou et al., 2018; Sanders and Laanbroek, 2018; Ming et al., 2020). Li et al. (2018) reported AOA more abundant than AOB in sediment and water from the Yellow River estuary, including sites of high salinity. The authors suggested that the relatively low ammonia concentration in their samples (<30 $\mu g.g^{-1}$) could explain why AOA was dominant. However, this value was higher than the ones reported in Zheng et al. (2013a,b) and these authors reported a dominance of AOB vs. AOA. Furthermore, the dominance of AOA in estuary sediments with high salinity and high ammonia concentration has been reported elsewhere (Moin et al., 2009; Cao et al., 2011). Overall, it therefore seems that ammonia and salinity alone might not be enough to explain the patterns of the dominance of AOA vs. AOB in estuaries and that other parameters could affect AOM abundances including pH (Zhang et al., 2018) and oxygen availability (Santoro et al., 2008; Magalhaes et al., 2009).

iii) Knowledge gap

Several questions therefore remain regarding the effect of key physiochemical parameters such as salinity, pH and ammonia concentration on niche differentiation of AOA and AOB in estuaries and coastal environments. Also, many published research has focussed on the numerical dominance of AOA vs. AOB (*i.e.* differences in gene copy number) but the correspondence between abundance and activity is unclear. For example, Zheng et al. (2013a)

found AOB dominant over AOA in most of their samples but they found no changes in PNR with or without ampicillin (bacterial inhibitor) suggesting that AOB, although being dominant, might not play an important role in ammonia oxidation. Duff et al. (2017) reported 10 fold differences in PNR between two bays that displayed similar *amoA* gene abundances. Therefore abundances alone have to be interpreted carefully in the context of complex environmental samples. Alternatively, this could reflect the inadequacy of the potential rate measurements methods that typically employ incubation conditions that do not reflect in situ physiochemical contexts.

Another caveat of measuring abundances alone is that it does not inform on the community composition. As different species within AOA and AOB likely have different tolerance for salinity, ammonia, high/low pH and different cell-specific ammonia-oxidizing rates, the question "how is there?" is arguably as important as "how abundant are they?". Zheng et al. (2020), for example, showed that although salinity was positively correlated with AOB abundance in the Yangtze estuary sediments, it was also correlated with a decrease in richness and inversely for AOA. Similarly, Gao et al. (2018) found a higher number of AOA OTUs in the samples with the lowest abundance of archaeal *amoA* genes, indicating that abundance (gene copy number) and richness (number of different OTUs) are not always correlated. Finally, Ming et al. (2020) found that the community composition but not the abundance of AOB was correlated to salinity changes along the Liaohe estuary. Together, these results illustrate the importance of AOM community composition together with abundance to determining their dynamics in the environment.

Finally, an important remaining question is how do changes in AOM gene abundance and composition actually reflect their dynamics compared to changes in their transcriptome? Indeed several studies report different responses of AOM communities at DNA and RNA levels when subjected to artificial or natural physiochemical gradients. For example, when following changes in AOM abundance and activity along two estuaries, Happel et al. (2018) found that although AOA was more abundant at both sites, AOB were more transcriptionally active. Duff et al. (2017) showed changes of up to 2 orders of magnitude in AOA *amoA* transcription between samples while gene abundance remained constant. In the same samples, they also found different patterns for AOB gene and transcript abundance. In a microcosms study of the effect of forced salinity on AOM from coastal sediment Zhang et al. (2015a) showed that, at DNA level, AOB was generally not affected by the increase in salinity with no significant

differences in *amoA* gene abundance from 5 to 15 and 30 PSU, while on the other hand, a significant decrease in AOB *amoA* transcript was observed at increased salinities. In a subsequent study, Zhang et al. (2015b) showed that AOA gene abundance generally dropped after 28 days of incubation under increased ammonia concentration (from 5.7µM up to 1142µM) but generally recovered to their initial number after 56 days. Gene abundance alone would therefore suggest that AOA is able to cope with an extreme increase in ammonia concentrations. However, no AOA *amoA* transcripts were detected at the highest ammonia concentration after 56 days indicating that these AOA were not transcriptionally active. Furthermore, DNA-SIP revealed that AOA was not incorporating $^{13}CO_2$ under these conditions suggesting the absence of autotrophic growth. This illustrates again how abundance based on gene copy number and activity based on transcripts abundance and/or SIP, when considered on their own, could lead to different conclusions. In this same study, AOB gene abundance and transcript/SIP data were more concordant, both showing a preference for higher ammonia conditions. However, gene abundance data revealed the strongest increase in abundance after 56 days under 685µM ammonia whereas transcript data revealed that AOB was more active at 114µM ammonia. These results again, if analysed individually, would lead to different conclusions regarding AOB adaptation to ammonia concentration. It is worth noting that the authors showed an increase in AOA *amoA* transcription between 5µM and 343µM ammonia and similar transcription between 5µM and 685µM ammonia. Both values are much higher than the K_m NH_3 reported for any AOA or AOB showing that inference of niche differentiation based ammonia affinity of pure culture might not be directly applicable when considering complex environments. Also, the authors concluded that AOA prefer ammonia limited environments; however, this study started from a low ammonia environment that was subjected to increases. Therefore it could be argued that their data instead indicated that the native AOA in their samples (*Nitrosopumilus maritimus*) prefers ammonia limited environments while it is still possible that other AOA, native from higher ammonia environments, might behave differently.

The often-recorded discrepancies between gene abundance and activity, as described above, could indicate that a high proportion of AOM present at a given time are not actively transcribing their nitrification genes. Indeed, in a survey of *amoA* gene and transcript abundances in relation to PNR, He et al. (2018) found ratios of active AOM to total AOM lower than 1% indicating that the majority of AOA and AOB genes were amplified from inactive cells. This strongly puts into question the validity of indexes such as the AOA/AOB

gene abundance ratio. Therefore, studies that link biochemical rate with microbial community compositions are needed (Damashek and Francis, 2018) but cannot be limited to the quantification and sequencing of genes alone. Identifying the main microbial drivers of nitrification in the environment requires studying AOM activities rather than their presence alone to gain insight into the parameters that regulate them.

1.4 Accessing Microbial Activity in the Environment

1.4.1 Rates measurement

One of the most common methods for assessing microbial activity is the measurement of biological activity rates. This technique is based on the measurement of the differences in the concentrations of nitrogen species (ammonia, nitrite, nitrate…) between the beginning and the end of the incubation period; the differences are then used to calculate a rate of oxidation or reduction per unit of time. If the sample was incubated in optimal substrate concentration, a potential rate is measured, if the sample was incubated in the same conditions as in the environment, an actual rate is measured (e.g. Eriksson et al., 2003; Laima et al. 1999). Generally, these rates are compared with rates obtained in samples incubated with inhibitors of the biological processes to take into account the abiotic transformations and/or to differentiate between pathways (van Groenigen et al., 2015; McCarty 1999; Prosser 2005; Rusmana and Nedwell 2004).

Heavy isotopes can also be used to determine the relative participation of different pathways to the production of a particular compound. ^{15}N-labelling techniques are used for example for determining the role of different microbial metabolisms in the production of nitrogen oxide species (e.g. Kool et al., 2011; Wrage et al., 2005; Zhu et al., 2013).

1.4.2 Functional genes as biomarkers

Phylogenetic studies based on the *16S rRNA* gene are useful techniques in microbial ecology to identify bacteria and archaea in natural systems much more accurately than culture-based approaches (Torsvik and Øvreås, 2002). The use of quantitative PCR (Q-PCR) facilitates the quantification of total bacteria and archaea from phylum to species level (Smith and Osborn, 2009). The nitrogen cycle genes encoding the enzymes directly involved in the biochemical reactions of the nitrogen cycle are also widely used as biomarkers (Table 1.2). Q-PCR based techniques are widely used to determining the relative importance of AOA vs. AOB in the first

step of ammonia oxidation (although this has its limitations when gene abundance is considered alone, as discussed previously). Several genes encoding enzymatic pathways of the nitrogen cycle are subjected to horizontal gene transfer (HGT) and, consequently, their phylogeny doesn't exactly match that of the *16S rRNA* genes (Fuhrman 2009; Wellington et al., 2003). Therefore, for surveying a specific reaction (*e.g.* nitrification), an approach based on functional genes is better suited than one based on *16S rRNA* genes. Table 1.2 lists a suite of commonly targeted functional genes of the nitrogen cycle that are used to differentiate different pathways (Figure 1.1). Following the distribution, diversity and abundance of functional genes in ecosystems helps understand the factors influencing microbial functioning. For example, changes in genomic potentials within a system reveal long-term adaptations of the community whereas short-term responses occur at the transcriptional level and can be assessed from the transcriptome.

Table 1.2 Summary of functional genes used as biomarkers for studding nitrogen cycle microbial processes. (*: Reference associated with Comammox)

Function	Target gene	Reference
Nitrogen fixation	*nifH*	Levy-Booth *et al*. (2014); Smith and Osborn (2009)
Autotrophic ammonia oxidation (Nitritation and Comammox*)	*amoA*	Daims *et al*. (2015)*; Junier *et al*. (2010); Levy-Booth *et al*. (2014); (Pinto *et al.*, 2016); Smith and Osborn (2009)
Autotrophic nitrite oxidation	*nxrA/ nxrB*	Pester *et al*. (2014)
Nitrate reduction	*narG* *napA*	Smith and Osborn (2009)
Nitrite reduction	*nirS* *nirK*	Smith and Osborn (2009)
Nitric oxide reduction	*norB*	Smith and Osborn (2009)
Nitrous oxide reduction	*nosZ*	Smith and Osborn (2009)
DNRA	*nrfA*	Smith and Osborn (2009)
Anammox	*hzo*	Junier *et al*. (2010)

1.4.3 Transcriptomics

Transcriptomics studies are used to study microorganisms in natural environments, and in particular, which fraction of the community is actively contributing to a specific process. Although the level of expression of a given mRNA might not always correlate with the level of the corresponding protein (Wang et al., 2012), these techniques are meaningful tools because responses of microbial communities to environmental changes may be better reflected by the transcriptome than the proteome (Carvalhais et al., 2012). Environmental transcripts cannot be directly quantified, and a step of Reverse Transcription (RT) is required to generate complementary DNA (cDNA). The RT reaction occurs as follows: A reverse-transcriptase enzyme (enzyme able to generate DNA using and RNA template), oligonucleotide DNA (primer) and the RNA preparation containing the transcript(s) of interest are mixed together. The oligonucleotide then binds to the RNA and the reverse transcriptase enzyme generates a complementary DNA strand, starting from the 3' end of the primers. At the end of the RT step, an RNA-cDNA structure is then generated, which can be subsequently be used in the same way as a DNA template. The RT step is therefore crucial, as it needs to faithfully convert the starting RNA targets into cDNA. Once the cDNA has been generated, several techniques can be used to identify and quantify the target transcript(s).

1.4.3.a Microarray

This technique is based on the hybridization of the target cDNA on an array containing multiple probes (*i.e.* a complementary sequence to the target). The intensity of the hybridisation then correlates to the level of expression of the transcript of interest. This technique has the advantage of being able to survey a large number of transcripts at once. Furthermore, amplification techniques can be used to overcome the limitations posed by the low concentrations of RNA typically obtained from natural environments. As such, microarray can be used to detect activities of microbial communities in natural environments (Gao et al., 2007). However, the degrees of sequence variation can be high among environmental samples, which can negatively affect the probe-target hybridisation. Furthermore, environmental samples such as soil and sediments can contain high levels of organic inhibitors that will also affect the assay (Saleh-Lakha et al., 2005; Parro et al., 2007).

1.4.3.b Metatranscriptomics

Recent advances in sequencing technologies and bioinformatics pipelines have allowed for the genome-scale investigation of complex microbial communities. The shotgun sequencing of all DNA and cDNA available from a sample provides information about the overall genes present (metagenomics) and expressed (metatranscriptomics), respectively, in the community. As such, metatranscriptomics permits an insight into the metabolic pathways that are triggered by environmental changes. It is now widely used in microbiology and is used to characterize the interaction between biogeochemical cycles (Hollibaugh et al., 2011), identify active members of the microbial community, detect regulatory antisense RNA or determine microbe-virus interactions (Carvalhais et al., 2012; Gutleben et al., 2018; Shakya et al., 2019). However, metatranscriptomics generally comes at a high cost and requires extensive computational power. Plus, as a non-targeted approach, it might miss the rare transcripts, depending on sequencing depth and removal of rRNA.

1.4.3.c RT-(Q)-PCR

Specific genes within the cDNA pool can also be quantified using Quantitative PCR (RT-Q-PCR). In RT-Q-PCR, only a single target from the cDNA is quantified in a given Q-PCR reaction. Alternatively, the amplicon generated by PCR amplification on the cDNA can undergo sequencing to characterize the active fraction of the microbial population. These methods are less high throughput compared to metatranscriptomics in terms of the number of

target transcripts that can be surveyed but has the advantage of being specific and more sensitive having lower detection limits. Plus, it generally comes at a lower cost, does not require high computational power and, for RT-Q-PCR, are quantitative unlike metatranscriptomics (only semi-quantitative). RT-Q-PCR and RT-PCR have therefore been extensively used in environmental microbiology to link gene expression to ecosystem processes (Smith and Osborn, 2009; Saleh-Lakha et al., 2011; Gadkar and Filion, 2013).

1.4.3.d Limitations of transcriptomics studies

Some limitations of RNA based techniques come from the extraction, storage, sequencing protocol and sequence analysis step (Carvalhais et al., 2012; McCarthy et al., 2015; Wang et al., 2012), as in DNA based studies, but an additional problem with RNA is its instability: the half- life is much shorter than DNA, within the range of seconds to minutes (Carvalhais et al., 2012; Steiner et al., 2019). The transcriptomic profile of cells might therefore be significantly affected by the changing conditions imposed by sampling. The determination of the quality of RNA before further processing is essential as the quantification of a transcript from degraded samples could bias the results (Die and Román 2012; Jahn et al., 2008). Another limitation of transcriptome-based studies is that mRNA levels might not always reflect protein concentrations due to post-translational events such as delay in protein synthesis and protein stabilisation. In the latter case, the organism might be active because the enzyme is still present long after transcription has stopped (Liu et al., 2016).

Further to the quality and integrity of the RNA samples, several different parameters such as the extraction/purification method (e.g. Lloyd, et al., 2010; Tournier et al., 2015) and the matrix from where the RNA is extracted can affect the result of the subsequent RT-Q-PCR (Carvalhais et al., 2012; Wang et al., 2012). The comparison of expression levels between different conditions or different time point can therefore be complicated. It is important to ensure that difference observed are due to actual differences in the environment and are not influenced by, for example, sample preparation. Some authors (*e.g.* Die and Román 2012; Fleige and Pfaffl 2006) suggested normalising the copy number of the transcript studied with the copy number of some reference transcript (Rocha et al., 2015). This approach generates more accurate results but, as shown by (Pérez-Novo et al., 2005), the comparison of gene expression levels between samples with very different RNA integrity will still be biased, even when using such normalizations. Therefore, a measure of RNA integrity of samples being compared is required.

The reverse transcription step can also have an influence on the level of transcripts detected. It is still unclear to what extent the variations generated by this step on the transcriptome of complex microbial communities are, but studies from the medical field have shown that the priming strategy (Bustin and Nolan 2004; Ståhlberg et al., 2004; Stangegaard et al., 2006), the enzyme (reverse transcriptase) (Bustin and Nolan 2004; Levesque-Sergerie et al., 2007) and the quantity of RNA used (Bustin and Nolan 2004; Ståhlberg et al., 2004; Levesque-Sergerie et al., 2007) have an effect on the level of transcripts detected. Therefore, gene expression levels determined with different RT techniques are generally not comparable.

1.4.4 Stable Isotope Probing

Stable Isotope Probing (SIP) is a method that allows the identification of functional groups that incorporate/degrade a particular substrate: First, stable isotopes of carbon (^{13}C) or nitrogen (^{15}N) are incorporated in a substrate and added to the environment of interest. Microbial groups that are able to utilise this substrate will incorporate it into their organic matter including their DNA if they are growing (*i.e.* replicating their genetic material and creating new cells) and into their RNA if they are active (*i.e.* transcribing their genes and producing enzymes). Nucleic acids are then extracted from this environment and labelled and unlabelled DNA/RNA are separated via a density gradient. The identification of source organisms within the labelled (heavy) fractions of the density gradient allows for the linking between the phylogeny of organisms and their function (Neufeld et al., 2007).

SIP has been widely used to study nitrification in the environment: Jia and Conrad (2009) used SIP to show that ammonia fertilization only stimulated CO_2 fixation by AOB and not AOA in agricultural soils. Later, in a study of this same soil environment, Pratscher et al. (2011) reported similar results with the absence of labelling of AOA DNA in soils under different ammonia fertilisation treatments. RNA-SIP on the other hand revealed that AOA were actively fixing CO_2, coupled to ammonia oxidation. The labelling was higher at the lower fertilisation used (15µg $N.g^{-1}$) and incomplete at 100µg$N.g^{-1}$. Bacterial *amoA* was labelled in RNA and DNA SIP and the labelling was stronger at the highest fertilization. From these results, the authors concluded that AOB are better adapted to ammonia-rich environments and could be the most important ammonia oxidizers in fertilised environments such as agricultural soils. These results are in good agreement with a long term study where DNA-SIP was used to show the differential selection of AOM in agricultural soils depending on the fertilisation practice: AOB and AOA were selected by the application of inorganic and organic fertilisers,

respectively (Jia et al., 2020), consistent with results from Xia et al. (2011) who showed AOA, AOB and NOB labelled in agricultural soils upon ammonia fertilisation, however to a lower extend for AOA after 28 days. Other studies have shown the preferential labelling of AOA in soils including paddy (Wang et al., 2015) and agricultural soils (Zhang et al., 2010) and to show that they respond more to warming compared to AOB (Hu et al., 2016). SIP has also been used to demonstrate the growth of COMAMMOX alongside AOA and AOB in soils (Zhao et al., 2020) and to study AOM in coastal ecosystems; for example, Wang et al. (2020) were able to determine the ratio between total and labelled gene copy number to estimate the ratio of active and inactive cells in coastal wetland and showed, overall, AOA more active than AOB.

1.5 Aims and Objectives

The first aim of this thesis is to address knowledge gaps in environmental transcriptomics related to the impact of RNA degradation and the RT workflow on subsequent quantification and sequencing results. The overall aim is to establish a robust and reproducible experimental workflow for transcripts extracted from complex environmental samples, in an effort to match the guidelines published in the molecular clinical research field (*e.g.* MIQE guidelines: http://miqe.gene-quantification.info/). Then, the second aim is to apply this knowledge to evaluate the effect of sedimentary structures on the distribution and activity of nitrifying microorganisms in coastal sediments, using the Montportail-Brouage intertidal mudflat, located on the French Atlantic coast, as a model ecosystem. This site was chosen due to its characteristic display of ridge/runnels sedimentary structures, which have been shown to significantly affect microbial nitrification (Laima et al., 1999, 2002) as measured by PNR assay. This site has also been the focus of microbial ecology studies investing carbon and nitrogen cycling (Lavergne et al., 2017; Lavergne et al., 2018 a,b), but never using transcriptomics based approaches to further link structure and function in an effort to explain the observed differences in nitrification rates. Studies that link microbial communities and biochemical rates are needed to identify the microbes driving nitrogen transformations in the environment, which requires robust transcriptomics workflows in order to target the active portion of the microbiome. Both these topics will be covered in this thesis, in experimental part II, and I respectively.

To address the first aim, two methodological studies are presented. The first study (Chapter II) aims to develop a novel method to measure RNA integrity and evaluate the effect of RNA

degradation on the quantification and sequencing of environmental transcripts. RNA integrity refers to how intact the RNA is after sample collection, storage and extraction (*i.e.* strand breaks and base degradation). As stated in Bustin and Nolan (2017), ensuring that the integrity of RNA extracted from different samples being compared is similar is paramount to ensuring the validity of biological results. In other words, ensuring that differences observed are biological and not methodological due to degraded RNA. Specifically, the following hypothesis will be tested:

Hypothesis 1: Current approaches to monitor RNA integrity (RIN/RQI) are based on ribosomal RNA (rRNA). However, results from the clinical filed (Brisco and Morley, 2012; Björkman et al., 2016) and our understanding of the robustness and stability of rRNA (*e.g.* Sidova et al., 2015) indicate that it is not a suitable proxy for messenger RNA (mRNA) integrity. **We therefore hypothesise that direct measure of messenger RNA is a better proxy for its integrity than rRNA**. This hypothesis will be tested in chapter II. To do so, we developed a method based on the differential amplification of mRNA targets of different length from the same transcripts as an indicator of the overall transcriptome integrity. The premise being that longer mRNA degrades more easily. We validate our approach by gradually degrading RNA using different methods (heat, UV, RNase) and comparing this approach with the commercial method (RIN). **We also hypothesise that the same sample with different degrees of degradation will result in different transcript quantification and alpha and beta diversity compared to the original sample**.

Next, in the second study (Chapter III), the methodological approach to generate complementary DNA (cDNA) from the extracted RNA is examined. The aim of the reverse-transcription (RT) reaction is to faithfully represent, quantitatively and qualitatively, the mRNA community structure and abundance. Specifically, the following hypothesis will be tested:

Hypothesis 2: Bustin and Nolan (2017), in their review of the use of RT in the molecular clinical field, recorded that the RT reaction can introduce differences of up to 100 fold for the same sample being reverse transcribed using different enzymes and priming. Consequently, authors declare: *"[…] the majority of published RT-qPCR data are likely to represent technical noise*". There is a complete lack of understanding of the effect of cDNA preparation in environmental transcriptomics and the impact it has on the subsequent ecological

interpretation. In this context, it is essential to establish the errors introduced by the RT reaction and identify what the most accurate and precise methodology is. As the target template is typically present at low abundances within a high background of other untargeted RNA, the RT priming strategy and the dynamic range of the enzyme used will affect the accuracy and precision of the recovery of transcripts. **Therefore, we hypothesise that the RT protocol (reverse transcriptase enzyme and priming strategy) used will significantly affect the quantification and sequencing results of environmental transcripts**. To test this, a combination of four commercial enzymes and two priming strategies were tested for their abilities to faithfully reverse-transcribe RNA into cDNA.

In Chapter IV, the optimised workflow for the quantification and sequencing of environmental transcripts is applied to reveal the impact of sedimentary structures on microbial activity and more generally, to understand if such heterogeneity in terrain characteristics are important when calculating overall nutrient processing at the ecosystem scale. Here, we focus on identifying the active microbial nitrifiers that are responsible for the previously observed differences in nitrification rates between ridges and runnels of the Montportail-Brouage intertidal mudflat, France. Specifically, the following hypothesis will be tested:

Hypothesis 3: Nitrification in coastal sediments is linked with the removal of excess nitrogen derived from anthropogenic activities inland and N_2O production. In this context, it is crucial to understand how community composition informs activity and what are the major environmental parameters to be taken into account when estimating global budgets. Previous studies have shown that the physical arrangement of the sediment in ridges and runnels can significantly enhance nitrification rates, which are higher in the runnels. Yet this is not taken into account when calculating global rates at the ecosystem scale. **Due to the close proximity between the semi-permanent ridges and runnels, we hypothesise that the higher nitrification in runnels is explained by a higher nitrifier activity (*i.e.* higher transcription of nitrification genes) as opposed to differences in nitrifier abundance or community composition**. To test this hypothesis, a sampling campaign on the Montportail-Brouage mudflat was carried out in summer 2019, with a focus on the differences in nitrification measured between ridges and runnels. Here, potential nitrification rates were measured in ridges and runnels, from three different sites. Differences in total and active microbial nitrifier were determined, using the optimised protocol developed in Chapters II and III. Finally, we

propose hypotheses to link the differences observed in physio-chemical conditions, active nitrifiers and nitrification rates in the Montportail-Brouage mudflat.

Experimental Part I: Development of a Robust Workflow for Environmental Transcriptomics

General Introduction

A key question in environmental microbiology is to determine the functioning and activity of microbial communities. While genomic approaches have resulted in an unprecedented understanding of their structure and complexity (Medini *et al*., 2008), they do not inform of the activity and functioning at a given time. In this case, targeting the transcriptome, which is the subset of genes that are actively transcribed at a given time, is more informative. While there can be substantial post−translational regulation that may prevent final protein synthesis and/or activity, gene expression is the direct link between the genome and the function it encodes and, therefore, a stronger link to activity than DNA approaches alone (Moran *et al*., 2013).

In environmental microbiology, transcriptomics is therefore essential to understanding which biochemical pathways are triggered by environmental conditions at a given time. RNAseq approaches facilitate primer free metatranscriptomics to reveal global gene expression profiles. It is now a widely used method in environmental microbiology and has allowed scientists to gain formidable insight into the genome-scale mechanisms used by microbes to adapt to changing environmental conditions (Shakya *et al.*, 2013; Gutleben *et al.*, 2018). However, it generally comes at a high cost and requires extensive data analysis. Plus, as an untargeted approach, it may require enrichment of the mRNA (via removal of ribosomal RNA) and will be dependent on sequencing depth to reveal rare transcripts among the diverse array of transcripts expressed in complex environmental samples. In contrast, Reverse-Transcriptase-Quantitative PCR (RT-Q-PCR) is directed via primers towards a single target. While this is much lower throughput in terms of a global overview of transcription, this approach facilitates transcript quantification that is specific, with high-sensitivity and low-detection limits over a wide dynamic range (Sanders *et al.*, 2014). RT-Q-PCR is high-throughput in terms of sample numbers, cost-effective (in comparison to metatranscriptomics) and subsequent data processing is fast without the requirement for high computational power and bioinformatics expertise needed for metatranscriptomics analysis. As a consequence, RT-(Q)-PCR is routinely used in most life science research fields including environmental microbiology to target and quantify specific transcripts.

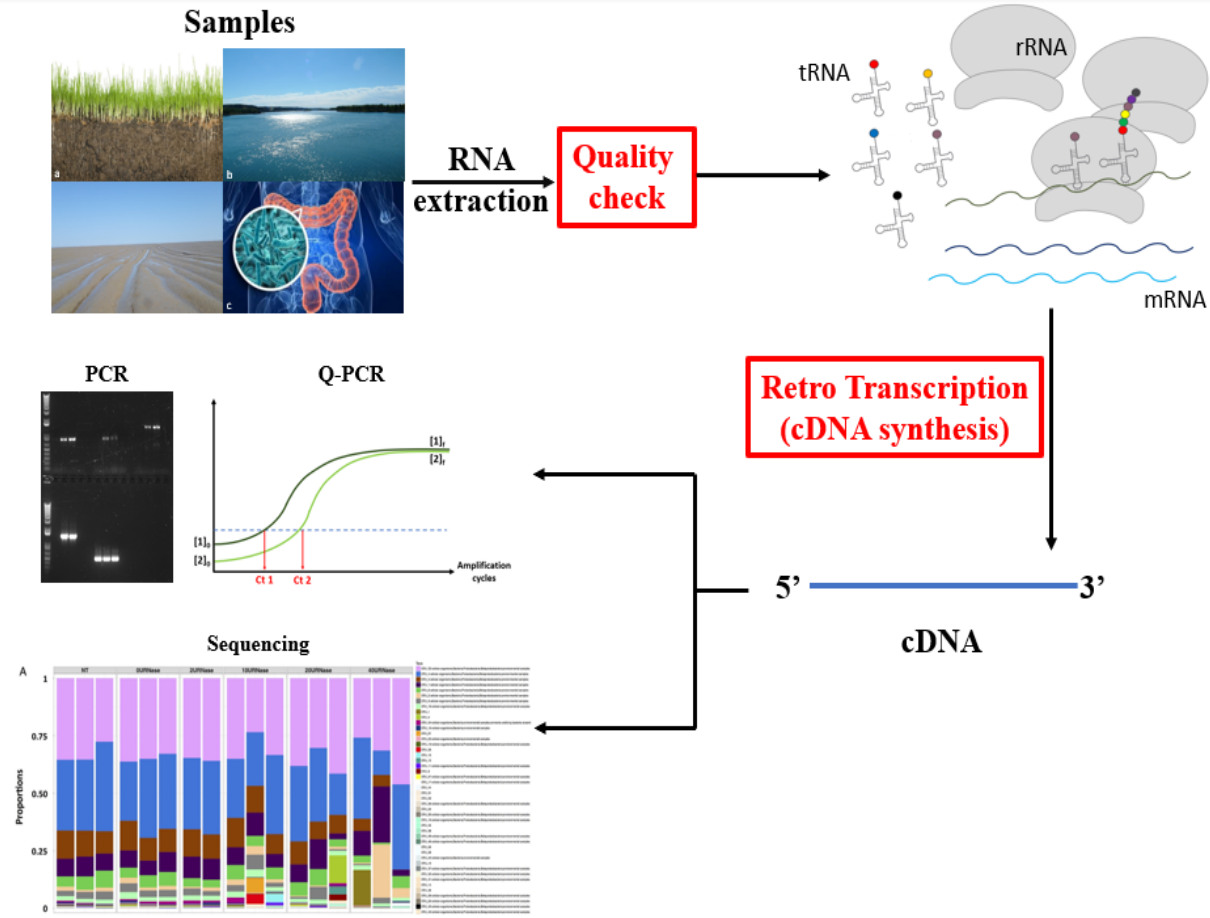


Figure I.1 Schematic representation of the RT-(Q)-PCR workflow, including RT-Q-PCR and cDNA amplicon sequencing. Steps highlighted in red represent the knowledge gaps in environmental microbiology that will be addressed to ensure reliable approaches to generate robust data from environmental samples.

As a result, the approach has been widely used to quantify transcripts to distinguish different pathways of the nitrogen cycle in sediments (Santoro *et al.*, 2010; Smith *et al.*, 2007; Zheng *et al.*, 2013; Damashek *et al.*, 2015; Duff *et al.*, 2017; Santos *et al.*, 2018; Zhang *et al.*, 2018), soil (Leininger, *et al.*, 2006; Graham *et al.*, 2011; Wang, Nagaoka, *et al.*, 2012; Li *et al.*, 2017; Pierre *et al.*, 2017), water column (Tolar *et al.*, 2016; Santoro *et al.*, 2010; Kapoor *et al.*, 2015; Posman *et al.*, 2017; Feng *et al.*, 2018; Gonçalves *et al.*, 2018; Liu *et al.*, 2018; Christiansen *et al.*, 2019) and other microbial processes including water treatment (Gadkar and Filion, 2013; Botes *et al.*, 2013; Wang *et al.*, 2016; Pelissari *et al.*, 2017, 2018) and bioremediation (Gadkar and Filion, 2013; Marzorati et *al.,* 2010; Yergeau *et al.*, 2009). In addition to this, cDNA from mRNA or rRNA can undergo PCR for amplicon sequencing to reveal actively transcribing organisms within the environment (Zhang *et al.,* 2018; Duff *et al.,* 2017).

The drawback of transcriptomics methods is that they are highly dependant on a robust methodology and require extra care compared to DNA-based methods. One of the main challenges of the RT-Q-PCR workflow (Figure I.1) is to efficiently extract the fragile RNA from the samples while preserving its integrity. Secondly, because the RNA cannot directly be quantified or sequenced, it has to be reverse-transcribed into complementary DNA (cDNA) before undergoing Q-PCR and PCR-sequencing. A reverse-transcription (RT) protocol that is accurate, precise and sensitive is thus essential to ensure that the information contained in the initial RNA has been conserved in the cDNA. In the field of clinical microbiology and cellular biology, these issues have been long recognised and addressed, leading to the publication of a list of recommendations for best practice when performing Q-PCR and RT-Q-PCR, the Minimal Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guideline (http://miqe.gene-quantification.info/). In environmental microbiology, efforts have been made to improve the extraction procedure and prove the robustness of subsequent Q-PCR (Smith et al 2006; Smith & Osborn, 2009), and PCR for amplicon sequencing (Marotz *et al.,* 2019). However, significant knowledge gaps remain in environmental microbiology if we want to reach similar standards of the MIQE guidelines (see Figure I.1); The work presented in the next two chapters is a much needed first step toward reaching such standards for microbial ecology studies. First, we develop an approach to monitor RNA integrity to evaluate mRNA quality prior to downstream processing and then we examine the impact of different RT approaches (enzyme and priming) on the quantification of RNA and mRNA transcripts. Based on these results, a number of recommendations for best practice in environmental microbiology transcriptomics will be presented that we hope will constitute the basis for establishing standards that match the ones expected in clinical transcriptomics studies.

Chapter II Differential Ratio Amplicon for the Evaluation of RNA Integrity Extracted from Complex Environmental Samples

This chapter is the basis for the peer-reviewed journal publication:

Cholet, F., Ijaz, U.Z., and Smith, C.J. (2019) Differential ratio amplicons (Ramp) for the evaluation of RNA integrity extracted from complex environmental samples. *Environ. Microbiol. 21(2): 827–844. doi: 10.1111/1462-2920.14516*

Fabien Cholet contributed to the study design, experimental work, data analysis and manuscript writing.

environmental microbiology

Environmental Microbiology (2019) **21**(2), 827–844 doi:10.1111/1462-2920.14516

Differential ratio amplicons (R_{amp}) for the evaluation of RNA integrity extracted from complex environmental samples

Fabien Cholet,* Umer Z. Ijaz and Cindy J. Smith
Infrastructure and Environment Research Division, School of Engineering, University of Glasgow, Glasgow, G12 8LT, UK.

Summary

Reliability and reproducibility of transcriptomics-based studies are dependent on RNA integrity. In microbial ecology, microfluidics-based techniques, such as the Ribosomal Integrity Number (RIN), targeting rRNA are currently the only approaches to evaluate RNA integrity. However, the relationship between rRNA and mRNA integrity is unknown. Here, we present an integrity index, the Ratio Amplicon, R_{amp}, adapted from human clinical studies, to directly monitor mRNA integrity from complex environmental samples. We show, in a suite of experimental degradations of RNA extracted from sediment, that while the RIN generally reflected the degradation status of RNA the R_{amp} mapped mRNA degradation better. Furthermore, we examined the effect of degradation on transcript community structure by amplicon sequencing of *16S rRNA*, *amoA* and *glnA* transcripts. We successfully sequenced transcripts for all three targets even from highly-degraded RNA samples. While RNA degradation changed the community structure of the mRNA profiles, no changes were observed for the 16S rRNA transcript profiles. Since both RT-Q-PCR and sequencing results were obtained, even from highly degraded samples, we strongly recommend evaluating RNA integrity prior to downstream processing to ensure meaningful results. For this, both the RIN and R_{amp} are useful, with the R_{amp} better evaluating mRNA integrity in this study.

Introduction

A key question in environmental microbiology is to determine the functioning and activity of microbial communities. While genomic approaches have resulted in an unprecedented understanding of their structure and complexity (Medini *et al*., 2008), they do not inform of the activity and functioning at a given time. In this case, targeting the transcriptome, that is the subset of genes that are actively transcribed at a given time, is more informative. While there can be substantial post-translational regulation that may prevent final protein synthesis and/or activity, gene expression is the direct link between the genome and the function it encodes and, therefore, a stronger link to activity than DNA approaches alone (Moran *et al*., 2013). As a result, transcriptomics-based approaches are widely used to assess microbial activity and functioning in the environment (Smith *et al*., 2006; Evans, 2015). The premise is that messenger RNA (mRNA) turn-over within cells is rapid, ranging from a few minutes to less than an hour (Laalami *et al*., 2014). As such, a snap-shot of the transcriptome reflects the cells transcriptional response to its surrounding environment and metabolic needs at a given time.

A challenge for all transcript-based studies, not least for those from environmental samples, is to ensure the quality and integrity of the RNA on which the results are based. Extracted RNA is prone to degradation both during the extraction procedure, post-extraction handling and over time. Factors such as *RNase* activity, physical degradation during extraction procedures and even storage can degrade RNA. If there is significant post-extraction degradation among different samples that are to be compared, the interpretation of results may be compromised. In other words, differences between samples may arise as a result of post-extraction degradation, as opposed to representing actual difference in gene expression. Indeed, meaningful and reproducible results can only be obtained when working with good quality, intact RNA, whether it is eukaryotic RNA (Fleige and Pfaffl, 2006; Fleige *et al*., 2006; Copois *et al*., 2007; Die and Román, 2012) or Prokaryotic RNA (Jahn *et al*., 2008). As such an initial quality check of extracted RNA, not least from complex environmental microbial communities should be the essential first step before proceeding to any downstream applications. This quality check would help to ensure that any differences observed between

Received 3 October, 2018; revised 7 December, 2018; accepted 20 December, 2018. *For correspondence. E-mail f.cholet.1@research.gla.ac.uk.

2.1 On the Issue of RNA Integrity

A challenge for all transcript−based studies, not least for those from environmental samples, is to ensure the quality and integrity of the RNA on which the results are based. Extracted RNA is prone to degradation both during the extraction procedure, post−extraction handling and over time. Factors such as *RNase* activity, physical degradation during extraction procedures and even storage can degrade RNA. If there is significant post−extraction degradation among different samples that are to be compared, the interpretation of results may be compromised. In other words, differences between samples may arise as a result of post−extraction degradation, as opposed to representing actual difference in gene expression. Indeed, meaningful and reproducible results can only be obtained when working with good quality, intact RNA, whether it is eukaryotic RNA (Fleige and Pfaffl, 2006; Fleige *et al*., 2006; Copois *et al*., 2007; Die and Román, 2012) or Prokaryotic RNA (Jahn *et al*., 2008). As such an initial quality check of extracted RNA, not least from complex environmental microbial communities should be the essential first step before proceeding to any downstream applications. This quality check would help to ensure that any differences observed between samples are due to actual changes in gene expression rather than differences in samples integrity as a result of degradation.

In microbial ecology, current methods to evaluate the integrity of extracted RNA are based on ribosomal RNA (rRNA). These approaches evaluate integrity as a ratio between the 23S and 16S ribosomal RNA: 23S, 16S and 5S rRNA are synthetized as one primary transcript and are separated upon maturation (Kaczanowska and Ryden−Aulin, 2007). The 23S and 16S ribosome should therefore be present at a ratio 1:1. However, as the 23S ribosome is approximately twice as long as the 16S ribosome, for intact, non−degraded RNA, the expected ratio of 23S:16S RNA is 2:1. However, the caveat of this approach is the assumption that the integrity of rRNA reflects that of the overall RNA, including mRNA. The relationship between the integrity of rRNA and that of mRNA has not been demonstrated (Die and Román, 2012). Indeed, the formation of secondary structures and the interaction with ribosomal proteins may help protect ribosomes from degradation and could explain the more stable properties of rRNA compared with mRNA (Bonincontro *et al*., 1998; McKillip *et al*., 1998; 1999; Fontaine and Guillot, 2003; Rathnayaka and Rakshit, 2010; Rhodes *et al*., 2012; Sunyer−Figueres *et al*., 2018). As such, the usefulness of this ratio to assess mRNA integrity is still unclear.

In its simplest form, evaluating ribosomal RNA integrity is an electrophoretic separation of RNA in a gel matrix. Essentially, a visual check for the presence of the characteristic bands corresponding to 23S and 16S rRNA. More advanced techniques based on microfluidics are better suited for assessing RNA quality, allowing for the calculation of integrity indexes, such as the RNA Integrity Number, RIN (Agilent Technologies) or the RNA Quality Score, RQS (BioRad). These scores vary between 0 (RNA totally degraded) and 10 ('perfect' RNA). A value of seven has been suggested as a limit between 'good' and 'bad' quality RNA extracted from bacterial pure cultures (Jahn *et al*., 2008). However, RNA extracted from natural environments such as soil or sediment will likely have lower quality due to the more complex matrixes and often harsh extraction techniques routinely used, for example bead beating (Hurt *et al*., 2001), but this information is not widely reported in the literature. Nevertheless, as highlighted above, even if reported, a shortcoming for RIN/RQI algorithms is that they are primarily based on rRNA (16S/23S ratio), which may degrade differently from mRNAs; the relevance of such indexes for gene expression analysis is therefore unknown.

In Eukaryotic gene expression studies, an alternative index often used to evaluate mRNA degradation is the 3′−5′ ratio (Die *et al*., 2011). This technique is based on the observation that Eukaryotic mRNAs generally degrade from the 5′ to the 3′ end, with the 3′ poly(A) tail acting as a protective agent. As a result, Reverse Transcriptase−PCR (RT−PCR) targeting the 5′ end of the transcript is less likely to produce amplicons than those targeting the 3′ end. A high 3′:5′ ratio (low 5′ copy number) is therefore an indication of mRNA degradation. This technique cannot be applied to prokaryotic mRNAs as they generally do not possess poly(A) tails, and when they do, the tail enhances mRNA degradation (Dreyfus and Régnier, 2002). Recently, a new approach called differential amplicon (Δamp) has been developed (Björkman *et al*., 2016). This technique is based on the differential amplification of RT−PCR amplicons of different lengths from the same mRNA target as a new means to determine RNA integrity (see also Karlsson *et al*., 2016). Here, it was observed that the copy number of long RT−Q−PCR targets correlated with mRNA degradation whereas short targets were more stable. Since this approach does not rely on the presence of the poly(A) tail, it could theoretically be adapted to prokaryotic mRNA. Although degradation of longer transcripts, faster, has not been directly observed previously in prokaryotes, Reck and colleagues (2015) showed a similar response of an exogenous green−fluorescent−protein mRNA (GFP), spiked into stool RNA, to monitor its integrity when subjected to different storage conditions. They showed that the copy number of

the spiked exogenous GFP correlated well with RNA integrity when targeting long amplicon (≥500 bp), whereas the short amplicon (≤100 bp) remained constant, even in highly degraded RNA preparations. This indicated that, as was observed by Björkman and colleagues, longer mRNA targets reflect degradation better. As such, the difference in RT−Q−PCR performance, reflected by the difference in cycle threshold (Ct) between a short and a long amplicon from the same cDNA target could be used as an index to reflect mRNA integrity.

Here, we propose to exploit the differential amplicon approach, initially developed by Björkman and colleagues, to develop a ratio of long to short amplicons of Bacterial mRNA transcripts using universal primers targeting conserved regions of the ubiquitous bacterial glutamine synthetase A transcript (*glnA*) as an indicator of overall mRNA integrity. Glutamine synthetase is a ubiquitous gene, found in Bacteria and Archaea (Kumada *et al*., 1993; Brown *et al*., 1994), with a role in assimilating inorganic nitrogen (ammonia) into amino acids (Reitzer, 2003). The *glnA* transcript has been used previously in RT−(Q)−PCR approaches to evaluate RNA extraction yield from soils (Sessitsch *et al*., 2002; Costa *et al*., 2004; Sharma *et al*., 2012). However, as the expression of *glnA* is regulated by ammonia concentration (Atkinson *et al*., 2002; Hua *et al*., 2004; Leigh and Dodsworth, 2007), the copy number of this transcript can vary making comparison between samples difficult. Our approach overcomes this difficulty by calculating the ratio of long to short *glnA* transcripts. We designate this the Ratio Amplicon (R_{amp}), and propose it as an indicator of mRNA integrity, independent of absolute gene expression.

Specifically, this study aims to design and test the Ratio Amplicon (R_{amp}) approach to evaluate bacterial mRNA integrity extracted from marine surface mud samples (0 to 2 cm; Rusheen Bay, Ireland) using a phenol−chloroform/bead−beating co−extraction method. Furthermore, we aim to compare and evaluate this approach against the conventional ribosomal based RNA Integrity Number, RIN. Comparison between the two approaches was conducted by monitoring how well both indexes reflected experimental RNA degradation (UV, heat, *RNase*, freeze/thaw and long−term storage). The impact of RNA degradation and the ability of the two indexes to predict ribosomal and mRNA integrity was evaluated via quantification of two commonly surveyed bacterial transcripts, the highly abundant ribosomal 16S rRNA and mRNA from the less abundant bacterial ammonia monooxygenase (*amoA*). Finally, the effect of RNA degradation on transcript community structure was evaluated by amplicon sequencing of the

cDNA obtained from sequentially degraded samples. A schematic representation of the experimental workflow undertaken in this chapter is presented in Figure 2.1.

We hypothesized that (i) the R_{amp} would be a better predictor of mRNA integrity than the RIN and (ii) RNA degradation would adversely affect both transcript quantification and community composition.

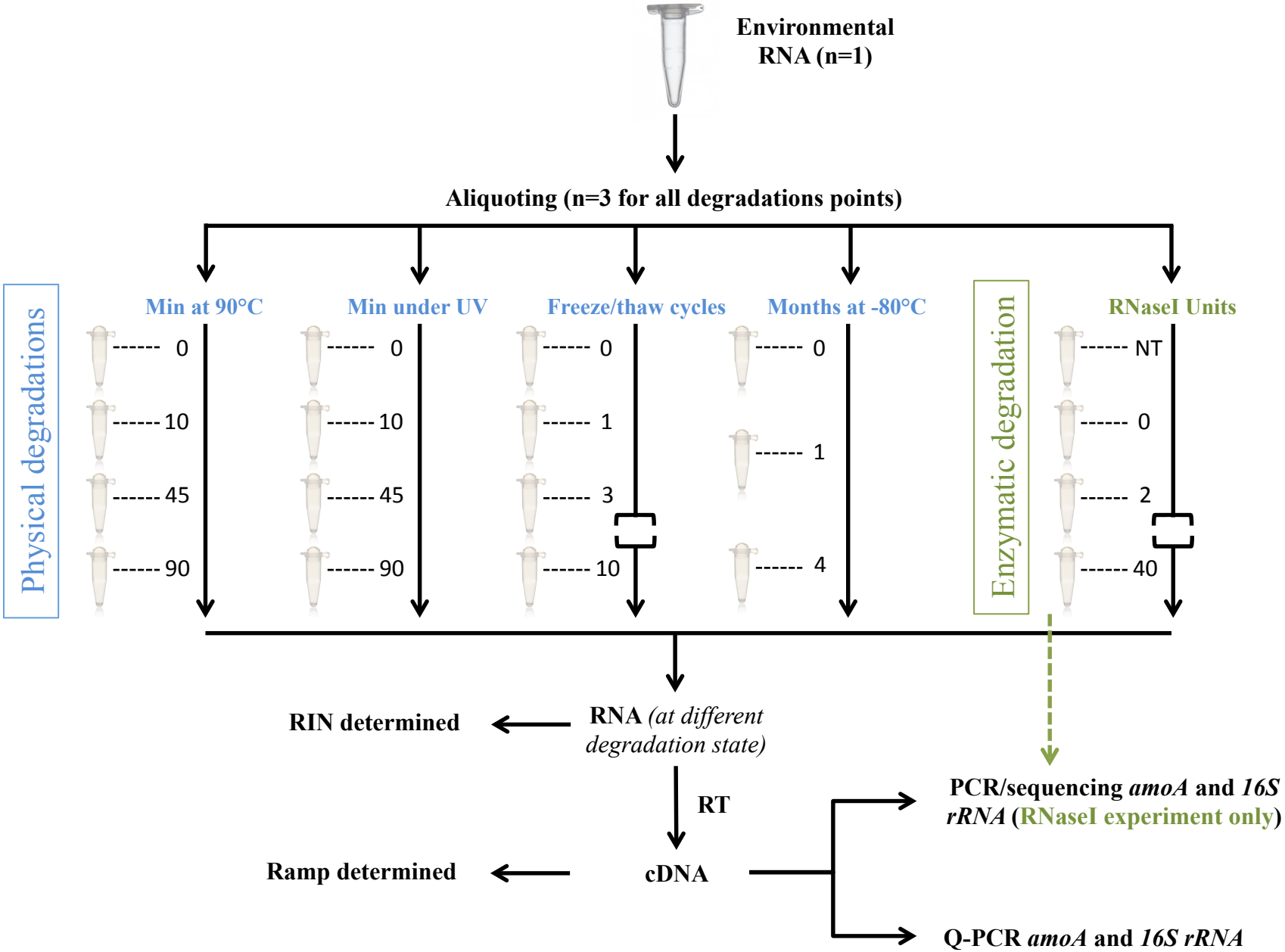


Figure 2.1 Schematic representation of the experimental workflow followed in chapter II.

2.2 Material and Methods

2.2.1 Sediment samples

Surface mud samples (0 to 2 cm) were collected on 11/01/2017 from Rusheen Bay, Ireland (53.2589° N, 9.1203° W) (presence of *amoA* genes/transcripts previously established (Duff *et al.*, 2017; Zhang *et al.*, 2018) in sterile 50ml Eppendorf tubes, flash frozen and stored at -80°C until subsequent use. Three biological replicate were used to test the effect of RNA degradation on the R_{amp}, RIN and *amoA* and *16SrRNA* Q-PCR/ PCR sequencing.

2.2.2 Design of new *glnA* primers

To design new primers, bacterial *glnA* sequences were downloaded from the GeneBank database (Clark *et al.*, 2016). Sequences related to environmental bacteria were subjected to BLAST search (Altschul *et al.*, 1990) in order to gather additional sequences. In total, 84 sequences (Supplementary Information 1) were aligned using MUSCLE (Edgar, 2004) and a phylogenetic neighbour joining tree was drawn in MEGA 7 (Kumar *et al.*, 2016). Based on sequence similarity, eight groups could be distinguished (Figure S.1). Primer sequences from Hurt and colleagues (2001) were aligned in each individual group to determine coverage and new primers were designed based on conserved regions to target the same groups with varying length primers.

Primers were tested on DNA and cDNA using environmental DNA/RNA extractions and environmental cDNA, as template. *glnA* genes were amplified (BIOTAQ DNA polymerase kit; Bioline) in a 25 µl final volume composed of 2.5 µl BioTaq10x buffer, 18 µl water, 1.5 µl $MgCl_2$ (50 mM), 0.5 µl of each primer (10 µM), 0.5 µl dNTPs (10 µM each), 0.5 µl *Taq* DNA polymerase and 1 µl of template. PCR conditions were as follow: 95°C 5 min, (94°C 30 s, 60°C 30 s, 72°C 30 s) × 30 and 72°C 5 min.

2.2.3 RNA preparation from sediment

All surfaces and equipment were cleaned with 70% ethanol and RNase Zap (Ambion) before sample processing. All glassware and stirrers used for solutions were baked at 180°C overnight to inactivate RNases. All plastic ware was soaked overnight in RNase away solution (ThermoFisher Scientific). Consumables used, including tubes and pipette tips were RNase free. All solutions were prepared using Diethylpyrocarbonate (DEPC) treated Milli-Q water. A simultaneous DNA/RNA extraction method, based on that of Griffiths and co-workers

(Griffiths *et al.*, 2000) was used to recover nucleic acids from sediment. Briefly, 0.5g of sediments were extracted using bead beating lysing tubes (Matrix tube E; MP Biomedical) and homogenised in 0.5ml CTAB/phosphate buffer (composition for 120 ml: 2.58g $K_2HPO_4.3H_2O$; 0.10g KH_2PO_4; 5.0g CTAB; 2.05g NaCl) plus 0.5ml Phenol:Chlorophorm:Isoamyl alcohol (25:24:1 v:v:v). Lysis was carried out on the FastPrep system (MP Biomedical) (S: 6.0; 40sec) followed by a centrifugation at 12,000g for 20 min 4°C). The top aqueous layer was transferred to a fresh 1.5ml tube and mixed with 0.5ml chloroform:isoamyl alcohol (24:1 v:v). The mixture was centrifuged at 16,000g for 5 min (4°C) and the top aqueous layer was transferred to a new 1.5ml tube. Nucleic acids were precipitated by adding two volumes of a solution containing 30% poly(etlyleneglycol)$_{6000}$ (PEG6000) and 1.6M NaCl for 2 hours on ice and subsequently recovered by centrifugation at 16,000 x g for 30 min (4°C). The pellet was washed with 1ml ice-cold 70% ethanol and centrifuged at 16,000g for 30 min (4°C). The ethanol wash was discarded, and the pellet was air dried and re-suspended in 40µl DEPC treated water. DNA/RNA preparations were stored at -80°C if not used immediately. RNA was prepared from the DNA/RNA co-extraction by DNase treating with Turbo DNase Kit (Ambion) using the extended protocol: half the recommended DNase volume is added to the sample and incubated for 30 min at 37°C, after which the second half of DNase is added, and the sample is re-incubated at 37°C for 1 hour. Success of the DNase treatment was checked by no PCR amplification of the V1-V3 Bacterial *16S rRNA* gene (Smith *et al.*, 2006).

2.2.4 RNA degradation experiments

2.2.4.a Physical degradation

To obtain RNA with controlled degradation status, DNA free RNA preparations (≈8 µl) were aliquoted from an initial extraction in separate 0.2 ml *RNase* free tubes and incubated at 90°C or under a UV lamp for 0, 10, 45 or 90 min. To determine the potential effect of repeated freeze–thaw on RNA preparations, the same 15 µl DNA−free RNA was exposed to cycles of freezing (at −80°C) and thawing (on ice) as follows – 0, 1, 3, 5, 7 and 10 freeze–thaw cycles. cDNA was then generated for each individual aliquot as described later. Aliquots of RNA were stored at −80°C for up to 4 months, at time 0, 1 month and 4 months samples were removed from the freezer to determine RNA integrity.

2.2.4.b Enzymatic degradation by RNase I

For *RNAse I* degradation experiment, 40 µl aliquots of DNA−free RNA were incubated at 37°C for 40 min in the presence of increasing concentrations of *RNase I* (Ambion): 0 (buffer only), 2, 10, 20 and 40 Units *RNase I* μg^{-1} RNA. The reaction was stopped by adding 10 µl β−mercaptoethanol and RNA was recovered by ethanol precipitation: 5 µl of 7.5 M ammonium acetate and 137.5 µl 100% ethanol was added and the mixture was precipitated overnight at −20°C. RNA was pelleted by centrifugation 16 000*g* for 40 min at 4°C and the pellet was washed with 480 µl ice cold 70% ethanol and pelleted by centrifugation at 16 000*g* for 30 min at 4°C. The pellet was air-dried and re−suspended in 40 µl of DEPC−treated water. An aliquot of RNA that did not undergo ethanol precipitation was also included for comparison (designated NT: 'Not Treated').

2.2.5 Reverse-Transcription

DNA-free RNA was used for glnA cDNA synthesis using Superscript III kit (Invitrogen) and gene specific priming. The initial RT mixture containing 3 µl water, 1 µl reverse primer GS1_new (10 µM), 1 µl dNTP's (10 mM each) and 5 µl template was incubated at 65 C for 5 min and quickly transferred on ice for 1 min. A second mix composed of 4 µl 5X first-strand buffer, 1 µl 0.1 mM dithiothreitol (DTT) and 1 µl SuperScript III (200 units /µl) was added and the resulting mixture was incubated at 55 C for 50 min and then at 72 C for 15 min. The primers and PCR conditions for the amplification of glnA targets from cDNA were similar to those used for DNA. For 16S rRNA and amoA genes, Superscript III kit (Invitrogen) and random hexamers priming was used. The initial RT mixture containing 3 µl water, 1 µl random hexamer (50 µM), 1 µl dNTP's (10 mM each) and 5 µl template was incubated at 65 C for 5 min and quickly transferred to ice for 1 min. A second mix composed of 4 µl 5X first-strand buffer, 1 µl 0.1 mM dithiothreitol (DTT) and 1 µl SuperScript III (200 units /µl) and 1 µl RNase inhibitor (40 U /µl) was added and the resulting mixture was incubated at 25 C for 5 min, 55 C for 50 min and then at 72 C for 15 min.

2.2.6 *amoA* and *16S rRNA* Q-PCR

For all degradation experiments, the Cts of the Bacterial *amoA* and the Bacterial 16S rRNA was determined by Q-PCR of the cDNA preparations. The *amoA* Q-PCR was carried out in a 20 µl reaction volume composed of 10 µl EVAGreen Supermixes (SsoFast; Bio-Rad), 0.4 µl of each primer (BacamoA-1F and BacamoA-2R) (10 µM each), 7.2 µl water and 2 µl of cDNA

template (1/10 diluted). The Q-PCR cycle was as follows: 95°C-5 min, (95°C-30sec, 47°C-30sec, 72°C-1min, 81°C-1sec and plate read) × 40 cycles. Melt curve analysis was performed from 65 C to 95 C with 0.5 C increment every 5 s. 16S rRNA cDNA targets were quantified in a 20 µl reaction volume composed of 10 µl Itaq Universal Probes Super- mix (Bio-Rad), 1.8 µl each primer (1369F and 1492r) (10 µM each), 0.4 µl probe (1389P) (10 µM), 5 µl water and 1 µl cDNA template (1/10 diluted). The Q-PCR cycle was as follows: 95 C-10 min, (95 C-10s, 60 C-30s) × 40 cycles and 40 C-10 min. All primers and PCR conditions are detailed in Table 2.1.

2.2.7 RNA integrity evaluation

2.2.7.a RNA integrity number

RINs were determined at all degradation points, using the automated 2100 Bioanalyser platform (Agilent Technologies) with the Prokaryote total RNA Nano chip, following the manufacturer's instructions.

2.2.7.b glnA Q−PCR and ratio amp (R_{amp}) calculation

glnA cDNA underwent Q−PCR, to amplify varying length amplicon fragments with primer combination as detailed in Table 2.2. Three *glnA* amplicons were produced (Figure 2.2), a 120 bp amplicon (amplicon 1) generated using the primer pair GS1_new/GSFw1200, a 170 bp amplicon (amplicon 2) generated using the primer pair GS1_new/GS2_new and a 380 bp amplicon (amplicon 3) generated using the primer pair GS1_new/GSFw900. Q−PCR reaction (10 µl) was composed of 5 µl EVAGreen Supermixes (SsoFast; Bio−Rad), 0.3 µl of each primers (10 µM) and 1 µl of cDNA template (1/10 diluted). The Q−PCR condition was as follows: 95°C−30s, (95°C−10s; 65°C−10 s) × 35 cycles; plate read at 65°C. Melt curve analysis was performed from 65°C to 95°C with 0.5°C increment every 5 s.

The Ct value of each assay was recorded and the differential amplicon ratios (R_{amp}) were calculated for each degradation point as follows:

$$R_{amp} = \frac{35 - Ct(\text{long amplicon})}{35 - Ct(\text{short amplicon})}$$

The value of 35 was chosen as the maximum number of Q−PCR cycles the reaction underwent. A transformation of the differential amplicon was applied in order to have a theoretical

maximal value of 1 (no degradation of RNA) and a theoretical minimal value close to 0 (totally degraded RNA).

Table 2.1 List of primers used and corresponding (Q)-PCR conditions used in Chapter II.

Primer	Sequence (5'→3')	Orientation	Target	Experiment condition	Reference
BacamoA-1F	GGGGHTTYTACTGGTGGT	Forward	Bacterial *amoA* gene (485bp)	**PCR:** 95°C-15min; (94°C-30sec; 55°C-30sec; 72°C-30sec) x 32 cycles and 72°C-10min **Q-PCR:** 95°C-5 min; (95°C-30sec; 47°C-30 sec; 72°C-1min; 81°C-1sec→ plate read) x 40; melt curve: 65°→95° 0.5° increment/ 5sec	Hornek *et al.*, 2006
BacamoA-2R	CCCCTCBGSAAAVCCTTCTTC	Reverse			
1369F	CGGTGAATACGTTCYCGG	Forward	Bacterial *16S rRNA* gene (123 bp)	**Q-PCR:** 95°C-10min; (95°C-10sec; 60°C-30sec) x 40; 40°C-10min.	Suzuki *et al.*, 2000
1492R	GGWTACCTTGTTACGACTT	Reverse			
1389P	CTTGTACACACCGCCCGTC	Probe			
515F	GTGYCAGCMGCCGC GTAA	Forward	Bacterial *16S rRNA* V4 (291bp)	**PCR:** 95°C-15min; (94°C-45sec; 50°C-30sec; 72°C-40sec) x 25; 72°C-10min	Parada *et al.*, 2016 Caporaso *et al.,* 2010
806R	GGACTACNVGGGTWTCTAAT	Reverse			

2.2.8 Illumina MiSeq amplicon library preparation

The qualitative effect of RNA degradation on the community composition of the three bacterial genes (*amoA*, *glnA* and *16S rRNA*) was determined by sequencing the amplicons generated from the cDNA preparations obtained after RNAse I degradation. For each PCR, amplification was carried out using the HotStartTaq PCR kit (Qiagen) in the following mix 25 µl volume: 19.8 µl water, 0.5 µl of each primer (10 µM each), 0.5 µl dNTPs (10 µM each), 0.2 µl HotStartTaq, 2.5 µl of 10x PCR buffer and 1 µl cDNA template (10^{-1} and 10^{-3} diluted for functional genes and 16S rRNA respectively). Primers used for sequencing were modified by adding Illumina adaptors at the 5' end: 5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG (forward adaptor); 5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G (reverse adaptor). The use of vector-targeting forward primer ensured that only the spiked mock communities were amplified. The specificity of this PCR assay was verified by the absence of amplification from the un-spiked reverse transcribed background. The PCR was carried out using the HotStartTaq PCR kit (Qiagen) in a 25µl volume: 19.8µl water, 0.5µl of each primer (10µM each), 0.5µl dNTPs (10µM each), 0.2µl HotStartTaq, 2.5µl of 10x PCR

buffer and 1µl cDNA template (10^{-1}). The PCR conditions for each target are detailed in Table 2.1. For the sequencing of *amoA*, PCR were carried out in triplicate and pooled together for cleaning. PCR amplicons were cleaned using the Agencourt AMPure XP beads (Beckman Coulter) following the manufacturer's recommendations. Illumina indexes were attached using the Nextera XT Index Kit with the following PCR condition: 95°C-15min, (95°C-30sec, 55°C-30sec, 72°C-30sec) x 8 cycles and 72°C-5min. The resulting amplicons were purified using the Agencourt AMPure XP beads (Beckman Coulter) and eluted in 25µl water. After this step, some preparations were randomly chosen (two per gene target) and analysed on the Bioanalyser using the DNA 1000 Assay protocol (Agilent Technologies) to determine the average length of the amplicons and to check for the presence of unspecific products. Finally, DNA concentration was determined using fluorometric quantification method (Qubit) and molarity was calculated using the following equation:

$$(\text{concentration in ng/µl}) \times 10^6 = (660 \text{ g/mol} \times \text{average library size}).$$

Libraries were pooled in equimolar amount and checked again on the Bioanalyser and the final library was sent to the Earlham Institute (Norwich Research Park, Norwich, UK) for Illumina MiSeq amplicon sequencing (300PE, 22 millions reads/ lane).

2.2.9 Processing of amplicon sequences

Construction of the reference databases

The following sequences were downloaded: *amoA* sequences from Fungene (http://fungene.cme.msu.edu/) alongside NCBI sequences (n = 642); and bacterial *glnA* sequences (n = 1330) as FASTA files from Microbial Genome Database (http://mbgd.genome.ad.jp). For *amoA* sequences, the NCBI taxonomy was given in the FASTA headers whereas for *glnA* sequences, the MBGD Archive (http://mbgd.genome.ad.jp/htbin/view_arch.cgi) was used to download annotations (mbgd_2016_01) associated with the sequences, and a custom script was written to identify and tag the sequences with NCBI taxonomy. Subsequently, R's rentrez (Winter, 2017) package was used to get taxonomic information at different levels to generate a taxonomy file for *glnA* sequences. The FASTA file and the corresponding taxonomy file was then formatted to work with Qiime. For *16S rRNA* we used the SILVA SSU Ref NR database release v123. (More details in Supplementary Information 2).

Bioinformatics pipeline

Abundance tables were obtained by constructing operational taxonomic units (OTUs) as follows. Paired-end reads were trimmed and filtered using Sickle v1.2 (Joshi & Sickle, 2011) by applying a sliding window approach and trimming regions where the average base quality drops below 20. Following this we apply a 10 bp length threshold to discard reads that fall below this length. We then used BayesHammer (Joshi & Sickle, 2011) from the Spades v2.5.0 assembler to error correct the paired-end reads followed by pandaseq v(2.4) with a minimum overlap of 20 bp to assemble the forward and reverse reads into a single sequence. The above choice of software was as a result of author's recent work (Schirmer *et al.*, 2015; D'Amore *et al.*, 2016) where it was shown that the above strategy of read trimming followed by error correction and overlapping reads reduces the substitution rates significantly. After having obtained the consensus sequences from each sample, the VSEARCH (v2.3.4) pipeline (all these steps are documented in https://github.com/torognes/vsearch/wiki/VSEARCH-pipeline) was used for OTU construction. The approach is as follows: the reads are pooled from different samples together and barcodes added to keep an account of the samples these reads originate from. Reads are then de-replicated and sorted by decreasing abundance and singletons discarded. In the next step, the reads are clustered based on 97% similarity, followed by removing clusters that have chimeric models built from more abundant reads (--uchime_denovo option in vsearch). A few chimeras may be missed, especially if they have parents that are absent from the reads or are present with very low abundance. Therefore, in the next step, we use a reference-based chimera-filtering step (--uchime_ref option in vsearch) using a gold database (https://www.mothur.org/w/images/f/f1/Silva.gold.bacteria.zip) for *16S rRNA* sequences, and the above created reference databases for *amoA* genes. The original barcoded reads were matched against clean OTUs with 97% similarity to generate OTU tables. The representative OTUs were then taxonomically classified using assign_taxonomy.py script from Qiime (Caporaso *et al.*, 2010) against the reference databases. To find the phylogenetic distances between OTUs, we first multi sequence aligned the OTUs against each other using Mafft (Katoh *et al.*, 2009) and then used FastTree v2.1.7 (Price *et al.*, 2010) to generate the phylogenetic tree in NEWICK format. Finally, make_otu_table.py from Qiime workflow was employed to combine abundance table with taxonomy information to generate biome file for OTUs.

2.2.10 Statistical analysis

All statistical analysis was carried out in R (R Development Core Team, 2008). For degradation experiments, RIN and R_{amp} values were compared between time points with one−way

ANOVA, when the ANOVA test was significant, differences between time points were investigated using Tuckey HSD post hoc test. For community analysis (including alpha and beta diversity analyses), the vegan package was used (Oksanen *et al*., 2005). To find OTUs that are significantly different between multiple conditions (Degradation), DESeqDataSetFromMatrix() function from DESeq2 (Love *et al*., 2014) package with the adjusted p−value significance cut−off of 0.05 and log2 fold change cut−off of 2 was used. Vegan's adonis() was used for analysis of variance (henceforth referred to as PERMANOVA) using distance matrices (BrayCurtis/Unweighted Unifrac/Weighted Unifrac for gene sequences) *i.e.*, partitioning distance matrices among sources of variation (Degradation). The scripts for above analysis can be found at http://userweb.eng.gla.ac.uk/umer.ijaz/#bioinformatics.

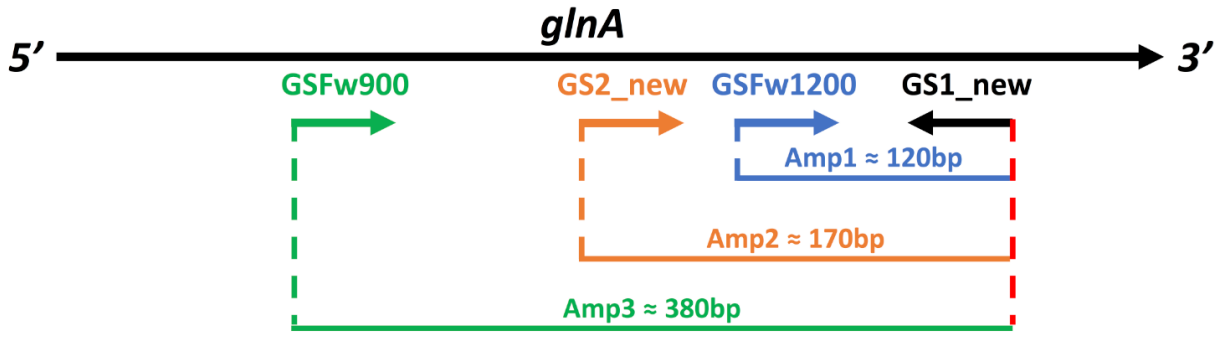
2.3 Results

2.3.1 Design and optimization of *glnA* primers

Three new forward *glnA* primers (GSFw1200, GSFw900 and GSFw800) were designed to target a conserved region in groups 3, 4, 5, 7 and 8 of the *glnA* alignment (Table 2.2, Figure S.1) at ≈120 bp, ≈380 bp and ≈500 bp, respectively, in front (closer to the 5′ end of the gene) of an updated reverse primer from Hurt and colleagues (2001) named, GS1_new primer. This resulted in three amplicon sizes to derive a ratio amplicon (R_{amp}) from (Figure 2.2). The newly designed primers (Table 2.2) were optimized for PCR and RT−PCR resulting in amplicons of the expected size for all primer pairs. Assays were subsequently optimized for SYBR Green Q−PCR. All primers except for GSFw800, producing the 500 bp amplicon were successfully optimized with diagnostic single peak melt curves. As such we preceded with two R_{amp} ratio primer sets the R_{amp} 380/120 and the R_{amp} 380/170.

Table 2.2 List of glnA primers used.

Primer	Sequence (5' → 3')	Orientation	Target	Experiment	Reference
GS1_new	GCTTGAGGATGCCGCCGATGT	Reverse	Bacterial glnA, all amplicons	Q-PCR and sequencing	This study, modified from Hurt et al., 2001
GSFw1200	GTTCGGGCATGCACGTGCA	Forward	Bacterial glnA, amplicon 1 (120 bp)	Q-PCR	This study
GS2_new	AAGACCGCGACCTTNATGCC	Forward	Bacterial glnA, amplicon 2 (170 bp)	Q-PCR	This study, modified from Hurt et al.,2001
GSfw900	GTCAARGGCGGYTAYTTCCC	Forward	Bacterial glnA, amplicon 3 (380 bp	Q-PCR and sequencing	This study
GSFw800	GAAGCCGAGTTCTTCSTCTTCG	Forward	Bacterial glnA, amplicon 4 (540 bp)	PCR	This study



$$R_{amp}\ 380/120 = \frac{35 - Ct\ Amp\ 3}{35 - ct\ Amp\ 1} \qquad R_{amp}\ 380/170 = \frac{35 - Ct\ Amp\ 3}{35 - ct\ Amp\ 2}$$

Figure 2.2 Schematic representation of primer binding sites along the Bacterial *glnA* gene. Primers are represented by arrows pointing to the right (forward primers) or to the left (reverse primer). The amplicons (Amp) generated by the different primer combinations are represented as coloured lines. The formulas used to calculate the two R_{amp} indexes are detailed under the figure.

2.3.2 Heat degradation

Incubation of RNA at 90°C had a strong and rapid impact on its integrity with a drop in the RIN from 7.5 to 4.7 after 10 min. At this point, the band corresponding to 23S rRNA had almost completely disappeared. Further exposure resulted in more pronounced degradation with accumulation of short RNA fragments and a RIN around 2 for both 45 min and 90 min exposure (Figure 2.3 A and C). One-way ANOVA revealed significant difference between all time points, except 45 and 90 min. A low and non-significant decrease in both R_{amp} indexes was

observed (−0.07 for 380/120 and − 0.11 for 380/170) between 0 and 10 min (Figure 2.3 C). This would tend to indicate that the R_{amp} was less sensitive than the RIN for monitoring RNA degradation by heat. However, interestingly the increase in Ct was also not significant for both *amoA* and *16S rRNA* between 0 and 10 min (Figure 2.3 B), showing that the R_{amp} reflected the outcome of the RT-Q-PCR assays better than the RIN. Further exposure to heat induced a more pronounced decrease in both R_{amp} (≈ −0.4 for 380/120 and ≈ −0.3 for 380/170) at 45 min compared with 0 min. Both R_{amp} indexes reached values around 0.15 at 90 min, which mapped well the behaviour of *amoA*, with a sharp increase in the Ct for this transcript between 10 and 45 min (≈4cts) and between 45 and 90 min (another ≈4cts). The *16S rRNA* transcript was also affected but to a smaller extent (increase in Ct of only ≈3ct between 0 and 90 min). Yet, in this case too, the increase was quite low between 0 and 10 min and sharper between 10 and 45 min and 45 and 90 min.

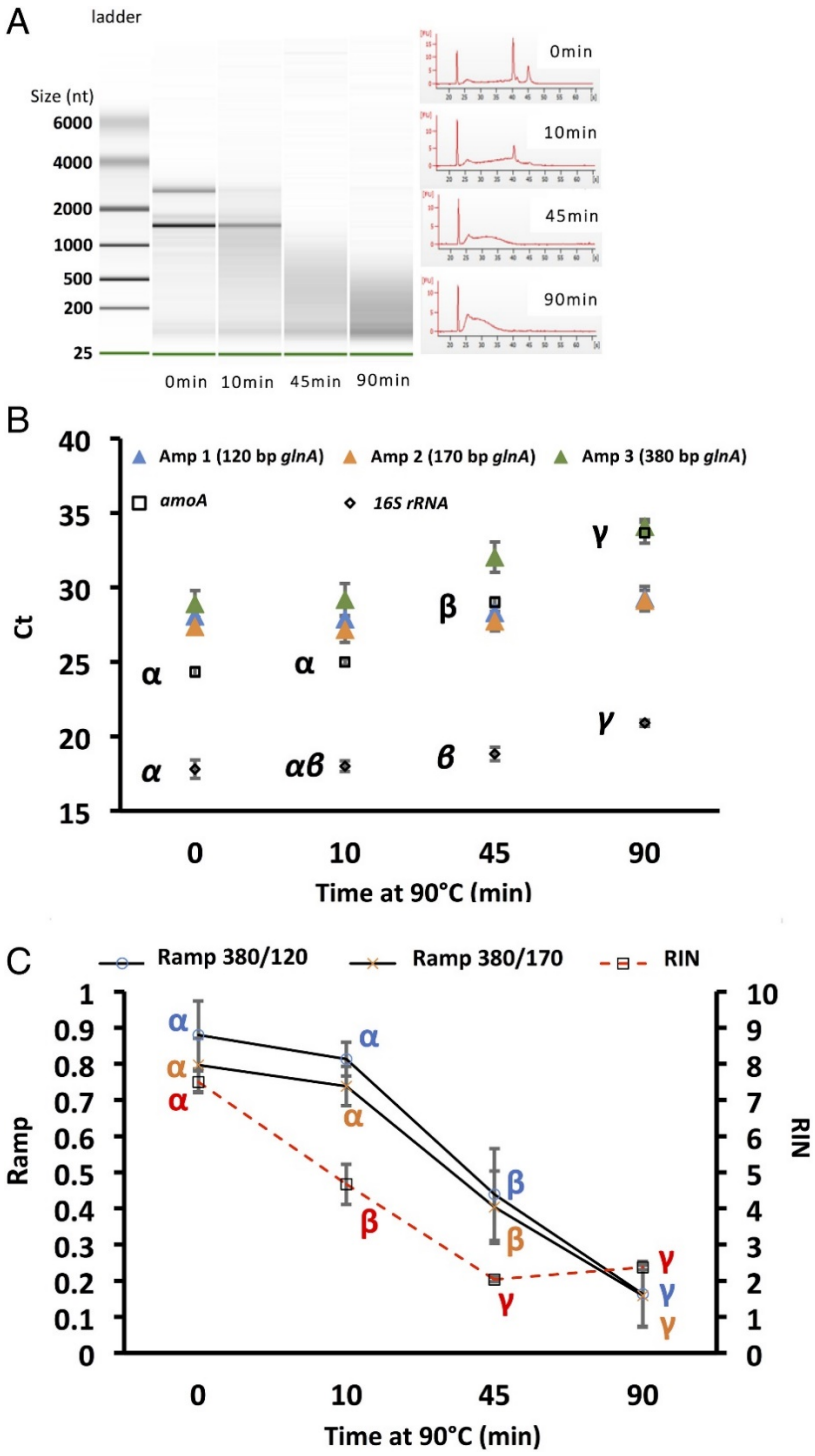


Figure 2.3 Effect of heat degradation on RNA integrity measured via the RIN (A), with RT-Q-PCR (B) and RIN versus R_{amp} (C). For RIN, RNA integrity visualized in virtual gels (A; left) and electropherogram (A; right) are displayed against incubation period at 90°C. RNA ladder shows size in nucleotides (nt). B. Effect of degradation on transcript quantification; Amp 1–3: average Ct (n = 3) of one of the three possible *glnA* amplicons; *amoA*: average *amoA* Ct (n = 3) of the Bacterial *amoA* transcript; *16S rRNA*: average *16S rRNA* Ct (n = 3) of the bacterial *16S rRNA* transcript. Effect of RNA degradation on R_{amp} index is presented in figure C; for comparison, RIN values were also plotted. Greek Letters indicate the result of TukeyHSD tests (points with different letters had values significantly different from each other using 0.05 as threshold for the p value).

2.3.3 UV degradation

The RIN was almost insensitive to UV radiation with an overall decrease of ≈1 at 90 min compared with 0 min (Figure 2.4 A and C). In contrast, UV radiation had a more pronounced effect on transcript quantification than heat as reflected by a quasi-linear increase in Ct of the *amoA* transcript between 0 and 45 min (Figure 2.4 B). Unlike heat exposure, 10 min under UV induced strong and significant increase in *amoA* Ct values (≈4cts). At 45 min, the Ct had increased by ≈9 compared with the starting point. After 90 min, the Ct of the *amoA* transcript almost reached 35, close to the detection limit. The Ct for *16S rRNA* transcript increased steadily from 18 at 0 min to 20 at 90 min, showing that this assay/transcript was less sensitive to UV degradation. The behaviour of the R_{amp}, again, mapped well onto *amoA* behaviour with a decrease of ≈0.2 after 10 min exposure for both indexes (although this was not significant) (Figure 2.4 C). A net decrease was observed at 45 min (≈ −0.6 compared with 0 min) and at 90 min both R_{amp} almost reached 0 since the Ct of the amplicon 3 *glnA* (380 bp) was very close to 35.

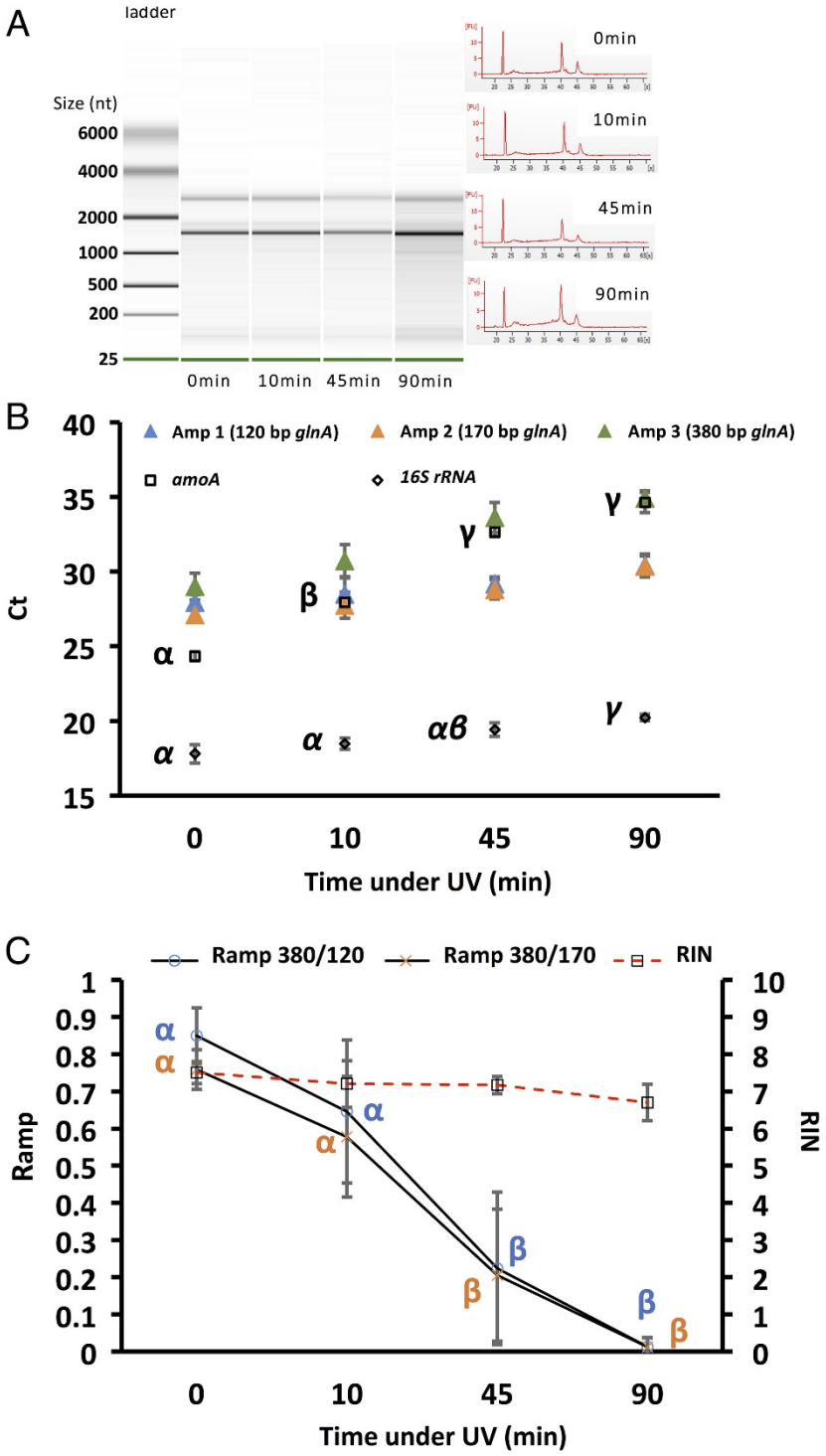


Figure 2.4 Effect of UV degradation on RNA integrity measured via the RIN (A), with RT-Q-PCR (B) and RIN versus R_{amp} (C). For RIN, RNA integrity visualized in virtual gels (A; left) and electropherogram (A; right) are displayed against incubation period under UV. RNA ladder shows size in nucleotides (nt).B. Effect of degradation on transcript quantification; Amp 1–3: average Ct (n = 3) of one of the three possible *glnA* amplicons; *amoA*: average *amoA* Ct (n = 3) of the Bacterial *amoA* transcript; *16S rRNA*: average *16S rRNA* Ct (n = 3) of the bacterial *16S rRNA* transcript. Effect of RNA degradation on R_{amp} index is presented in figure C; for comparison, RIN values were also plotted. Greek Letters indicate the result of TukeyHSD tests (points with different letters had values significantly different from each other using 0.05 as threshold for the p value).

2.3.4 Degradation by RNase I

The RIN showed a rapid response to *RNase I* degradation with a decrease from 7.1 to 6 between 0 and 2 U μg^{-1} (Figure 2.5 A and C.) as seen on virtual gels and electropherograms with an almost complete disappearance of the 23S rRNA. When using 10 U μg^{-1} and higher concentrations, the RIN decreased and remained stable at approximately 2.5 indicating advanced/almost complete degradation of the RNA. Complete destruction of both rRNA and an accumulation of small size RNA molecules on the electropherogram can be observed (Figure 2.5 A). In contrast, enzymatic degradation by *RNase I* had a relatively small effect on the Ct of the *amoA* transcript at low concentration (only 0.2 Ct increase between 0 and 2 U μg^{-1} treatments) (Figure 2.5 B). Ct values for *amoA* increased with greater degradation of the parent RNA (3 Cts difference at 10 and 20 U μg^{-1} and 5 Cts at 40 U μg^{-1} compared with 0 U μg^{-1} control). Of note, *amoA* transcripts were still quantified from the degraded 40 U μg^{-1} treatment with a mean Ct of 31.8. *RNase I* seemed to be the most effective treatment for the destruction of rRNA. Indeed, an increase of ≈ 3.2 Cts for the *16S rRNA* transcript was observed between 0 and 40 U μg^{-1} treatments whereas an increase of only 2.2 Cts was observed between 0 and 90 min for both physical degradation techniques (heat and UV). R_{amp} indexes were only slightly affected by 2 U *RNase I* μg^{-1} (decrease of ≈0.015 for 380/120 and ≈0.03 for 380/170) (Figure 2.5 C). The decrease was more pronounced for both R_{amp} at higher concentrations of *RNase I* (≈0.25 decrease at 20 U μg^{-1} compared with 0 U control). Even at concentrations as high as 40 U μg^{-1}, the R_{amp} indexes only reached 0.3. This indicated that at the high nuclease concentrations, even the small amplicons (120 and 170 bp) were starting to degrade. In this experiment, the R_{amp} 380/170 seemed to be more sensitive than the R_{amp} 380/120 in mapping RNA degradation, with significant differences between 0 and 10 U μg^{-1} treatments whereas R_{amp} 380/120 values only became significantly different from 0 U control from 20 U μg^{-1}. Again, as observed in the other degradation experiments, the behaviour of the *amoA* Ct was better reflected by changes in R_{amp}, especially R_{amp} 380/170, rather than by changes in the RIN.

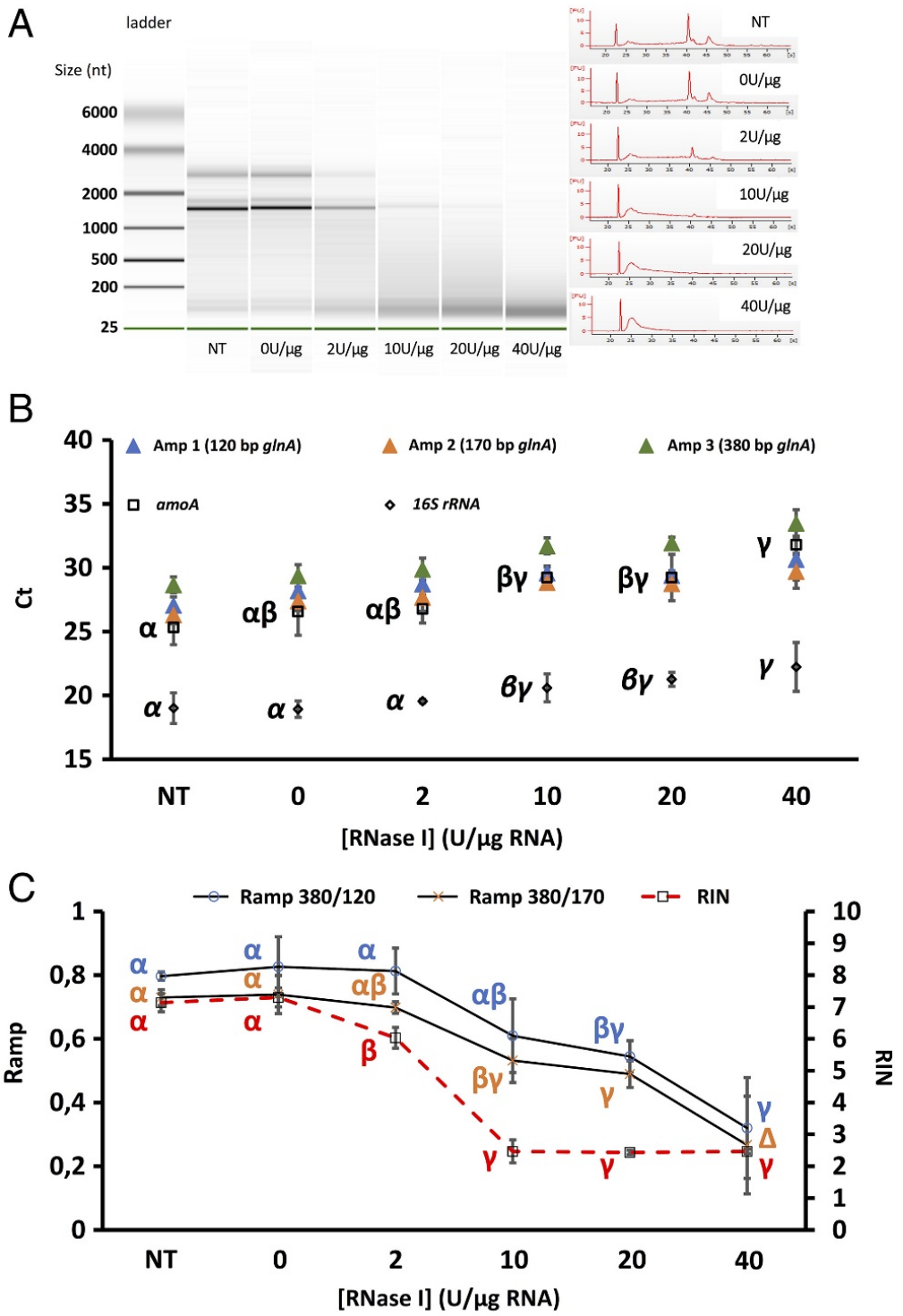


Figure 2.5 Effect of *RNase I* degradation on RNA integrity measured via the RIN (A), with RT-Q-PCR (B) and RIN versus R_{amp} (C). For RIN, RNA integrity visualized in virtual gels (A; left) and electropherogram (A; right) are displayed against incubation period with *RNase I*. RNA ladder shows size in nucleotides (nt). B. Effect of degradation on transcript quantification; Amp 1–3: average Ct (n = 3) of one of the three possible *glnA* amplicons; *amoA*: average *amoA* Ct (n = 3) of the Bacterial *amoA* transcript; *16S rRNA*: average *16S rRNA* Ct (n = 3) of the bacterial *16S rRNA* transcript. Effect of RNA degradation on R_{amp} index is presented in figure C; for comparison, RIN values were also plotted. Greek letters indicate the result of TukeyHSD tests (points with different letters had values significantly different from each other using 0.05 as threshold for the p value).

2.3.5 Effect of freeze/thaw cycles and storage

The effect of repeated cycles of freeze thaw on RNA is still poorly understood (and rarely studied) as conflicting results are reported, yet this is a common cause for concern when working with RNA. In our experiments, repeated freeze/thaw cycle (up to 10) did not induce any noticeable effects on RNA integrity, whether monitored via RIN or R_{amp} (data not shown). The effect of long-term storage was also investigated, by monitoring the RIN and R_{amp} of the same RNA after 0, 1 and 4 months stored at −80°C. No statistically significant change in RIN or R_{amp} was observed (data not shown).

2.3.6 Comparison between R_{amp} and RIN

Data generated from all of the degradation experiments undertaken (UV, heat and *RNase I*) was compiled to determine which of the two integrity indexes (RIN vs. R_{amp}) reflected the degradation status of the *amoA* and 16S rRNA transcripts more closely as determined by RT-Q-PCR. This was done by calculating Kendall correlations between either the R_{amp} or the RIN and the Cts of the two gene transcript targets (Figure 2.6). When considering all three degradation experiments, that is UV, heat and *RNase I*, the RIN was not significantly correlated with 16S rRNA nor *amoA* Ct values (p value > 0.05). In contrast, both R_{amp} ratios resulted in a significant correlation with both *amoA* and 16S rRNA transcripts (Figure 2.6). However, as the RIN was almost insensitive to UV, with a decrease of only about ≈1 after 90 min exposure (Figure 2.4), Kendall correlations were repeated without the inclusion of the UV data set. In this case, both the RIN and the R_{amp} were significantly correlated with 16S rRNA and *amoA* transcript abundances within the degraded RNA samples (Figure 2.6). In fact, the RIN was better correlated with *amoA* than 16S rRNA Cts. Nevertheless, both R_{amp} ratios were more highly correlated with *amoA* Cts than the RIN. Furthermore, the R_{amp} ratios were more highly correlated with the 16S rRNA than the RIN. Taken together, these two observations confirm that the R_{amp} indexes better reflected RT-Q-PCR changes induced by RNA degradation than the RIN.

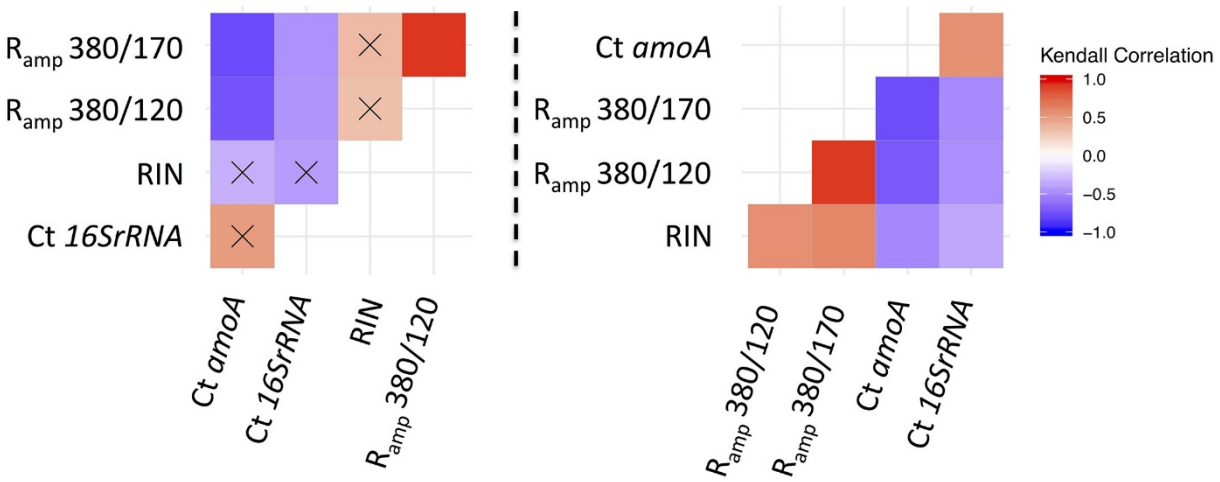


Figure 2.6 Kendall correlations between integrity indexes and Cts of the two reference gene used in this study. The correlations coefficients were calculated using all data generated from UV, heat and *RNase I* degradation experiments (left) and from the heat and RNase I only (right). The colour of squares represents the value of the correlation coefficients as explained on the colour scale. Black crosses indicate absence of significant correlation (threshold: *p* value > 0.05).

2.3.7 Effect of RNA degradation on transcript community composition

RNA degradation impacted upon *amoA, glnA* and *16S rRNA* gene quantification, as demonstrated previously. However, whether all members of the community were affected equally was still to be determined. To answer this question, cDNA amplicons of the Bacterial 16S rRNA, *amoA* and *glnA* transcripts underwent Illumina MISeq amplicon sequencing from all degradation points of the *RNase I* experiment representing RNA with RIN values from 7.5 to 2.4 and R_{amp} values from ≈0.8 to ≈0.3 and from ≈0.7 to ≈0.3 for R_{amp} 380/170 and R_{amp} 380/120 respectively. The effect of *RNase I* treatment on community evenness was tested using PERMANOVA. Results are presented in Figures 2.7, 2.8 and 2.9. Interestingly, the community structure of the three transcripts studied responded differently.

Strikingly, *RNase I* treatment had little effect on *16S rRNA* transcript community evenness (Figure 2.7). Indeed, for individual OTU, none of the members of the community were significantly differentially represented (*p* value $\log_2$ difference > 0.05) within highly degraded samples in comparison to controls (Figure 2.7). For individual OTU, at least 90% had their relative expression change over the degradation experiment fall within the [−log2(1.5); log2(1.5)] interval, even when comparing controls to the completely degraded 40 U *RNase I* sample (Figure 2.7 B). This indicates that *16S rRNA* OTU transcript community was responding evenly to degradation, with each member having the same chance to be affected regardless of its abundance or sequence.

For bacterial *amoA* transcript community, there was no change in the overall composition with increasing degradation as reflected by the non−significant PERMANOVA (p value > 0.05) (Figure 2.8). However, with increasing degradation, there was an increasing difference in the community evenness among replicates. Furthermore, unlike 16S rRNA transcripts, when examining individual *amoA* OTUs it was evident that in the degraded samples some OTUs were differentially represented at a significant level compared with controls (Figure 2.8 B). In fact, some OTUs in the highly degraded samples (10, 20 and 40 U *RNase I*) had a fold change difference of up to two orders of magnitude compared with the controls and in most cases, resulting in their over representation in degraded samples. Moreover, in the more highly degraded treatments (10, 20 and 40 U *RNase I*), up to 44% of *amoA* OTUs had their relative expression outside the [−log2(1.5); log2(1.5)] interval, compared with the starting RNA (Figure 2.8 B). So, while there was not an overall significant difference in *amoA* community structure with increasing RNA degradation, there were changes in the relative expression of individual OTU. The overall lower numbers of amoA OTUs for comparison and the increasing difference among replicates in the degraded samples may in fact explain the lack of overall statistical significance in community structure.

The effect of *RNase I* treatment was much more pronounced for *glnA* transcripts, than for *amoA,* and a significant change in community composition with increasing degradation was observed (p value < 0.05 for PERMANOVA with both Bray−Curtis and Unifrac distances) (Figure 2.9 A and B). As seen with *amoA*, the difference in community composition between replicates also increased with increasing *RNase I* treatment. Moreover, this effect was also observed at individual OTU level with a large fraction of the individual OTU showing different expression levels in treated samples compared with controls (Figure 2.9 B). As seen for *amoA*, some *glnA* OTUs were highly over represented in degraded samples by 2 to 3 orders of magnitudes, *e.g.* when comparing the untreated samples (NT) to the 40U*RNase* samples, 0.28% (3 sequences) were over represented by two orders of magnitude. When comparing the samples treated with buffer only to the 40U*RNase* samples, 2.43% (19 sequences) were over represented by two orders of magnitude and 0.13% (1 sequence) by three orders of magnitude.

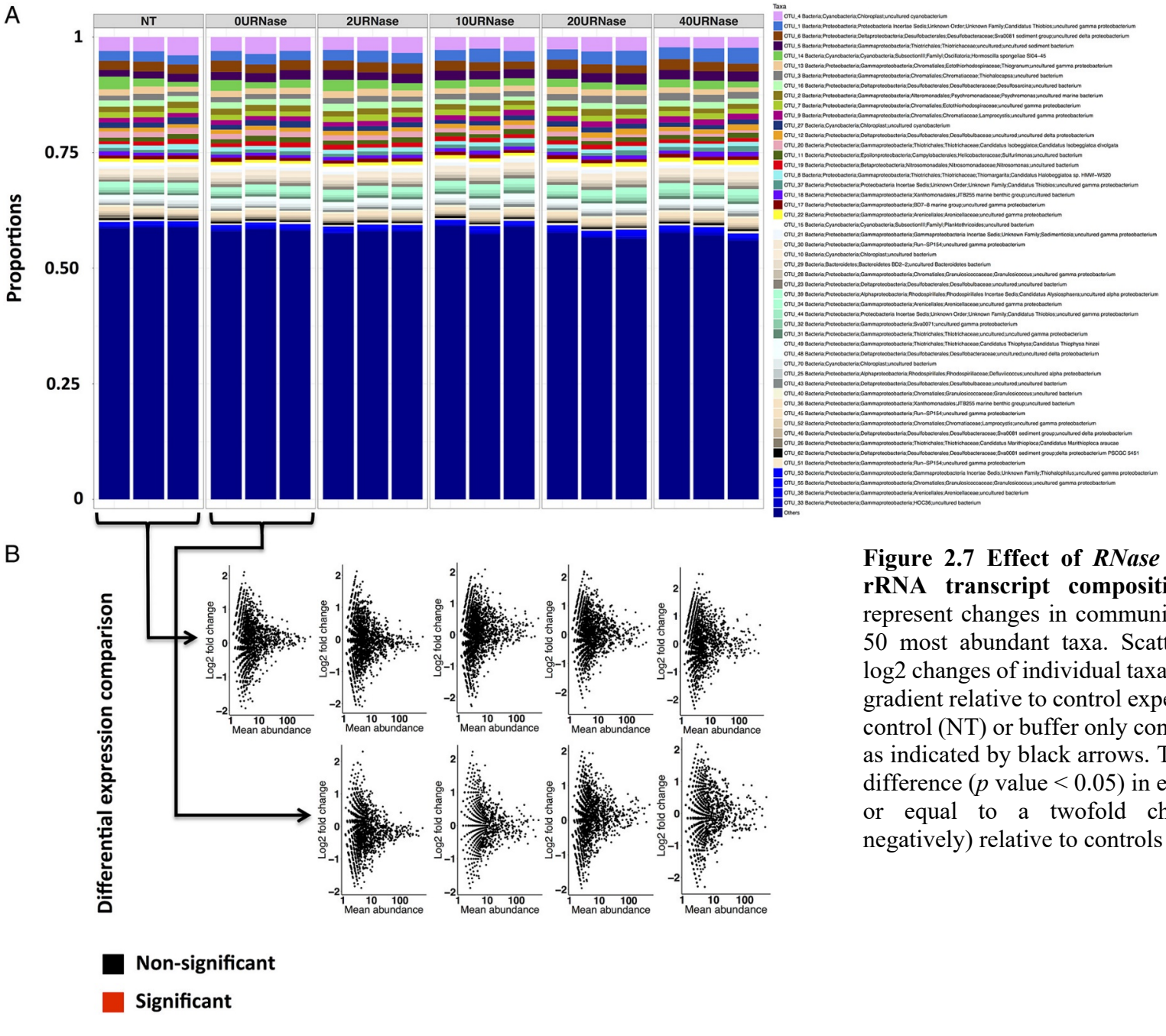


Figure 2.7 Effect of *RNase I* treatment on 16S rRNA transcript composition. Bar charts (A) represent changes in community composition of the 50 most abundant taxa. Scatterplots (B) represent log2 changes of individual taxa along the degradation gradient relative to control experiments (no treatment control (NT) or buffer only control (0U*RNase I* μl^{-1})) as indicated by black arrows. Taxa with a significant difference (p value < 0.05) in expression greater than or equal to a twofold change (positively or negatively) relative to controls are indicated in red.

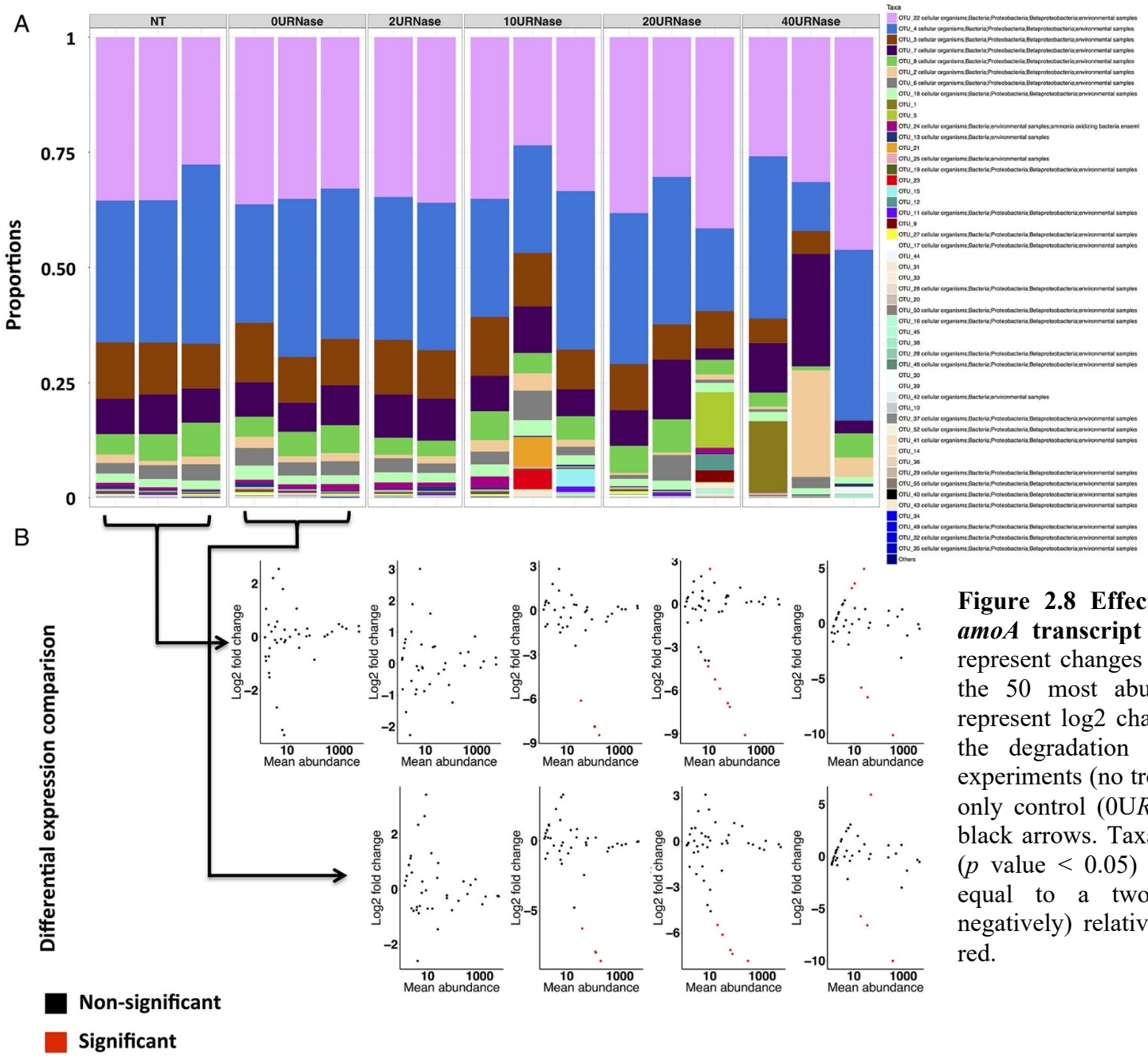


Figure 2.8 Effect of *RNase I* treatment on *amoA* transcript composition. Bar charts (A) represent changes in community composition of the 50 most abundant taxa. Scatterplots (B) represent log2 changes of individual taxa along the degradation gradient relative to control experiments (no treatment control (NT) or buffer only control (0U*RNase I* μl^{-1})) as indicated by black arrows. Taxa with a significant difference (p value < 0.05) in expression greater than or equal to a twofold change (positively or negatively) relative to controls are indicated in red.

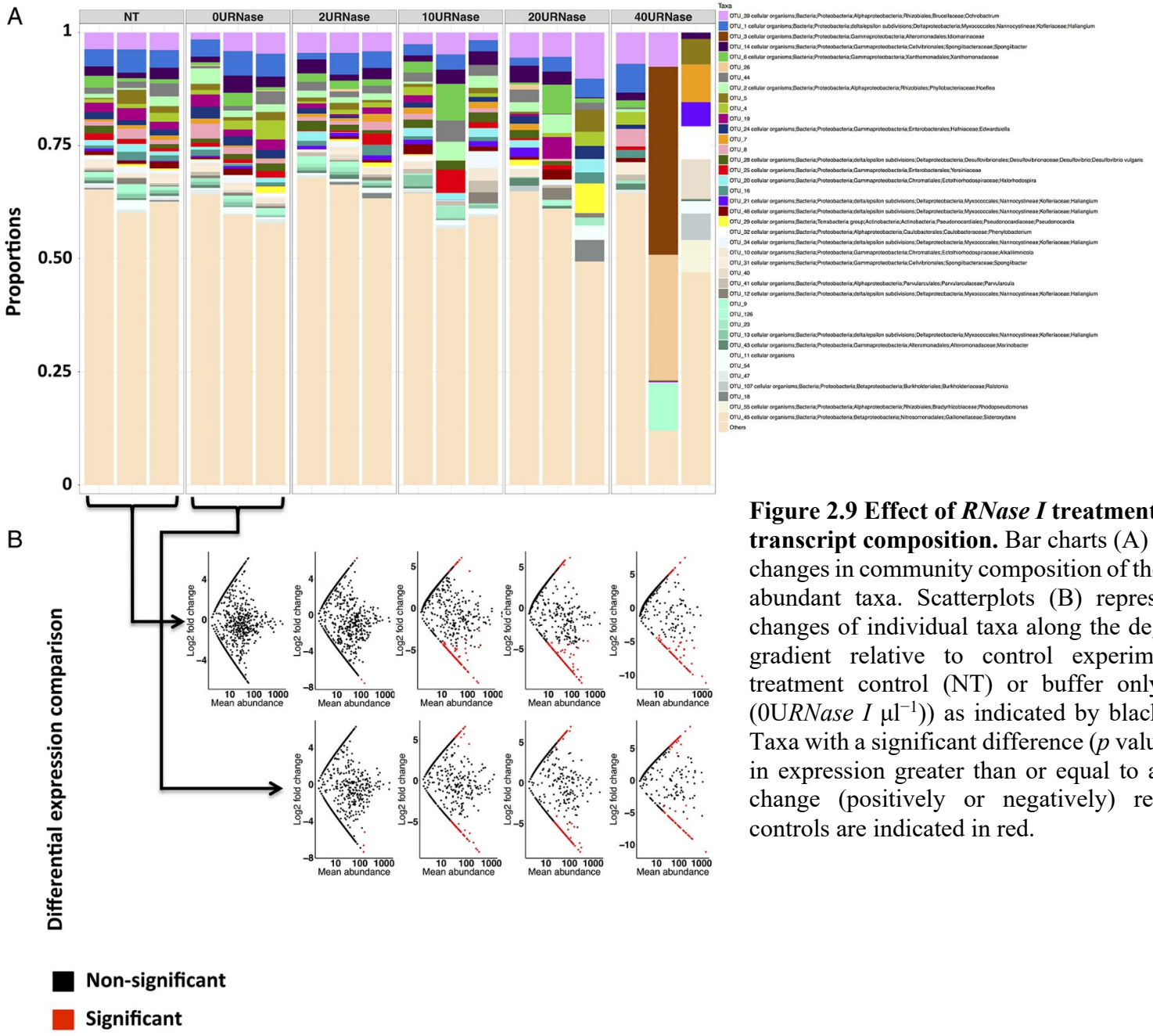


Figure 2.9 Effect of *RNase I* treatment on *glnA* transcript composition. Bar charts (A) represent changes in community composition of the 50 most abundant taxa. Scatterplots (B) represent log2 changes of individual taxa along the degradation gradient relative to control experiments (no treatment control (NT) or buffer only control (0U*RNase I* μl^{-1})) as indicated by black arrows. Taxa with a significant difference (p value < 0.05) in expression greater than or equal to a twofold change (positively or negatively) relative to controls are indicated in red.

2.4 Discussion

Here, we successfully designed and tested the Ratio Amplicon, $R_{amp,}$ index. The concept is that as RNA degrades, longer strands are preferentially affected and the abundance of the longer amplicon relative to the shorter amplicon will decrease with increasing RNA degradation (Björkman *et al*., 2016). Using experimentally degraded environmental RNA, we have shown that the newly developed R_{amp} index was a better predictor of the Ct of the target mRNA transcript used in this study, *amoA,* than the ribosome−based RIN approach. In fact, when data from the three degradation experiments carried out was considered together only the R_{amp} statistically correlated with *amoA* Cts. As the RIN failed to detect UV degradation, we removed this data from the correlation calculation to determine if this data set was biasing the results towards the R_{amp} approach. In this case, there was also a significant correlation between the RIN and *amoA* Ct (−0.51). However, the R_{amp} index still reflected the fate of the mRNA better than the RIN (−0.72 and −0.77 for R_{amp} 380/120 and R_{amp} 380/170 respectively).

Taking the different RNA degradation approaches used individually, the RIN and R_{amp} ratios responded differently. As noted above, the RIN did not change over a 90-min exposure to UV. UV causes intramolecular cross-linking of thymines but does not cause strand breaks (Kladwang *et al*., 2012) while the RIN monitors stand break. Similar results were obtained by Bjorkman *et al* (Björkman *et al*., 2016) who reported a lack of response for the RIN and the RQI when human RNA preparations were degraded by UV radiation, even after 120 min of exposure. As such RNA damage by UV cannot be detected by electrophoresis separation but is recorded by RT-Q-PCR R_{amp} index. Other RNA degradation processes that result in base destruction but not necessarily strand break include oxidative damage (Rhee *et al*., 1995) or chemically-induced radical formation (Hawkins and Davies, 2002).

In contrast, the RIN was the most efficient method to detect heat degradation. There was a strong and significant decrease in this index after 10 min whereas the R_{amp} indexes only became significantly different from the controls after 45 min. Moreover, there was very little effect on the direct quantification of the transcripts by RT-Q-PCR with very little change in the Ct of either *amoA* or 16S rRNA in the first 10 min at 90°C. Initially, heat degradation caused a rapid decrease in the RIN. However, at this point the RT-Q-PCR targets were actually responding more slowly and were more closely mapped by the R_{amp} than the RIN. Björkman *et al*., 2016 showed a similar response of their differential amplicon, the ΔΔamp index, that did not change

much between 2 and 10 min at 95°C whereas the RIN rapidly reduced from 7 to 2. Moreover, Gingrich *et al* (Gingrich *et al*., 2008) showed that transcripts could be quantified from RNA preparations incubated at 90°C for several hours. This relatively low impact of heat on RNA quantification may be due to modification of RNA secondary structures which could result in more efficient cDNA synthesis and mask the effect of the heat-induced reduction of RNA integrity. More likely it is due to the small amplicon size of the targets that are unaffected by degradation. This essentially illustrates the difference in the methods used to monitor RNA degradation – the RIN detects strand break no matter where the fracture occurs along the transcript while the R_{amp} will only detect degradation if the break occurs between primer binding sites.

RNA degradation using the nuclease enzyme *RNase I* was monitored using both RIN and R_{amp}. A similar behaviour could be observed here as in the heat degradation experiment with the RIN responding more quickly but loosing sensitivity when RNA was highly degraded whereas the R_{amp} responded slightly later but remained sensitive when RNA was extensively degraded. *RNase I* was the degradation method that had the strongest effect on the 16S rRNA Ct. *RNase I* activity is dependent on the concentration of the substrate. If rRNA and mRNA are considered as two distinct substrates, it can be expected that *RNase I* will have a greater impact on ribosomes as they constitute 80%–85% of total RNA. Furthermore, cDNA synthesis from mRNA would be enhanced in preparations where rRNA was depleted (Petrova *et al*., 2017). This dynamic may mask and change the effect of degradation over time, which would explain the relatively low increase in Ct for *amoA* at the beginning of the *RNase I* degradation experiment. Nevertheless, in this experiment and generally, for all degradation tests carried out, the behaviour of the *amoA* Ct was better predicted by the R_{amp}, as reflected by the higher correlation coefficient between R_{amp} indexes and *amoA* Ct than the RIN (Figure 2.6). As the *in vitro* half-life of different transcripts is not well-understood and has been shown to vary (Selinger *et al*., 2003; Belasco, 2010; Evguenieva-Hackenberg and Klug, 2011) further work is required to test the correlation of the R_{amp} against a larger range of mRNAs. For ribosomal RNA, while the correlation between the R_{amp} index and 16S rRNA Ct was lower than for *amoA,* it still correlated better with RNA degradation than the RIN. This indicates that the outcome of 16S rRNA analysis was less affected by degradation than our mRNA targets. There are two factors that may contribute to this, the reported greater robustness of ribosomal RNA than mRNA and the shorter (~103 bp) 16S rRNA amplicon. That ribosomal RNAs behave the same as mRNA has never been proven. On the contrary, Sidova and colleagues (2015) showed that

when natural *post mortem* degradation occurs, rRNA is more stable than mRNA. In this case, rRNA is a poor predictor of degradation of the mRNA fraction, as supported by this work. As mRNA is subjected to more rapid decay to adjust to the needs of the cell whereas rRNA are degraded only under certain stress conditions or when defective (Deutscher, 2006), then these intrinsic differences in stability properties may also affect degradation rates of the different class of RNA post-extraction. Therefore, based on this work we can conclude that the R_{amp} was a better predictor of mRNA integrity than the RIN. However, as we and others (Björkman *et al*., 2016) have shown RNA responds differently to different types of degradation e.g. strand break versus intramolecular cross-linking of thymines, and as the exact and likely multiple causes of post-extraction degradation are unknown, we recommend that the RIN is used in conjunction with the R_{amp} to monitor RNA integrity.

Which R_{amp} to use?

In theory, the greater the difference between the two amplicons the more sensitive the R_{amp} index. However, as the R_{amp} approach is based on RT-Q-PCR it is restricted by the presence of conserved sites for primer design, and the success and efficiency of the RT and qPCR reactions. We initially designed a 500 bp *glnA* PCR amplicon however, the Q-PCR assay failed to produce a single diagnostic melt curve analysis. Of the remaining shorter R_{amp} sets, in practice, only one R_{amp} index is necessary, we recommend using the R_{amp} 380/170. The R_{amp} 380/170 always had a higher value than the R_{amp} 380/120, which would indicate that the number of 170 bp targets is higher than the 120 bp. Since both are amplified from the same target, this is not possible and the explanation for this observation is the lower efficiency of the 120 bp Q-PCR compared with the 170 bp assay. In spite of this, both R_{amp} correlated similarly well overall with each degradation experiment, with R_{amp} 380/170 slightly more sensitive in the *RNase I* experiment.

Impact of experimental degradation of environmental RNA on ribosomal (16S rRNA) and mRNA (amoA and glnA) community diversity

For complex environmental communities, the integrity of RNA is not only important to evaluate quantitative gene expression but is also of significance if it adversely affects the relative abundance of transcript diversity. To examine this, we assessed changes in the community structure of the 16S rRNA, *amoA* and *glnA* transcripts from all fractions of the *RNase I* sequentially degraded RNA.

The results were surprising with successful amplicon sequencing even from highly degraded samples. Nevertheless, the data did suggest a different response of *16S rRNA* and mRNA transcripts to degradation, with *16S rRNA* community structure unaffected over the range of degraded RNA samples. That is a statistically similar community was present in the control non-degraded samples as in the totally destroyed 40 Units *RNase I* (with a mean RIN of 2.5 and R_{amp} of 0.32 and 0.27 for R_{amp} 380/120 and R_{amp} 38/170 respectively). This indicates that while total RNA was degraded, the small transcript fragments required for RT-PCR and amplicon sequencing remained intact. In fact, so much so that no significant change in the relative abundance of individual OTU was observed.

On the other hand, RNA degradation had a greater influence on both *amoA* and *glnA* mRNA targets. While, again surprisingly, transcript amplicons were successfully detected from all degradation status samples, greater variability between degraded replicates was observed. This resulted in statistically different communities for *glnA* but not *amoA* when compared with the same non−degraded control samples. However, the low number of *amoA* OTUs and increased variability between replicates contributed to the lower statistical power resulting in no statistical difference between treatments (Figure 2.8). Furthermore, there were significant, sometimes up to two to three orders of magnitude change in the relative abundance of individual *glnA* and *amoA* OTUs in the degraded samples versus control samples. So, while we could successfully amplify mRNA transcripts from degraded environmental samples, we have shown that the relative composition of the community members was adversely affected by degradation and was not representative of the initial starting point. While further work is needed to determine the impact of degradation across the entire transcriptome to see if all mRNAs respond in a similar manner, it is clear from our mRNA amplicon sequencing that RNA degradation will alter the outcome of community analysis. It is therefore necessary to ensure the RNA integrity of the sample is known prior to interpretation of results. For this, our data indicates that a combination of approach targeting both ribosomal (the RIN) and mRNA (the R_{amp}) is needed.

Chapter III Reverse Transcriptase Enzyme and Priming Strategy Affect Quantification and Diversity of Environmental Transcripts

Cholet, F., Ijaz, U.Z., and Smith, C.J. (2020) Reverse transcriptase enzyme and priming strategy affect quantification and diversity of environmental transcripts. *Environ. Microbiol. (2020) 22(6), 2383–2402 doi:10.1111/1462-2920.15017*

Fabien Cholet contributed to the study design, experimental work, data analysis and report writing.

environmental microbiology

Environmental Microbiology (2020) **22**(6), 2383–2402 doi:10.1111/1462-2920.15017

Reverse transcriptase enzyme and priming strategy affect quantification and diversity of environmental transcripts

Fabien Cholet,* Umer Z. Ijaz and Cindy J. Smith
Infrastructure and Environment Research Division, James Watt School of Engineering, University of Glasgow, Glasgow, Scotland, G12 8LT, UK.

Summary

Reverse-transcriptase-quantitative PCR (RT-Q-PCR) and RT-PCR amplicon sequencing, provide a convenient, target-specific, high-sensitivity approach for gene expression studies and are widely used in environmental microbiology. Yet, the effectiveness and reproducibility of the reverse transcription step has not been evaluated. Therefore, we tested a combination of four commercial reverse transcriptases with two priming techniques to faithfully transcribe *16S rRNA* and *amoA* transcripts from marine sediments. Both enzyme and priming strategy greatly affected quantification of the exact same target with differences of up to 600-fold. Furthermore, the choice of RT system significantly changed the communities recovered. For *16S rRNA*, both enzyme and priming had a significant effect with enzyme having a stronger impact than priming. Inversely, for *amoA* only the change in priming strategy resulted in significant differences between the same samples. Specifically, more OTUs and better coverage of *amoA* transcripts diversity were obtained with GS priming indicating this approach was better at recovering the diversity of *amoA* transcripts. Moreover, sequencing of RNA mock communities revealed that, even though transcript α diversities (i.e., OTU counts within a sample) can be biased by the RT, the comparison of β diversities (i.e., differences in OTU counts between samples) is reliable as those biases are reproducible between environments.

Received 16 September, 2019; revised , ; accepted 8 April, 2020.
*For correspondence.E-mail f.cholet.1@research.gla.ac.uk

Introduction

Whereas modifications of the genome can reflect adaptations of living organisms over evolutionary time scales, changes in the transcriptome reflect short-term responses of cells (López-Maury *et al*., 2008; Browning and Busby, 2016). In environmental microbiology, transcriptomics is essential to understanding which biochemical pathways are triggered by environmental conditions at a given time. RNAseq approaches facilitate primer free metatranscriptomics to reveal global gene expression profiles. It is now a widely used method in environmental microbiology and has allowed scientists to gain formidable insight into the genome-scale mechanisms used by microbes to adapt to changing environmental conditions (Shakya *et al*., 2013; Gutleben *et al*., 2018). However, it generally comes at high cost and requires extensive data analysis. Plus, as an untargeted approach, it may require enrichment of the mRNA (via removal of ribosomal RNA) and will be dependent on sequencing depth to reveal rare transcripts among the diverse array of transcripts expressed in complex environmental samples. In contrast, reverse-transcriptase-quantitative PCR (RT-Q-PCR) is directed via primers towards a single target. While this is much lower throughput in terms of a global overview of transcription, this approach facilitates transcript quantification that is specific, with high-sensitivity and low-detection limits over a wide dynamic range (Sanders *et al*., 2014). RT-Q-PCR is high-throughput in terms of sample numbers, cost effective (in comparison to metatranscriptomics) and subsequent data processing is fast without the requirement for high computational power and bioinformatic expertise needed for metatranscriptomics analysis. As a consequence, RT-(Q)-PCR is routinely used in most life science research fields including environmental microbiology to target and quantify specific transcripts.

In environmental microbiology it is an invaluable approach to further link microbial activity, via gene expression, to microbial and ecosystem processes, compared to DNA approaches alone (Smith and Osborn, 2009; Saleh-Lakha *et al*., 2011; Gadkar and Filion, 2013). As a result the approach has been used to

3.1 Reverse-Transcription: the Information Conservation Challenge

In the previous chapter, we have shown that RNA degradation is a major issue in environmental microbiology. Only RNA of high quality should be used as template for the RT reaction. Still, RNA extractions are complex mixtures of ribosomal, messenger, transfer and small RNAs. This represents a huge amount of information, of which only a small and diverse fraction, the mRNAs, informs on gene expression. The challenge when converting RNA into cDNA via the Reverse-Transcription (RT) reaction, is therefore to conserve this information as best as possible, both quantitatively (The proportions of different RNAs should be represented in the cDNA) and qualitatively (the same nucleotides that constitute the RNA should be found in the cDNA) (Figure 3.1).

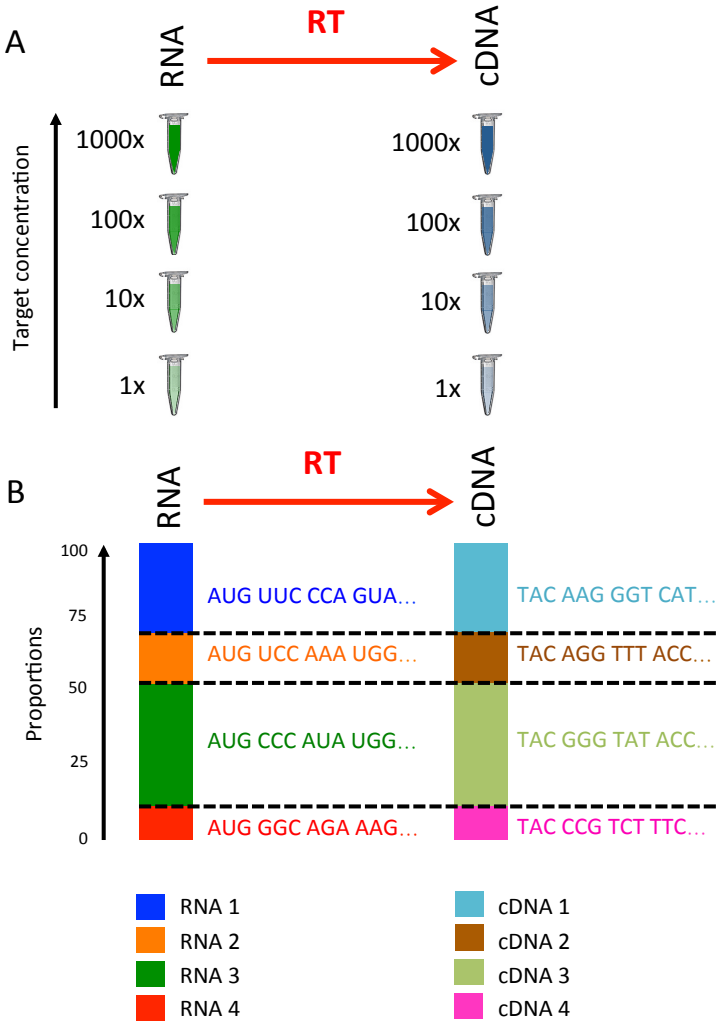


Figure 3.1 The information conservation challenge of the RT reaction. In this figure, perfect quantitative (A) and qualitative (B) information conservation is represented.

The RT reaction requires a reverse transcriptase enzyme, of which there are a number commercially available, and a reverse primer to initiate the RT reaction. There are two main priming strategies, random or gene specific priming. For random priming, short oligonucleotides (*e.g.* hexamer or decamer) consisting of all possible sequence combinations for that size, are used to randomly initiate the RT across the entire transcriptome. Gene specific, as the name implies, target specific transcripts of interest. A number of studies in the wider field of molecular biology indicate that the RT reaction has a significant impact on the final results for the same sample. Indeed within clinical studies, the inherent variability of cDNA synthesis has been reported in some cases to be greater than the differences between biological samples (Sanders *et al.*, 2014). This level of variability implies that comparison of results between different studies using different approaches is near impossible (Bustin, 2002). Moreover, the sources of RT variability have been attributed to a wide range of factors including: the choice of reverse transcriptase (Ståhlberg *et al.*, 2004; Stangegaard *et al.*, 2006; Levesque-Sergerie *et al.*, 2007; Werbrouck *et al.*, 2007; Okello *et al.*, 2010; Sieber *et al.*, 2010; Miranda and Steward, 2017); priming (Lekanne Deprez *et al.*, 2002; Ståhlberg *et al.*, 2004; Stangegaard *et al.*, 2006; Werbrouck *et al.*, 2007; Sieber *et al.*, 2010; Miranda and Steward, 2017); background RNA concentration (Bustin & Nolan, 2004; Levesque-Sergerie et al., 2007; Miranda & Steward, 2017); cleaning of the RT reaction (Okello *et al.*, 2010); RNaseH treatment (Polumuri et al., 2002); RT reaction composition and conditions (Ståhlberg *et al.*, 2004; Werbrouck *et al.*, 2007) and dilution of cDNA (Smith et al 2006).

In environmental microbiology applications, the effect of the initial RT reaction on quantification and amplicon sequencing of environmental transcripts has yet to be determined. Indeed environmental samples may provide a number of further challenges for efficient and reproducible RT reactions due to the presence of co-extracted inhibitors; variable target expression (high to low) in a background of high non-target template concentration and low RNA quality and integrity (Cholet *et al.*, 2019). Moreover, there is the need for the RT reaction to faithfully transcribe the diversity of target of interest. A small number of studies investigating primer-free approaches to characterise *16S rRNA* transcripts revealed better accuracy (Mäki and Tiirola, 2018) and sensitivity (Hoshino and Inagaki, 2013) with primer-free approaches for amplicon sequencing than PCR of the cDNA. Nonetheless, these primer-free approaches still rely on an initial RT reaction, which could impact the outcome of the *16S rRNA* transcript sequencing. Moreover, our own personal observations in the laboratory have indicated that RT enzyme and priming strategy greatly impact the results of environmental

transcript studies, often meaning the difference between detection or not of a given transcript that in turn results in different ecological interpretation.

Therefore, to improve reproducibility and inform best practice and standardisation of RT-(Q)-PCR approaches in environmental microbiology, we have undertaken a detailed study of the effect of the RT reaction on RNA extracted from environmental samples. We aimed to determine the impact of enzyme and priming strategy on quantification and amplicon sequencing of transcripts (spiked artificial RNA, *16S rRNA* and ammonia monooxygenase (*amoA*)). We therefore examined a combination of four commonly used commercial reverse transcriptases (Superscript III, Superscript IV, Omniscript and Sensiscript; designated SSIII, SSIV, Omni and Sensi, respectively, thereafter) and two priming strategies (random hexamer and gene specific; designated RH and GS, respectively, thereafter). We hypothesized that both quantification and alpha diversity (*i.e.* OTU counts within a sample) of transcripts from the same samples will be affected by RT enzyme and priming strategy.

.

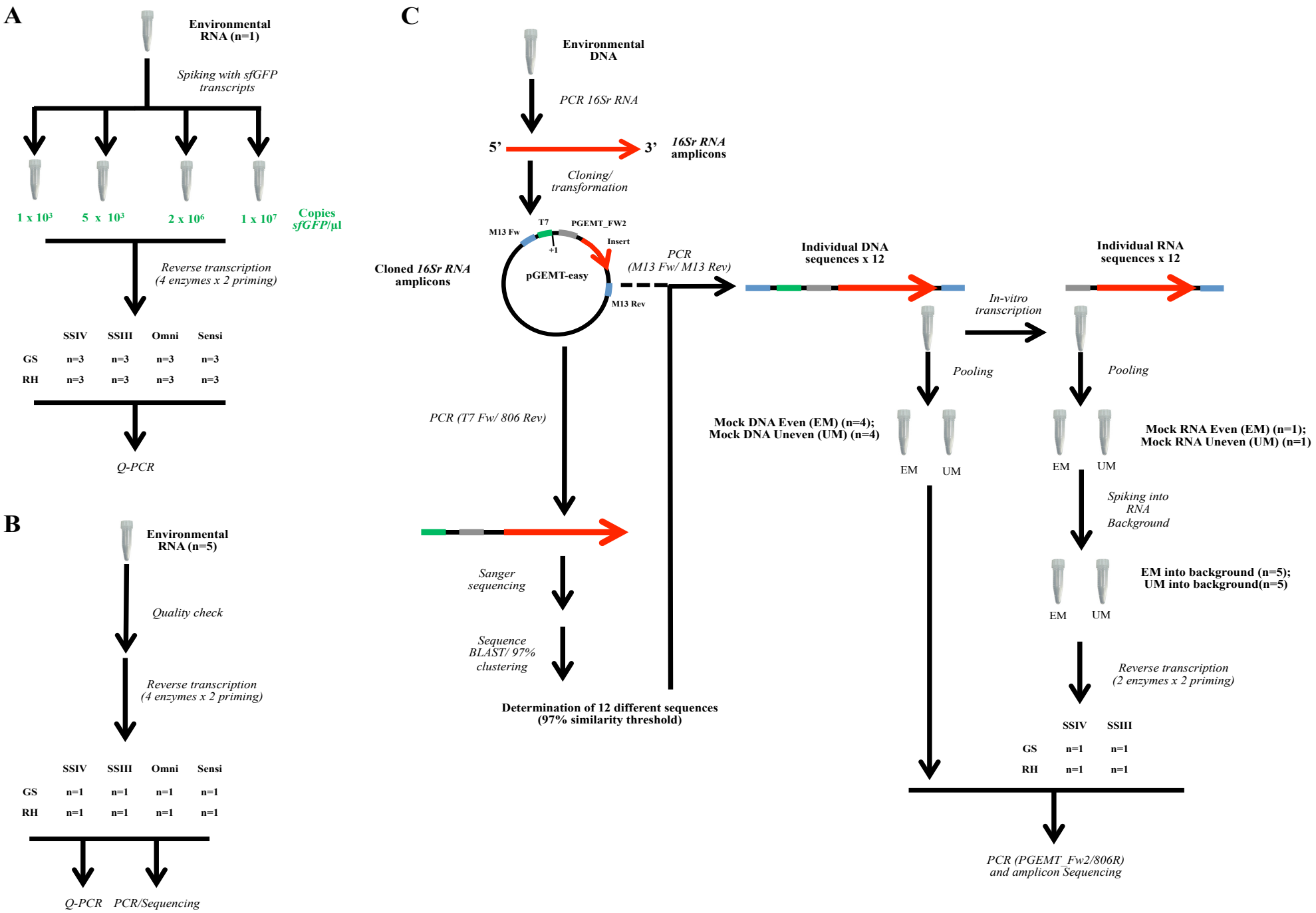

A
Environmental RNA (n=1)
Spiking with sfGFP transcripts
1 x 10³
5 x 10³
2 x 10⁶
1 x 10⁷
Copies sfGFP/µl
Reverse transcription (4 enzymes x 2 priming)
SSIV SSIII Omni Sensi
GS n=3 n=3 n=3 n=3
RH n=3 n=3 n=3 n=3
Q-PCR
B
Environmental RNA (n=5)
Quality check
Reverse transcription (4 enzymes x 2 priming)
SSIV SSIII Omni Sensi
GS n=1 n=1 n=1 n=1
RH n=1 n=1 n=1 n=1
Q-PCR
PCR/Sequencing
C
Environmental DNA
PCR 16Sr RNA
5'
3'
16Sr RNA amplicons
Cloning/ transformation
Cloned 16Sr RNA amplicons
M13 Fw
T7
PGEMT_FW2
Insert
+1
pGEMT-easy
M13 Rev
PCR (M13 Fw/ M13 Rev)
PCR (T7 Fw/ 806 Rev)
Sanger sequencing
Sequence BLAST/ 97% clustering
Determination of 12 different sequences (97% similarity threshold)
Individual DNA sequences x 12
Pooling
Mock DNA Even (EM) (n=4);
Mock DNA Uneven (UM) (n=4)
EM
UM
In-vitro transcription
Individual RNA sequences x 12
Pooling
EM
UM
Mock RNA Even (EM) (n=1);
Mock RNA Uneven (UM) (n=1)
Spiking into RNA Background
EM into background (n=5);
UM into background(n=5)
EM
UM
Reverse transcription (2 enzymes x 2 priming)
SSIV SSIII
GS n=1 n=1
RH n=1 n=1
PCR (PGEMT_Fw2/806R) and amplicon Sequencing

Figure 3.2 Schematic representation of the experimental workflow followed in chapter III. The effect of the RT reaction was evaluated on: A) the quantification of an exogenous transcript spiked at known concentrations; B) the quantification of two endogenous transcripts and the subsequent sequencing of these transcripts and C) the sequencing of mock communities composed of 12 transcripts with known sequences for this last experiment, DNA mocks were also included as controls. Replicates are indicated by "n=".

3.2 Methods

3.2.1 Sediment Samples

[See 2.3.1 for origin of the samples]. Five biological replicate sediments, designated Env1, Env2, Env3, Env4 and Env5 respectively were used for testing the effect of the RT reaction on RT-Q-PCR and RT-amplicon sequencing of the endogenous *amoA* and *16S rRNA* transcripts. An additional sample was used for preparing the RNA background for the *sfGFP* spiking experiment

3.2.2 RNA Preparation from Sediment

[Same as 2.3.3]

3.2.3 RNA quality check

The quality, purity and integrity of extracted environmental RNA was determined as follows: *Quantity/purity*: Total RNA was quantified using three different approaches: spectrophotometry (NanoDrop; Life Technologies), fluorometry (Qubit broad Range RNA; Life Technologies) and microfluidics (Bioanalyser 2100 RNA Nano; Agilent Technologies). Purity was determined by spectrophotometry (NanoDrop; Life Technologies) with the 260nm/230nm and 260nm/280nm band absorption ratios.

Integrity: RNA integrity was determined using two different approaches: the RNA Integrity Number (RIN), based on the 23S/16S rRNA ratio and the electropherogram of the extracted RNA (Bioanalyser 2100 RNA Nano; Agilent Technologies) and the R_{amp} approach, based on the differential amplification of glnA mRNA amplicons of different length (Chapter II and Cholet *et al.,* 2019).

3.2.4 Evaluation of RT reaction on transcript quantification via a *sfGFP* RNA spike

3.2.4.a Preparation of the sfGFP RNA spike and standard curves

A plasmid containing the *sfGFP* gene (designed by Segall-Shapiro et al. (2014)) was ordered from the Addgene plasmid repository web site (https://www.addgene.org/59948/) as a bacterial stab (*E. coli*). The bacterial stab was streaked on LB agar + Ampicillin (100 µg/ml) and incubated overnight at 37°C. A single colony was re-grown in LB ampicillin (100 µg/ml) and used to generate glycerol stocks and subsequently used for PCR and Q-PCR validation of the primers (see main document). For all PCR amplifications, three negative control were included: *E. coli* DNA, environmental DNA and a no template control. The primer sF500_R (used for gene specific reverse transcription) was also tested at PCR level to ensure its specificity for subsequent RT experiments. After validation of the primers, a fragment of the plasmid including the T7 promoter site and the quasi full-length *sfGFP* gene (40 bp at the 3' end was not included) was PCR amplified using primers pBRforEco and GFP-Frc (reverse complement of GFP-F) (Table 3.1). The PCR product was purified using Agencourt AMPure XP beads (Beckman Coulter) following the manufacturer's recommendations and then used for *in vitro* transcription using the MEGAscript T7 transcription kit (Invitrogen) to prepare *sfGFP* RNA. The RNA preparation was treated with the Turbo DNase kit (Ambion) and the full digestion of the DNA template was confirmed by the absence of PCR amplification of a 300bp *sfGFP* fragment using the sF300_F and sF300_R primer pair (Table 3.1). Production of target RNA of the correct length was confirmed on the 2100 Bioanalyser RNA nano (Agilent) and concentration determined using fluorometric quantification method (Qubit RNA BR assay; ThermoFisher Scientific). The number of RNA transcripts was calculated using EndMemo RNA copy number Calculator (http://endmemo.com/bio/dnacopynum.php). A 10 points serial dilution was prepared by successive 1/10 dilutions. RNA dilutions were reverse transcribed using the combination of four enzymes and two priming (Table 3.2) and the resulting cDNA was diluted 1/10 before Q-PCR amplification.

3.2.4.b sfGFP Q-PCR standard curves

sfGFP RNA dilutions (10^{10}→10^{1} transcript copies/µl) were prepared and individually reverse transcribed (RT) using four different enzymes: Superscript IV (SSIV), Superscript III (SSIII) (Invitrogen), Sensiscript (Sensi) and Omniscript (Omni) (Qiagen) and two priming strategies - gene specific (GS) and random hexamer (RH). Each RT was done in duplicate. A summary of the protocol for each system is presented in Table 3.2. The resulting cDNA preparations underwent Q-PCR using the primer pair sF300_Fand sF300_R (Table 3.1).

3.2.4.c Spiking experiment

In order to determine which enzyme/priming combination was the most accurate, the exogenous RNA spike (*sfGFP* RNA) was seeded into a background of environmental RNA ($[RNA]_{background}$ =70.7 ng/µl; ratio $260/280_{background}$ = 1.63; ratio $260/230_{background}$ = 0.87) at known concentrations: 10^3, 5 x 10^3, 2 x 10^6 and 10^7 copies/µl. The RNA background was same for all spikes. These concentrations were chosen to mimic five-fold changes in gene expression at both low and high expression level. After the *sfGFP* spike was added, total RNA was reverse transcribed in triplicate, using different combinations of enzymes and priming (four different RT enzymes; two different priming strategies) in the same manner as illustrated in Figure 3.2 and Table 3.2. A 300bp fragment of the *sfGFP* cDNA was then quantified from the cDNA preparations using quantitative PCR (one Q-PCR reaction for each of the 3 RT replicates) with the primer pair sF300_F/ sF300_R. The Q-PCR mix was composed of 10µl EVAGreen Supermixes 2X (SsoFast; Bio-Rad), 0.5µl each primer (10µM each), 8µl water and 1µl cDNA template.

3.2.4.d Differential Expression (DE) between consecutive spike concentrations

The fold difference between consecutive spike concentrations was then calculated as the ratio of the mean copies/µl exogenous spike detected: DE "Low" corresponds to the ratio of mean copies/µl detected in the 10^3 spike versus the 5 x 10^3 spike. DE "High" corresponds to the ratio of mean copies/µl detected in the 2 x 10^6 spike versus the 10^7 spike. The standard deviations of the ratios were calculated as:

$$\mathrm{sd}\,\frac{m2}{m1}\sqrt{\frac{m2^2}{m1^2}\,x\,(\frac{sd2}{m1^2}+\frac{sd1}{m2^2})}$$

where m1 is the mean copies/µl at concentration C and m2 the mean copies/µl at concentration C x 5; sd1 and sd2 the standard deviations of m1 and m2 respectively

3.2.5 Effect of RT reaction on quantification of endogenous *amoA* and *16S rRNA* transcripts

3.2.5.a amoA and 16S rRNA Q-PCR standard curves

RNA standard curves were constructed by serial dilution of the target RNA, reverse transcription of the individual dilutions and Q-PCR amplification of the resulting cDNA: First, the genes of interest (Bacterial *amoA* and *16S rRNA*; Table 3.1) were amplified and cloned into

the pGem-T-Easy Vector Systems (Promega). The resulting ligation was transformed into the competent cell *E.coli* JM109 by heat shock according to manufacturer's instruction. White colonies were then PCR screened using T7 (5'-TAATACGACTCACTATAGGG–3') forward and gene specific reverse primer to ensure the amplicon had been cloned in the 5' → 3' orientation. Colonies that gave positive results were re-amplified using M13F (5'-GTAAAACGACGGCCAGT-3') and M13R (5'-CAGGAAACAGCTATGAC-3') and the resulting PCR products were purified using the SureClean Plus DNA purification kit (Bioline), then quantified using Qubit DNA High Sensitivity (ThermoFisher Scientific) and used as template for *in-vitro* transcription using the MEGAscript T7 transcription kit (ThermoFisher Scientific). RNA was then purified using the RNA clean-up protocol of the RNeasy Mini Kit (Qiagen) and DNase treated using an extended protocol: for 20µl purified RNA, 1µl DNase was added and incubated for 1h at 37°C. Then, another 1µl DNase was added and further incubated for 1h at 37°C. RNA was recovered using the phenol-chlorophorm/Isopropanol method (as recommended by the MEGAscript instruction manual) and complete digestion of the DNA template was confirmed by no amplification of the insert using T7F and M13R primers. The concentration and size of the DNA-free RNA were then checked using the Bioanalyzer 6000 RNA Nano (Agilent) assay. The number of RNA transcripts was calculated using EndMemo RNA copy number Calculator (http://endmemo.com/bio/dnacopynum.php). For *amoA*, an eight points serial dilution (5^{10} to 5^3 copies/µl) was prepared by successive 1/5 dilutions. For *16SrRNA*, a five points serial dilution (10^9 to 10^5 copies/µl) was prepared by successive 1/10 dilutions. RNA dilutions were reverse transcribed using the combination of four enzymes and two priming (Table 3.2) and the resulting cDNA was diluted 1/10 before Q-PCR amplification.

3.2.5.b RT-Q-PCR endogenous amoA and 16S rRNA transcripts

RNA extracted from the five biological replicates was reverse transcribed using eight different RT protocols (four enzymes and two priming strategies) as detailed in Table 3.2. The endogenous *amoA* and *16S rRNA* targets were then quantified from the resulting cDNA via Q-PCR [same Q-PCR protocol as in section 2.3.7].

3.2.6 Evaluation of RT reaction on transcript community composition

3.2.6.a Effect of reverse transcription on sequencing of endogenous transcripts

Besides evaluating the effect of the RT enzyme and priming strategy on the quantification of the endogenous *amoA* and *16S rRNA* transcripts, the effect of these on community composition,

as determined by amplicon sequencing of cDNA was also studied. To this end, *amoA* and *16Sr RNA* amplicons were generated from the cDNA preparations used in the previous experiment (Figure 3.2 B). For the *amoA* transcript, only SSIII and SSIV enzymes with both random hexamer and gene specific priming strategy were considered, as the other enzyme systems failed to work. For *16S rRNA* all four enzymes and both priming strategies produced amplicons and were therefore tested. Details for MiSeq-Illumina amplicon library preparation are provided in the section "MiSeq Illumina sequencing".

Table 3.1 List of primers used and corresponding (Q)-PCR conditions used in Chapter III.

Primer	Sequence (5'→3')	Orientation	Target	Experiment condition	Reference
pBRforEco	AATAGGCGTATCACGAGGC	Forward	*sfGFP* with T7 promoter	**PCR:** 95°-5min; (94°-30sec; 56°-30sec; 72°-30sec) x 30; 72° -10 min	https://www.addgene.org/browse/sequence/95181/
GFP-Frc	GTTACAAACTCAAGAAGGCC	Reverse			
sF300_F	TCACATGAAACGGCATGACT	Forward	Portions of the *sfGFP* gene	**PCR:** 95°-5min; (94°30sec; 55.3°-30sec; 72°-15sec) x 30; 72° -5min **Q-PCR:** 95°-3min; (95°-10sec ; 60°-15sec) x 40; melt curve: 65°→95° 0.5° increment/ 5sec	Reck *et al.,* 2015
sF300_R	GAACGGAACCATCTTCAACG	Reverse			
sF500_R	TAAAAGGACAGGGCCATCGC	Reverse			
BacamoA-1F	GGGGHTTYTACTGGTGGT	Forward	Bacterial *amoA* gene (485bp)	**PCR:** 95°C-15min; (94°C-30sec; 55°C-30sec; 72°C-30sec) x 32 cycles and 72°C-10min **Q-PCR:** 95°C-5 min; (95°C-30sec; 47°C-30 sec; 72°C-1min; 81°C-1sec→ plate read) x 40; melt curve: 65°→95° 0.5° increment/ 5sec	Hornek *et al.*, 2006
BacamoA-2R	CCCCTCBGSAAAVCCTTCTTC	Reverse			
1369F	CGGTGAATACGTTCYCGG	Forward	Bacterial *16S rRNA* gene (123 bp)	**Q-PCR:** 95°C-10min; (95°C-10sec; 60°C-30sec) x 40; 40°C-10min.	Suzuki *et al.*, 2000
1492R	GGWTACCTTGTTACGACTT	Reverse			
1389P	CTTGTACACACCGCCCGTC	Probe			
515F	GTGYCAGCMGCCGC GTAA	Forward	Bacterial *16S rRNA* V4 (291bp)		Parada *et al.*, 2016 Caporaso *et al.,* 2010
806R	GGACTACNVGGGTWTCTAAT	Reverse			
pGEMT_FW2	CGGCCGCGGGAATTCGAT	Forward	pGEM-T Easy vector (60 to 77)	**PCR:** 95°C-15min; (94°C-45sec; 50°C-30sec; 72°C-40sec) x 25; 72°C-10min	This study
M13 Forward	GTAAAACGACGGCCAGT	Forward	pGEM-T Easy vector (Forward: 2976 to 2993; Reverse: 176 to 193)		https://openwetware.org/wiki/Common_primer_sequences
M13 Reverse	CAGGAAACAGCTATGAC	Reverse			
T7 Forward	TAATACGACTCACTATAGGG	Forward	pGEM-T Easy vector (1 to 20)		

Table 3.2 Summary of the Reverse-transcription protocols tested. The black arrow indicates the procedure through time. RH = Random Hexamer; GS = Gene Specific. Provider for Superscript Enzymes: ThermoFisher Scientific; Provider for Sensiscript and Omniscript: Qiagen

Superscript III and IV (named SSIII and SSIV respectively thereafter)		**Sensiscript and Omniscript** (named Omni and Sensi respectively thereafter)
Cocktail 1: - 1µl primer (10µM GS; 50µM RH) - 1µl dNTPs (10µM each) - 5µl template RNA - 6µl water		Template RNA: 65°C-5min; ice-1min
65°-5min; ice-1min		Cocktail: - 2µl buffer 10x - 1µl primer (10µM GS; 50µM RH) - 5µl template RNA - 1µM dNTPs (10µM each) - 1µl enzyme (80U/µl) - 1µl RNAse inhibitor (40U/µl) - 9µl water
Cocktail 2: - 4µl Buffer 5X - 1µl DTT (0.1mM) - 1µl enzyme (200U/µl) - 1µl RNAse inhibitor (40U/µl)		
RH	GS	RH and GS
- 25°C (SSIII)/ 23°C(SSIV) for 10min - 50°C for 50min (SSIII)/ 55°C for 10min (SSIV) - 72°C for 15min (SSIII)/ 80°C for 10min (SSIV)	- 55°C for 50min (SSIII)/ 55°C for 10min (SSIV) - 72°C for 15min (SSIII)/ 80°C for 10min (SSIV)	37°C for 60min

3.2.6.b Effect of reverse transcription on sequencing of exogenous tagged-mock community

Preparation of the mock communities. 16S rRNA amplicons were generated by PCR amplification of the V4 region of the Bacterial *16S rRNA* gene from environmental samples using primers 515F and 806R (Table 3.1). PCR products were purified using the SureClean Plus kit (Bioline) and ligated into the pGEM-T Easy Vector using the pGEM-T Easy Vector System I (Promega) following manufacturer's instructions. The resulting constructions were transformed into *E.coli* MDS42 competent cells using the heat shock method (50µl *E.coli* MDS42 competent cells culture was incubated with 2µl of the ligation reaction on ice for 20min and transferred at 42°C for 50sec, then back on ice for 2min). The transformation solution was then added to 950µl SOC medium and incubated at 37°C for 1 hour. 100µl was then plated on LB agar/ampicillin (100µg/ml)/Xgal (20µg/ml)/IPTG (0.5 mM). After overnight incubation at 37°C, white colonies were picked up from the plates and grown for 2h30min in LB ampicillin (100µg/ml). A colony PCR was then carried using the BiotaqRed kit (Bioline) in a 50µl final reaction composed of 25µl 2X BiotaqRed Buffer, 1µl T7 forward primer (10µM), 1µl 806

reverse primer (10µM) and 1µl Bacterial culture (the T7 forward primer was used instead of the 515 forward to ensure that the amplicon had been cloned in the 5'→3' orientation relative to the T7 promoter). The rest of the culture was stored as glycerol stock at -80°C. When band of the expected size were detected on agarose gels, the PCR reaction was purified using the SureClean Plus kit (Bioline) and send for Sanger sequencing. In total 47 purified PCR products were sent for Sanger sequencing.

The sequences obtained were submitted to BLAST (Altschul *et al.*, 1990) search and then clustered at 97% similarity. A phylogenetic tree was constructed using MEGA7 (Kumar *et al.*, 2016) and the twelve most dissimilar sequences were selected for the mock community construction. Colony PCRs were then carried out using the glycerol stock of these 12 clones as template. The amplicons were generated using M13 forward and M13 reverse primers (Table 3.1). The resulting amplicons were composed of the full-length insert (≈290bp) plus a vector sequence of 100bp at the 5' end containing the T7 promoter site, and 137bp at the 3' end. After purification, the twelve PCR products were quantified using fluorometric quantification method (Qubit DNA High Sensitivity assay) and the corresponding copy number was calculated using EndMemo DNA copy number Calculator (http://endmemo.com/bio/dnacopynum.php). Finally, they were pooled together as explained in Figure 3.2 and Table 3.3 to generate the DNA Even (EM) and Uneven (UM) mock communities. Once pooled together, the DNA mock communities were diluted in order to reach $\approx 10^8$ copies/µl for each sequence.

The RNA mock communities were constructed by *in-vitro* transcription of the individual 12 PCR products (from DNA mock) using the MEGAscript T7 transcription kit (Invitrogen). The resulting RNAs were then gel-purified and DNase treated using the Turbo DNase Kit (Ambion). Complete digestion of the DNA template was confirmed by the absence of PCR amplification (using 515F and 806R primers; Table 3.1). The 12 individual RNA preparations were quantified using fluorometric quantification method (Qubit RNA broad range assay) and the corresponding copy number was calculated using EndMemo RNA copy number Calculator (http://endmemo.com/bio/dnacopynum.php). Each RNA preparation was diluted to 10^{10} copies/µl. Even (EM) and Uneven (UM) RNA mock communities were prepared as explained in Table 3.3 and Figure 3.2. C. Once pooled together, the RNA mock communities were diluted to reach $\approx 10^8$ copies/µl for each sequence.

Table 3.3 Composition of the mock communities.

Sequence ID	Even Mock (EM)		Uneven Mock (UM)	
	µl into pool	% of the community	µl into pool	% of the community
1			0.523	4.36
2			0.407	3.39
3			1.020	8.57
4			0.880	7.33
5			0.662	5.52
6	1	100/12≈ 8.33	0.815	6.79
7			1.452	12.1
8			2.170	18.08
9			1.490	12.41
10			0.690	5.73
11			0.955	7.96
12			0.931	7.76

Spiking and recovery of the RNA mock communities. Once constructed, the RNA mock communities (EM and UM) were diluted 1/10 into environmental RNA background (background: [RNA] = 102.3 ng/µl; 260/280 = 1.65; 260/230 = 1.28). This step was repeated five times. Once seeded with the mock communities, the environmental RNA preparations were reverse transcribed using two different enzymes (SSIII and SSIV) and two different priming strategies (RH and GS) (Figure 3.2. C). The GS priming was carried out using the 806R primer (Table 3.1). The procedure followed was the same as described in Table 3.2. After reverse transcription, the spiked mocks sequences were recovered from the total RNA pool by PCR amplification using the 806R reverse primer and a custom vector-specific forward primer pGEMT_FW2 (Table 3.1), designed to amplifying pGEMT vector sequence located between the T7 promoter site and the beginning of the insert, hence insuring the specific amplification of the mock sequences from the background (Figure 3.2). The specificity of the pGEMT_FW2 forward primer was checked by the absence of amplification of environmental *16Sr RNA* genes when used in combination with 806R.

3.2.7 Illumina MISeq amplicon library preparation

[see section 2.2.8].

3.2.8 Processing of amplicon sequences

[see section 2.2.9].

In this chapter, to ensure all *amoA* OTUs were valid *amoA* sequences they were translated using BLASTx to proteins and the match recorded for each individual OTU. Results of this search were used to filter the OTU_table before further processing, and non-translated amplicons removed from further analysis.

3.2.9 Statistical analysis

All statistical analyses were carried out in R (R Core team 2013). The effect of enzyme and priming on RT-Q-PCR result were tested after a log10 transformation of copy number data for 2-way ANOVA tests because the assumption of homogeneity of variances between groups was violated when using copy number directly. When the two-way ANOVA was significant, differences between enzymes/priming strategies were investigated using Tuckey HSD post-hoc test.

3.3 Results

3.3.1 Effect of enzyme and priming on the detection of exogenous spike

First, the impact of enzyme and priming on the quantification of RNA was determined for an exogenous target that was spiked at known concentrations into a background of environmental RNA. Artificial RNA (*sfGFP* RNA) that could be distinguished from environmental RNA background, was produced by *in vitro* transcription of a PCR product amplified from the pTHSSd_8 plasmid (Segall-Shapiro *et al.*, 2014). The resulting RNA was mixed with environmental RNA at different concentrations (10^3, 5 x 10^3, 2 x 10^6 and 10^7 copies/µl) and quantified using RT-Q-PCR (Figure 3.2 A).

3.3.2 *sfGFP* standard curves

Artificial RNA (*sfGFP* RNA) that could be distinguished from environmental RNA, was produced by *in vitro* transcription of a PCR product amplified from the pTHSSd_8 plasmid. Standard curves were constructed using 10-fold dilutions of *sfGFP* RNA from 10^{10} to 10^1 copies/µl that underwent individual reverse transcription in duplicate using each of the four RT enzymes with two priming strategies. The Cycle thresholds (Cts) of the same sample derived from different enzyme/priming strategies were obtained by Q-PCR amplification of the resulting cDNA. Amplification of the no template controls and $\log_{10}$[*sfGFP*]= 1 and 2 gave no signal. The Limit of Detection (LD) and Quantification (Forootan *et al.*, 2017) was $\log_{10}$[*sfGFP*]= 3 for all RT systems except for Sensi-RH which had a LD at $\log_{10}$[*sfGFP*]= 4.

Excluding the Cts obtained for $\log_{10}$[*sfGFP*]=10 resulted in an improvement of the regression fit (slopes closer to the expected -3.332 and R squared closer to 1; data not shown). As such standard curves ranged between $\log_{10}$[*sfGFP*] = [3;9] (or [4;9] for Sensi-RH). For all enzymes, the use of RH resulted in higher Cts than GS at all *sfGFP* concentrations. SSIV-GS resulted in the best efficiency (99.3%) while the lowest was obtained by Sensi-RH (84.2%) (Table 3.4).

Table 3.4 Description of the *sfGFP* standard curve regressions.

Enzyme	Priming	Slope	Efficiency (%)	Y-Intercept	R squared	LD
SSIV	**GS**	−3.34	99.3	43.55	0.993	10^3
	RH	−3.57	90.6	45.44	0.993	10^3
SSIII	**GS**	−3.39	97.2	43.62	0.996	10^3
	RH	−3.49	93.4	44.9	0.996	10^3
Sensi	**GS**	−3.61	89.2	43.87	0.999	10^3
	RH	−3.77	84.2	49.51	0.998	10^4
Omni	**GS**	−3.63	88.6	43.68	0.998	10^3
	RH	−3.67	87.3	46.23	0.997	10^3

The regressions coefficients were calculated based on the linear model (y= ax+b; a=Slope and b=Intercept). The fit of the regression is represented by the R squared values. Efficiency of the Q-PCR was calculated as Efficiency (%) = 100 x ($10^{-1/slope}$ – 1). LD = Limit of Detection.

Comparison of regressions between RT methods (enzyme/priming) revealed significant difference between slopes (F=3.29; Df=7; *p.*value=0.0036) (GraphPad Prism6, www.graphpad.com). The effect of the RT approach on standard curve construction was further investigated using multilevel linear model analysis. Three different models were tested where 1) intercepts only, 2) slopes only and 3) slopes and intercepts varied between groups (*i.e*. RT method). Models 1 and 2 were then tested against model 3 using an ANOVA. Model 3, allowing for variations in both slopes and intercepts, resulted in a better fit than model 1 (intercept only) or model 2 (slopes) (Table 3.5) indicating that the effect of enzyme and priming impacted both the slope (*i.e.* efficiency) and the intercept (*i.e.* signal at No Template Control (NTC)) of the standard curves.

Table 3.5 Comparison of models for standard curve.

Model	Random Effect	df	Fit Criteria				Comparison	ChiSq	Df	Pr(>ChiSq)
			AIC	BIC	logLik	Deviance				
Model1	Intercept	4	227	238	-109	219	1 VS 3	7.38	2	0.025
Model2	Slope	4	301	312	-146	293	2 VS 3	80.86	2	2.2-16
Model3	Intercept and Slope	6	224	240	-106	212				

For each model, the fit was assessed using four criteria namely; AIC (Akaike Information Criteria), BIC (Bayesian Information Criteria), LogLik (log Likelihood) and deviance. Models 1 and 2 were compared to model 3 and the statistic of the test is reported in the last column (Pr(>ChiSq))

3.2.3. RT-Q-PCR detection of the RNA *sfGFP* spike

The exogenous RNA spike, *sfGFP,* was added to environmental RNA at known concentrations and the resulting preparations were reverse-transcribed using the eight different combinations of RT enzymes and priming strategies (Figure 3.2 A). Four different concentrations of spike were added to environmental RNA background: two low and two high, with a five-fold difference in *sfGFP* copy number between the two low and the two high spikes respectively (10^3 and 5 x10^3 copies/µl for low spikes; 2 x10^6 and 10^7 copies/µl for high spikes). After cDNA synthesis, the spiked target was quantified by Q-PCR and Cycle thresholds (Ct) were converted to copies/µl using standard curves.

Both enzyme and priming had a strong effect on the copy number of exogenous target detected (log_{10} transformed) at all spike concentrations (Table 3.6; Figure 3.3). Overall, SSIII and SSIV enzymes were the closest to the expected value. SSIV was slightly more accurate than SSIII, especially at spike concentrations > 5 x 10^3. The use of Omni resulted in an underestimation of the spike concentration with factors ≈4 to ≈50 depending on the concentration of the target (higher differences at higher concentrations). Similarly, the use of Sensi also resulted in an underestimation of the exogenous target concentration with factors ≈3 to ≈30 (higher differences at higher concentrations) (Figure 3.3).

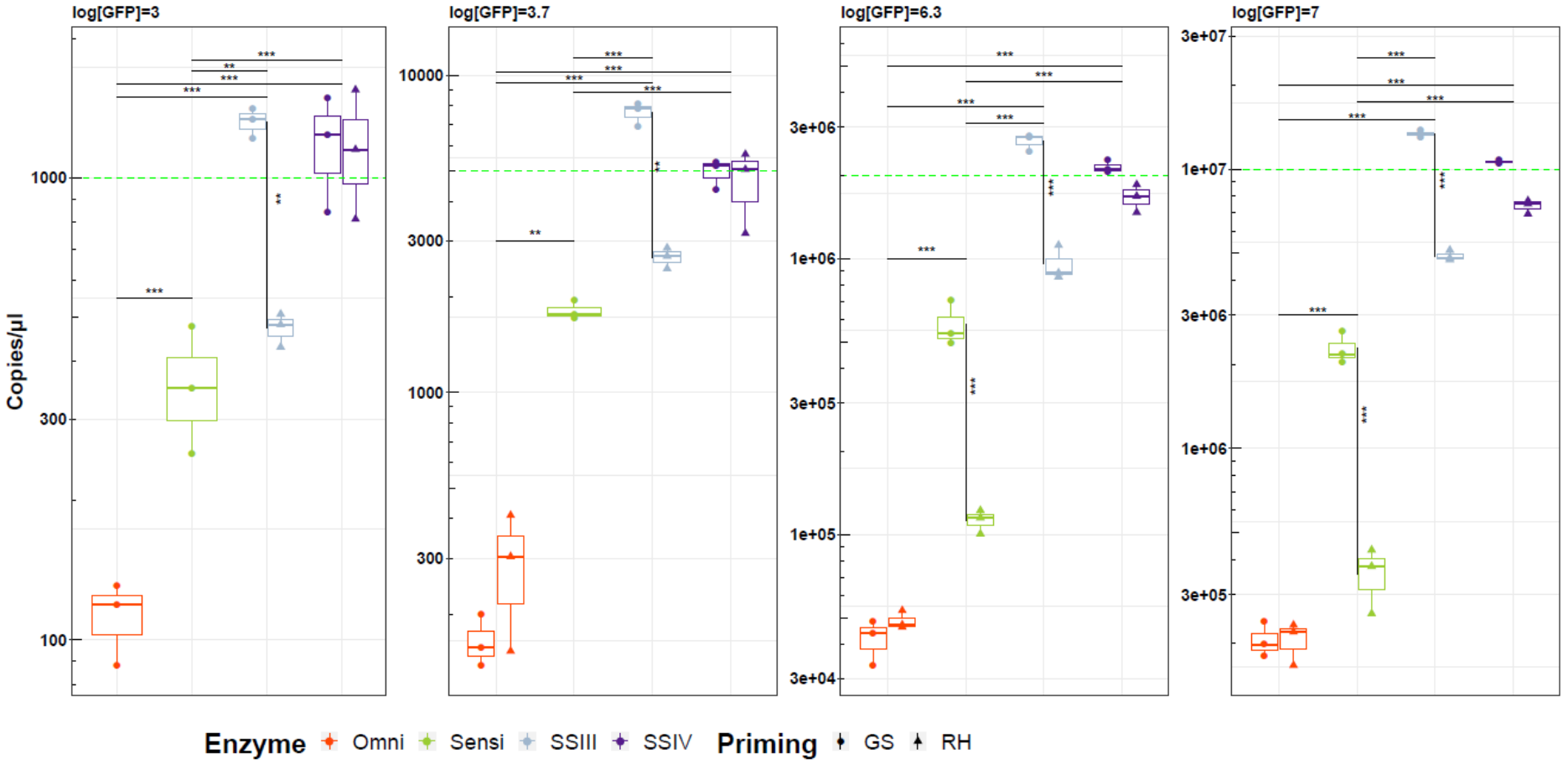


Figure 3.3 Effect of reverse-transcriptase enzyme and priming on the quantification of the same *sfGFP* spike in environmental RNA background. The concentration of the RNA spike (log[GFP]= 3; 3.7; 6.3and 7) inoculated is indicated at the top of each plot. The results of the two-way ANOVA, showing statistical differences between priming and enzyme for the same template, are presented as vertical and horizontal lines respectively. *: *p*.value <0.05; **: *p*.value <0.01; ***: *p*.value <0.001. A green dashed line indicates the actual concentration of spike for each experiment. GS = gene specific; RH = Random Hexamer

Table 3.6 Two-way ANOVA showing the impact of RT system on the quantification of the *sfGFP* spike.

Spike concentration (copies/µl)	Enz	Prim	Enz:prim
1×10^3	6.62×10^{-8}	2.57×10^{-3}	3.45×10^{-3}
5×10^3	5.92×10^{-12}	0.08	2.94×10^{-4}
2×10^6	$<2 \times 10^{-16}$	4.98×10^{-10}	7.38×10^{-9}
1×10^7	$<2 \times 10^{-16}$	1.24×10^{-10}	1.35×10^{-8}

The use of GS priming resulted in more accurate quantification for all enzymes except Omni for which it had no effect. For Sensi, RH priming failed at the low spike concentrations while at the high concentration of spike, RH was significantly lower (6-fold) than GS. For the Superscript enzymes, the use of RH versus GS generally resulted in lower quantification of the same target, except when using SSIV at low concentrations where the priming strategy had no effect. Of the two Superscript enzymes, SSIII with GS always overestimated the concentration of spike whereas RH always underestimated it (≈2 fold or less). Priming had the least effect with SSVI, but more accurate quantification was achieved when using GS priming (Figure 3.3).

Next, we tested the ability of the RT systems to faithfully report a 5-fold difference in the *sfGFP* spike concentration between the two low and two high concentrations respectively. For this, the differential expression (DE), *i.e.* the ratio of the average transcript number/µl between the two low or the two high spikes respectively, was calculated (Figure 3.4). The DE does not report how accurately the system quantifies the spike but rather its ability to reflect the 5-fold change. Again, the choice of enzyme and priming had an effect on the observed DE. All systems were better at detecting actual differences (DE closer to 5) at high spike concentrations. The most accurate system, *i.e.* giving DE values closer to the expected 5, at high spike concentration was SSIII-GS (DE = 5.03), followed by SSIV-GS (DE = 4.96) and Omni-GS (DE = 4.91). All enzymes gave less accurate results when used in combination with RH priming at high spike concentration. Still, SSIII and SSIV were the most accurate enzymes, with SSIII better than SSIV. At low spike concentration, SSIII always overestimated the DE whereas SSIV always underestimated it. The use of RH made SSIII slightly more accurate (DE=5.64 with RH versus 5.75 with RH) whereas it made SSIV slightly less accurate (DE=3.94 with RH versus 4.16 with GS). Interestingly, Sensi performed the best at low spike concentration when used with GS priming (DE = 5.03) whereas Omni performed the worst (DE = 1.47). Both Sensi and Omni failed at low spike concentrations when used with RH priming. Therefore, the

superscript enzymes preformed best overall, and the DE was improved when SSIII and IV were used in combination with GS priming (Figure 3.4).

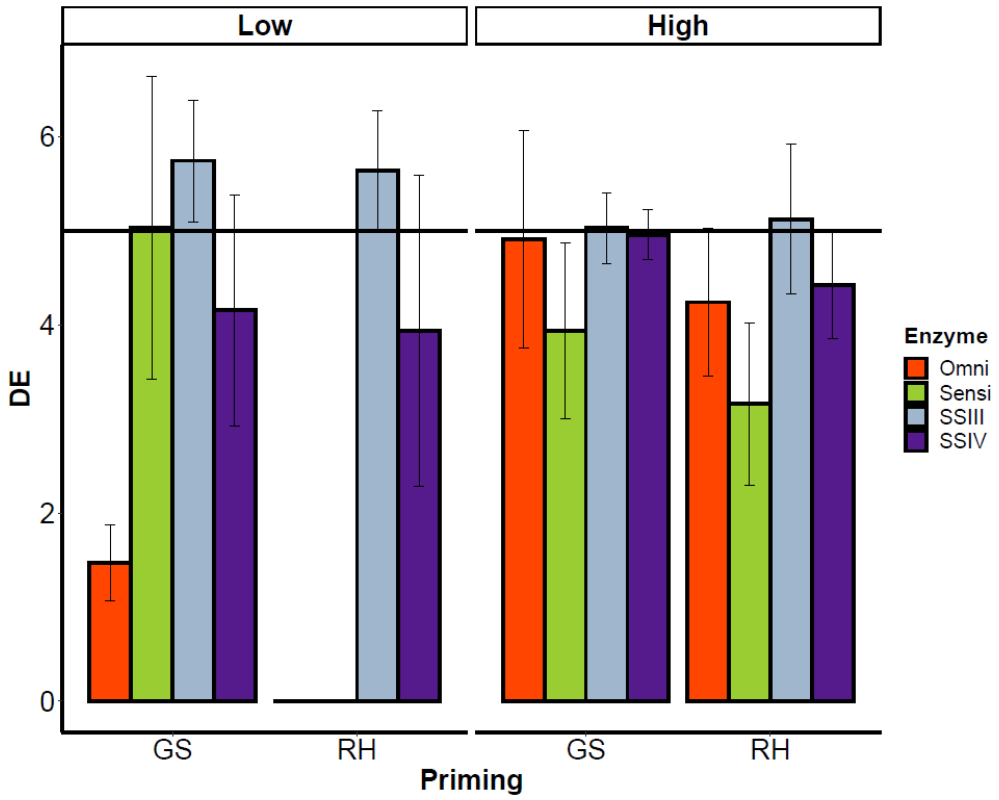


Figure 3.4 Ability of the RT systems to detect a 5-fold difference of target at low or high concentration in a background of environmental RNA. Differential expression (DE) is the ratio of the average copies/µl between "Low" (1 x 10^3 and 5 x 10^3 copies/µl) or "High" (2 x 10^6 and 1 x 10^7 copies/µl) *sfGFP* spikes. The expected DE is represented by the horizontal black line at y=5 while the measured DE by each RT system is shown by the bar-plot. GS = gene specific; RH = random hexamer

3.3.4 Effect of standard curve construction on *sfGFP* quantification

As there are two approaches to constructing RNA standard curves (Smith *et al.,* 2006), we tested if this had any impact on quantification of the spike and the above results. A standard curve can be made by serial dilution of RNA with individual RTs or via a single RT of RNA followed by serial dilution of cDNA. Standard curves for each enzyme and primer combination were made using these two approaches to quantify the spiked *sfGFP* (Figure 3.2. A). The percentage error was calculated between the observed and expected copies/µl for each *sfGFP* spike generated from each standard curve. The standard curve constructed from the dilution of cDNA generally increased the percentage error, and therefore dilutions of RNA for individual RTs were used for subsequent standard curves (Figure 3.5).

3.3.5 Effect of enzyme and priming on the quantification of endogenous environmental transcripts

3.3.5.a RNA Quality check

Before proceeding with the quantification of endogenous transcripts, RNA extracted from sediment underwent a quality check (Table 3.7) (Chapter II; Cholet *et al.*, 2019). All samples had good integrity as shown by the RIN (always > 7) and R_{amp} (R_{amp} 380/120 ≈ 0.8 or higher and R_{amp} 380/170 ≈ 0.7 or higher).

3.3.5.b Quantification of endogenous transcripts

Next, the effect of RT enzyme and priming strategy on the quantification of transcripts from the same sediment sample were tested. For this we targeted *in situ* highly abundant *16S rRNA* and less abundant mRNA from the bacterial ammonia monooxygenase subunit A, *amoA,* for quantification from cDNA generated using the different combinations of reverse transcriptases and priming (Figure 3.2. B). Results were converted into copies/µl using paired standard curves, normalized per µg of extracted RNA and log10 transformed (for parametric 2-ways ANOVA tests). The results clearly showed the effect of the RT system was target dependent: for *16S rRNA* both enzyme and priming significantly affected the results whereas for *amoA* only the effect of enzyme was significant (Figure 3.6; Table 3.8).

For *amoA*, the choice of RT system resulted in differences of up to 600-fold in the detected copies/µg RNA in the samples tested (Omni-RH versus SSIV-RH) and, in the most extreme case, the difference between detection of the target or not (Sensi). For this assay, only the choice of enzyme significantly affected the results whereas priming did not. A clear difference between Omni/Sensi and the Superscript enzymes (SSIII and SSIV) was observed with, on average, 150 times more *amoA* transcripts per µg RNA with the Superscript enzymes. For Sensi and Omni, the choice of GS priming resulted in better results, especially for Sensi, which failed at producing reliable results with RH. For Omni, the use of GS priming resulted in 6 times more copies of *amoA* transcripts compared to RH, although this difference was not statistically significant. SSIII and SSIV performed relatively similarly, with no statistical differences between the two, although the use of SSIV resulted in higher numbers of *amoA* transcripts detected (≈+2.4 fold with GS and ≈+1.7 fold for RH; *p*.value=0.512). Interestingly, although the use of RH priming resulted in higher Cts on Q-PCR (*i.e.* lower quantification), conversion to copies/µg RNA via the standard curve resulted in a higher quantification than that achieved

with GS priming (≈+1.2fold for SSIV (*p*.value=0.99) and ≈+1.7fold for SSIII (*p*.value=0.99)). In summary, SSIV was the best choice for the detection of *amoA* transcripts as it resulted the highest numbers of transcripts detected and produced consistent results between GS and RH. When used in combination with GS as opposed to RH the results were more precise (*i.e.* lower standard deviation) (Figure 3.6).

Table 3.7 Sediment RNA quality check.

	Quantity (ng/µl)			Purity			Integrity	
Sample	[RNA] NanoDrop	[RNA] Qubit	[RNA] Bioanalyser	260/280	260/230	RIN	R_{amp} 380/120	R_{amp} 380/170
Env1	87.8	54.6	55.8	1.69	1.51	7.85	1.01	0.77
Env2	301.2	196	135	1.74	2.02	7.6	0.88	0.73
Env3	164	117	67	1.74	1.73	7.9	0.92	0.73
Env4	201	140	91.5	1.76	1.87	7.9	0.78	0.65
Env5	218.2	156	91.5	1.78	1.84	7.1	0.82	0.72

Extracted RNA quantity, purity and integrity were determined. RIN = RNA Integrity Number, as determined from Agilent Bioanalyser; R_{amp} was calculated as described in Cholet *et al.,* 2019.

For *16S rRNA*, both enzyme and priming had a strong effect on quantification (Figure 3.6; Table 3.8). The choice of RT system resulted in differences of ≈4000, ≈3500 and ≈2300 fold between highest (Omni-RH) and lowest quantification (SSIV-RH > SSIII-RH > Sensi-RH). Omni actually behaved differently from the other enzymes as it was the only one for which the use of RH resulted in higher detected copies/µg RNA compared to GS and indeed, statistical differences were found only between Omni and the other three enzymes. For SSIV, SSIII and Sensi, the use of RH always resulted in lower detected copies *16S rRNA*/µg RNA (≈ -120fold for SSIV and Sensi; ≈ -400fold for SSIII). Results between enzymes were more consistent when used with GS priming, with an average difference in detected copies/µg RNA between enzymes of 2.18-fold (max: 4.01-fold between SSIII and SSIV; min: 1-fold between Sensi and Omni). With this priming, SSIII resulted in the highest number of copies of *16S rRNA*/µg RNA (+4.01fold versus SSIV and +2.22fold versus Omni and Sensi). It is worth mentioning that, even though the use of Omni-GS and Sensi-GS resulted in more copies *16S rRNA*/µg RNA on average compared to SSIV-GS (≈+1.8fold), SSIV-GS was more precise (*i.e.* lower standard deviation), as was the SSIII-GS combination (Figure 3.6).

Table 3.8 Two-ways ANOVA showing the impact of RT system on the quantification of endogenous transcripts.

Transcript	Enz	Prim	Enz:Prim
amoA	6.61×10^{-13}	0.26	0.03
16S rRNA	2.24×10^{-8}	7.55×10^{-8}	2.31×10^{-9}

p.value for the effect of enzyme (Enz), priming (Prim) and the interaction between the two (Enz:prim) on the Ct of the target transcript.

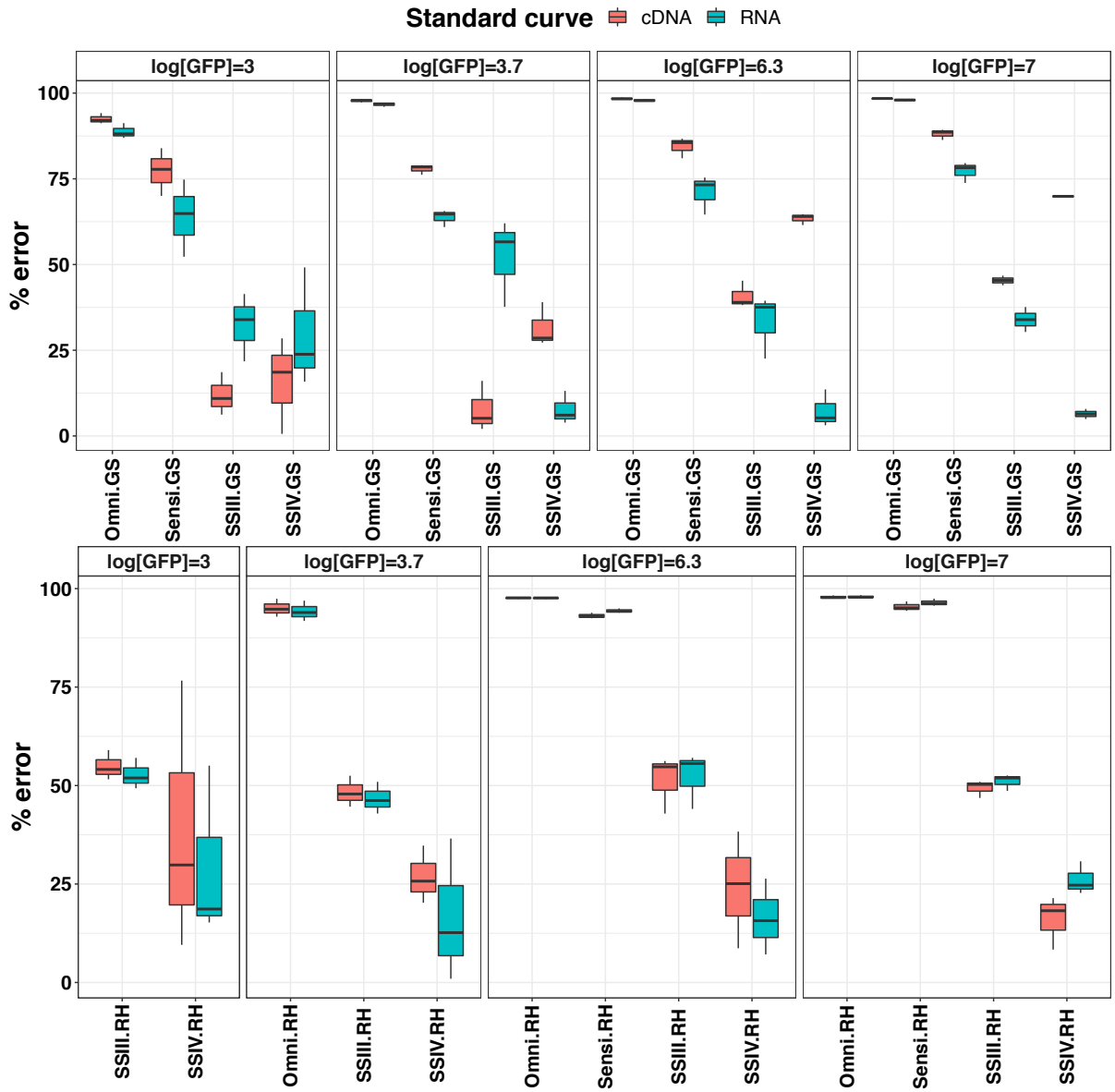


Figure 3.5 Percentage error in the detected copies/µl of spike *sfGFP*. The percentage error has been calculated for the three replicates and expressed in absolute value regardless of the type of error (over or under estimation of the expected value), for GS (top) and RH (bottom) priming. Standard curve RNA: standard curve made by serial dilution of RNA and individual RT, Standard curve cDNA: standard curve made by a single RT and serial dilution of cDNA.

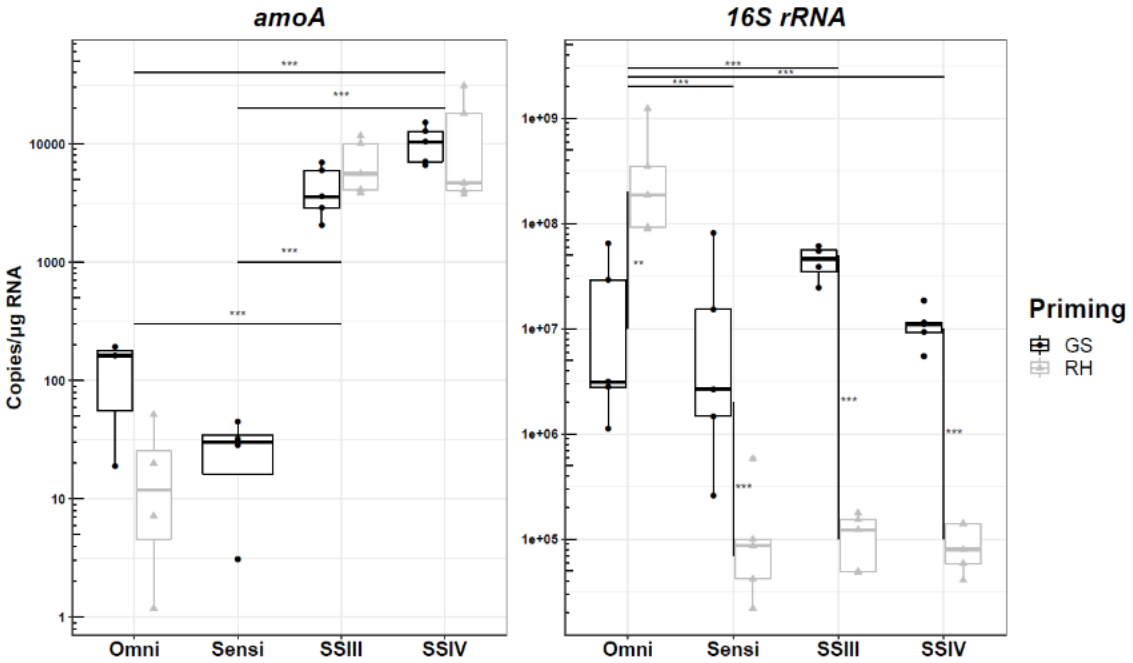


Figure 3.6 Impact of RT system quantification of two endogenous transcripts from environmental samples. The effect of the RT system on quantification of Bacterial *amoA* (left) and *16S rRNA* (right) transcripts. The results of the two-way ANOVA, showing statistical differences between priming and enzyme for the same template, are presented as vertical and horizontal lines respectively. *: *p*.value<0.05; **: *p*.value<0.01; ***: *p*.value<0.001.

3.3.6 Effect of enzyme and priming on cDNA amplicon sequencing data

While the quantitative work clearly shows dramatic and significant differences when using different RT enzyme and priming strategies for quantification of the same template, it does not inform if these impact upon community transcript diversity. To examine this, RNA and DNA mock communities of known composition were examined in addition to endogenous *16S rRNA* and *amoA* transcripts from marine sediments.

3.3.6.a Effect on mock community composition

As the actual composition of the transcriptome of the environmental samples is unknown, it is virtually impossible to determine which RT system most closely represents the starting RNA. We thus tested the effect of enzyme and priming on known RNA mock communities. Two mocks community (one even, with all 12 sequences at the same relative proportion, designated EM and one uneven, with the 12 sequences at different relative proportions, designated UM, Table 3.3), each composed of twelve different *16S rRNA* transcripts were constructed as detailed in Figure 3.2. C. To further tease apart the effects of the PCR from the RT, both DNA and RNA mock communities were constructed. Of the twelve mock community sequences, one

(S9) was over-represented in the DNA mock community but under-represented in the RNA mock community sequencing data. In contrast sequence S10 showed the opposite trend (over-represented in the RNA mock community but under-represented in the DNA mock community). As a result, sequences (S9 & S10) were removed from further analysis.

DNA mock. In the EM community, there were variations from the expected proportions (10%). Some members of the community: S1, S2, S4, S8 and S12 were underrepresented whereas S3, S5, S6, S7, S11 were overrepresented. The most underrepresented member, S4, represented only 4% of the total community whereas the most overrepresented, S7, represented 14%. Yet, even though the observed proportions deviated from the expected 10%, they were within the same order of magnitude (Figure 3.7. A). For the UM mock communities, the observed proportion of each member was plotted against the expected proportion (Figure 3.7. B). A regression line with equation y=x is expected if each sequence is faithfully represented. The UM community results were consistent with those of the EM, with S6, S7, S11 overrepresented and S1, S2, S4 and S12 underrepresented. S5 was at the correct proportion in both EM and UM. For most sequences, the errors of representations were consistent between EM and UM (*i.e.* a sequence overrepresented in the EM was generally also over-represented in the UM and vice versa) indicating a sequence specific bias of the PCR/sequencing workflow. And indeed, when the proportions of the UM were corrected by those of the EM, the fit of the regression improved (Figure 3.7.C).

RNA mock. As seen for the DNA mock community, the proportions observed in the RNA EM differed from the expected 10% (Figure 3.7. D, G, J, M). When analysing the sequences abundances in the EM using PERMANOVA test (Bray-Curtis distances), it was found that the priming strategy used had a significant effect on the proportions retrieved (*p.*value = 0.001), with RH being more accurate than GS (proportions closer to the expected 10%). On the other hand, neither enzyme nor the interaction between enzyme and priming had a significant effect (*p.*value=0.208 and *p.*value=0.194 respectively). The data set containing the highest errors was SSIII GS, followed by SSIV GS, SSIII RH and the lowest errors were found for SSIV RH (similar amount of error than for the DNA mock) (Figure 3.7 and Figure 3.8). For individual sequences, there was a sequence-specific bias: some sequences (S1, S2 and S6) were always over-represented in the data sets but this was different depending on the priming used (proportion S1 < proportion S2 with GS and inversely with RH). Other members of the mock communities (S3, S4, S7 and S8) had proportions lower than the expected 10%; Though S7 was

very close to the expected 10% with SSIV RH. Finally, the results from the other members of the mock community were dependent on the enzyme and priming strategy used. For example, S5 and S12 were always over-represented in the RH-prepared libraries but not in the GS ones. Inversely, S11 was over-represented in the GS libraries (especially with SSIII) and its proportions decreased in the RH ones

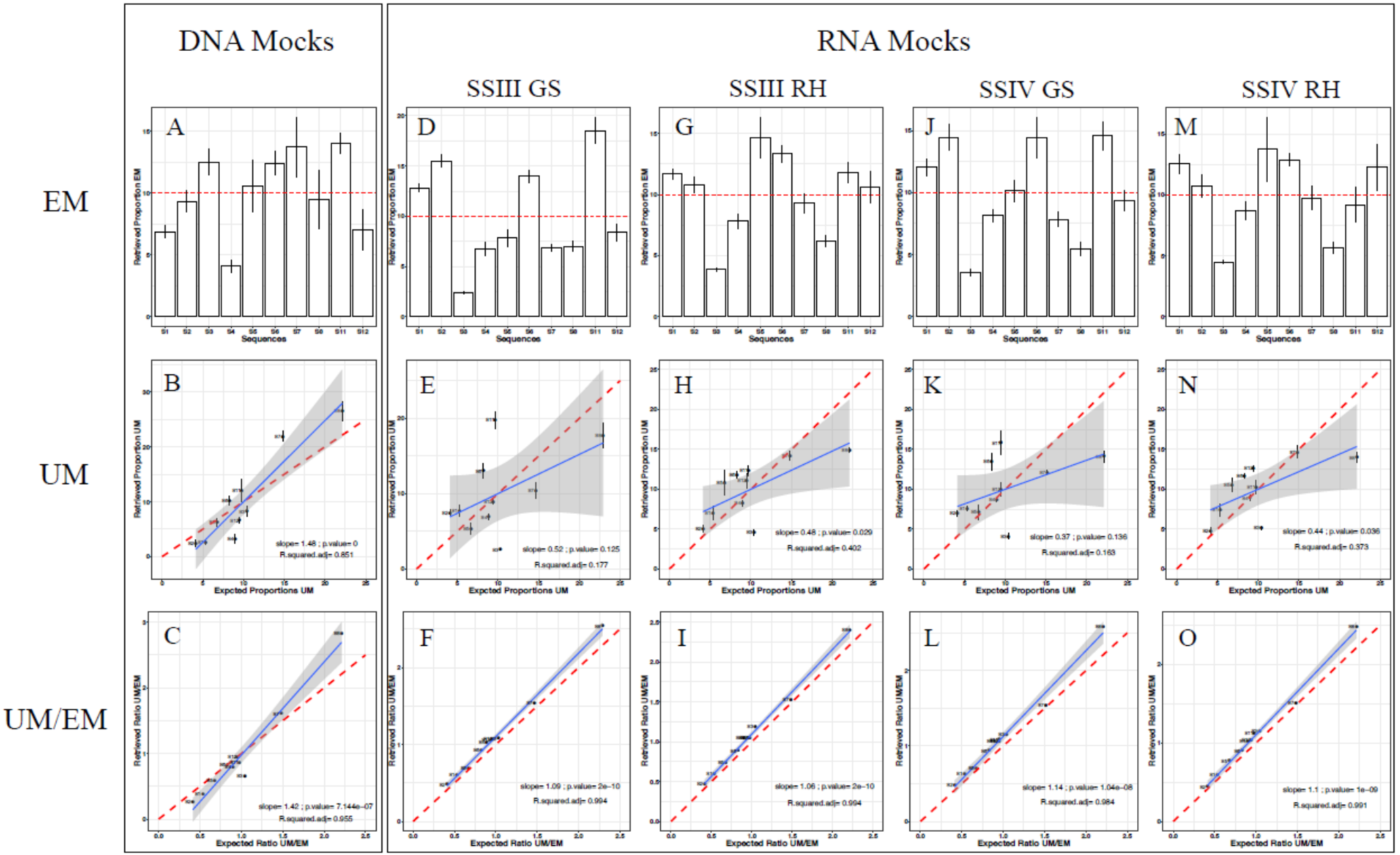

DNA Mocks
RNA Mocks
SSIII GS
SSIII RH
SSIV GS
SSIV RH
EM
UM
UM/EM
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
Retrieved Proportion EM
Retrieved Proportion UM
Retrieved Ratio UM/EM
Sequences
Expcted Proportions UM
Expected Ratio UM/EM
slope= 1.48 ; p.value= 0
R.squared.adj= 0.851
slope= 0.52 ; p.value= 0.125
R.squared.adj= 0.177
slope= 0.48 ; p.value= 0.029
R.squared.adj= 0.402
slope= 0.37 ; p.value= 0.136
R.squared.adj= 0.163
slope= 0.44 ; p.value= 0.036
R.squared.adj= 0.373
slope= 1.42 ; p.value= 7.144e−07
R.squared.adj= 0.955
slope= 1.09 ; p.value= 2e−10
R.squared.adj= 0.904
slope= 1.06 ; p.value= 2e−10
R.squared.adj= 0.904
slope= 1.14 ; p.value= 1.04e−08
R.squared.adj= 0.984
slope= 1.1 ; p.value= 1e−09
R.squared.adj= 0.991

Figure 3.7 Impact of RT system on the reproducibility of mock community composition. Row EM: observed proportions of each sequence within the Even Mock (EM) communities. The dashed red line indicates the expected proportion. Row UM: regression of observed versus expected proportions of each sequence within the Uneven Mock (UM) communities. Row UM/EM: the observed proportion of each sequence in the UM has been divided by its observed proportion in the EM and plotted against its actual ratio. Row UM and UM/EM the expected regression (y=x) is represented by a dashed red line. The actual regression is represented by a solid blue line with the 95% confidence interval (grey area). Individual regression statistics are reported on individual plot

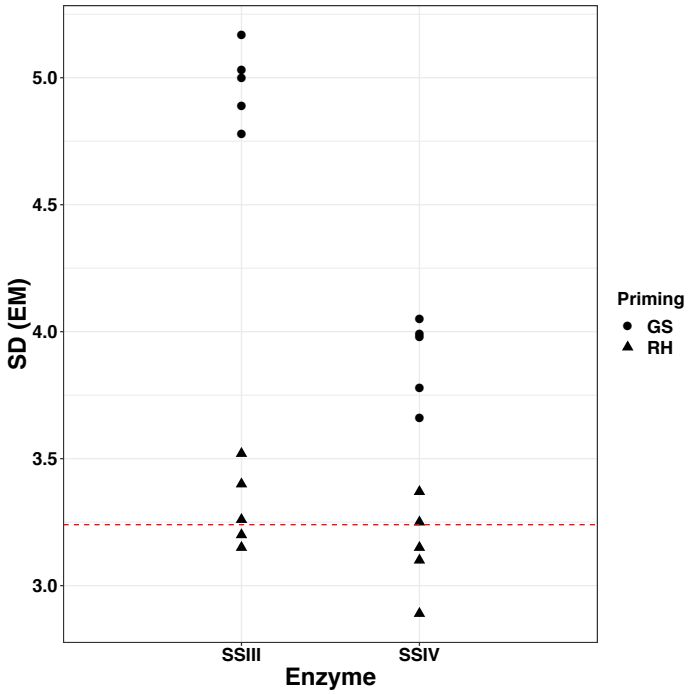


Figure 3.8 Errors in the sequencing of the cDNA even mocks. The errors are represented as the standard deviations (sd) of the 5 EM replicates. For reference, the mean sd observed in the DNA mock is indicated as a red dashed line.

As for the EM, the proportions retrieved for the UM mocks differed from the expected proportions (y=x) (Figure 3.7. E, H, K, N). As for the EM, RH seemed to perform better than GS with better fits for the regressions (for SSIII: R-squared GS = 0.177 versus 0.402 for RH; for SSIV: R-squared GS = 0.163 versus 0.373 for RH). The use of RH priming also resulted in slopes closer to the expected value of 1 compared to GS indicating a better conservation of the relative proportions using this priming strategy.

However, as observed for the DNA mock, the errors were consistent between EM and UM: A sequence over-represented in the EM would also be over-represented in the UM and inversely. As a consequence, when UM reads were corrected by EM reads (Figure 3.7. F, I, L and O), the calculated slopes were very close to the expected value (y=x) and the R-squared values also

improved (close to 1) indicating a better fit of the regression. This observation indicates that the same sequences were misrepresented in both EM and UM communities.

3.3.6.b Effect of RT enzyme and priming strategy on endogenous community composition

16S rRNA and (Bacterial) *amoA* PCR amplicons were generated from cDNA prepared using the different combinations of enzymes and priming (Figure 3.2. B). For *amoA*, only SSIII and SSIV were compared as the Sensi and Omni enzymes failed to produce PCR amplicons for sequencing (as reflected by Ct values above 30). The combination of Sensi and RH priming also failed to reliably amplify *16S rRNA* transcripts and was therefore also excluded from further analysis.

Effect on 16S rRNA community composition. When all four enzymes were taken into account, the effect of enzyme on OTU community composition was always significant (Table 3.9 and Figure 3.9). In addition, the priming strategy had a significant effect on community composition but only when the Bray-Curtis dissimilarity matrix was considered (Table 3.9 and Figure 3.9). Still, the choice of priming strategy had less of an effect on *16S rRNA* community composition than for *amoA,* (Figure 3.9) as GS priming did not systematically result in more OTUs (Figure 3.10). For the *16S rRNA* dataset, the combined effect of enzyme and priming depended on the specific combination. Specifically, for SSIII and SSIV, there was no difference between enzymes, but there was a significant difference in the Bray-Curtis distance matrix due to priming (richness RH>richness GS for SSIII and inversely for SSIV (Figure 3.9, 3.10, Table 3.9)) albeit marginally significant (*p.*value = 0.047). When Sensi and Omni were compared, both enzyme and priming had an effect on community composition (Table 3.7, Table 3.9).

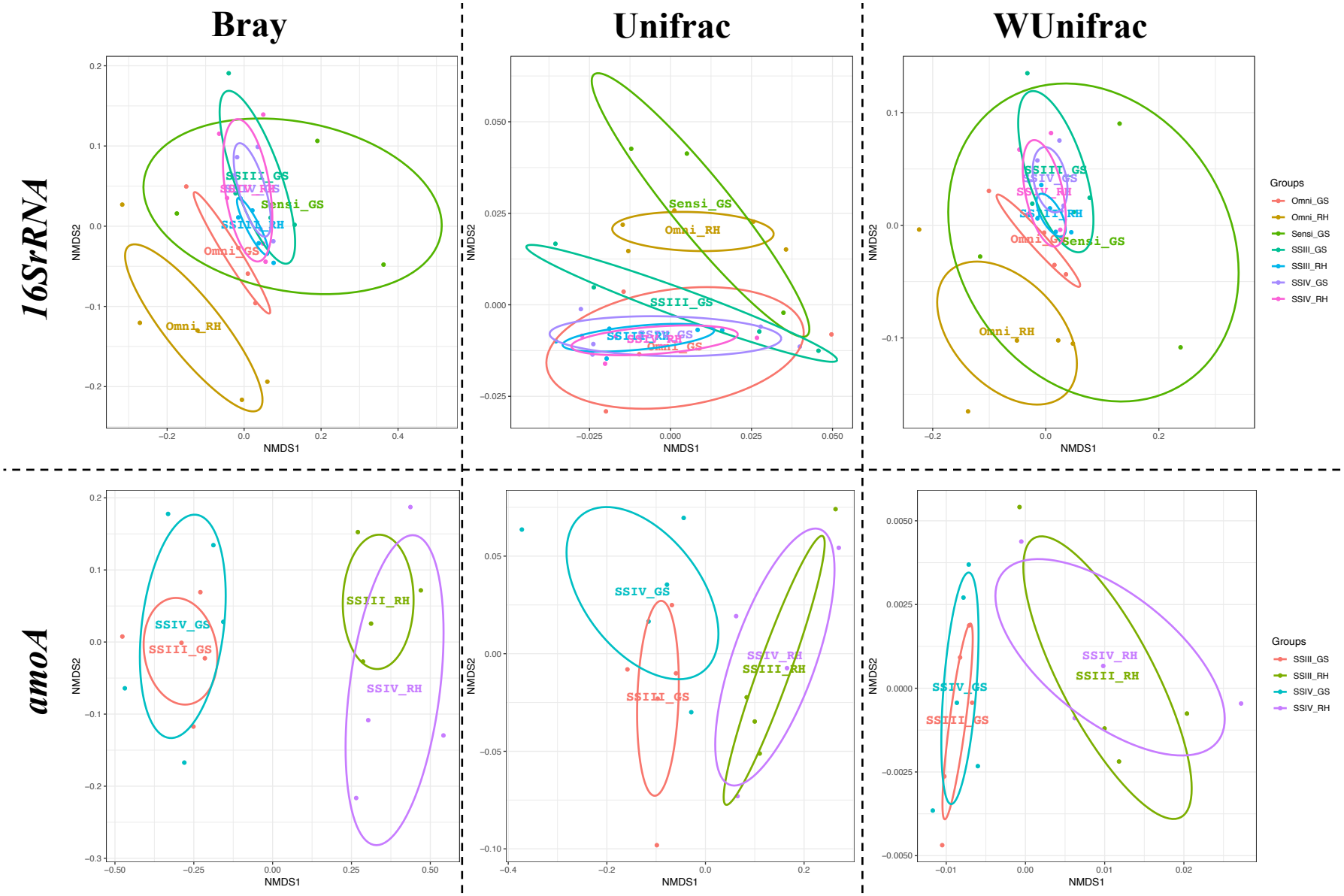

Bray
Unifrac
WUnifrac
16SrRNA
amoA
NMDS1
NMDS2
Omni_GS
Omni_RH
Sensi_GS
SSIII_GS
SSIII_RH
SSIV_GS
SSIV_RH
Groups

Figure 3.9 Effect of enzyme and priming on *amoA* and *16S rRNA* transcript community composition. NMDS clustering of *16S rRNA* (top) and *amoA* (bottom) cDNA community composition of the same sample derived from different enzyme and primer strategies, using Bray-Curtis (left), Unifrac (middle) and WUnifrac (right) distances. Corresponding groups are indicated in the legend.

We further tested the effect of the RT system on the recovery of the *16S rRNA* transcripts at different taxonomic levels (Table 3.10). The effect of enzyme was stronger than priming and was more important when the individual OTUs had well resolved taxonomy (at Family, Order and Class level). On the other hand, at lower taxonomic levels, the effect of the RT system became non-significant as a lot of OTUs could not be assigned to a species or genus and were therefore classified as unknown. Interestingly, the effect of both enzyme and priming became significant again at the kingdom level (Table 3.10).

amoA OTU check. amoA OTUs sequences were checked by BLASTx to ensure they translated into AMO A proteins. Results of this search revealed that, out of the 202 *amoA* OTUs, 63 did not correctly translate (*e.g.* "hypothetical protein" or "low quality protein") and were therefore removed from the data set. In terms of percentage of reads, these non-translating OTUs represented 0.017% to 4.6% of the total. As shown in Figure 3.11, the amount of "incorrect OTUs" found was higher in the data set obtained when using GS priming. However, as the number of reads obtained with GS was generally higher, they did not represent a significantly higher percentage of the community, except for the replicate 1 and 3 with SSIV GS where the percentage of incorrect OTUs represented 4.6% and 1.3% respectively (Figure 3.11). These OTUs were removed before further processing, and therefore didn't impact on the subsequent analysis.

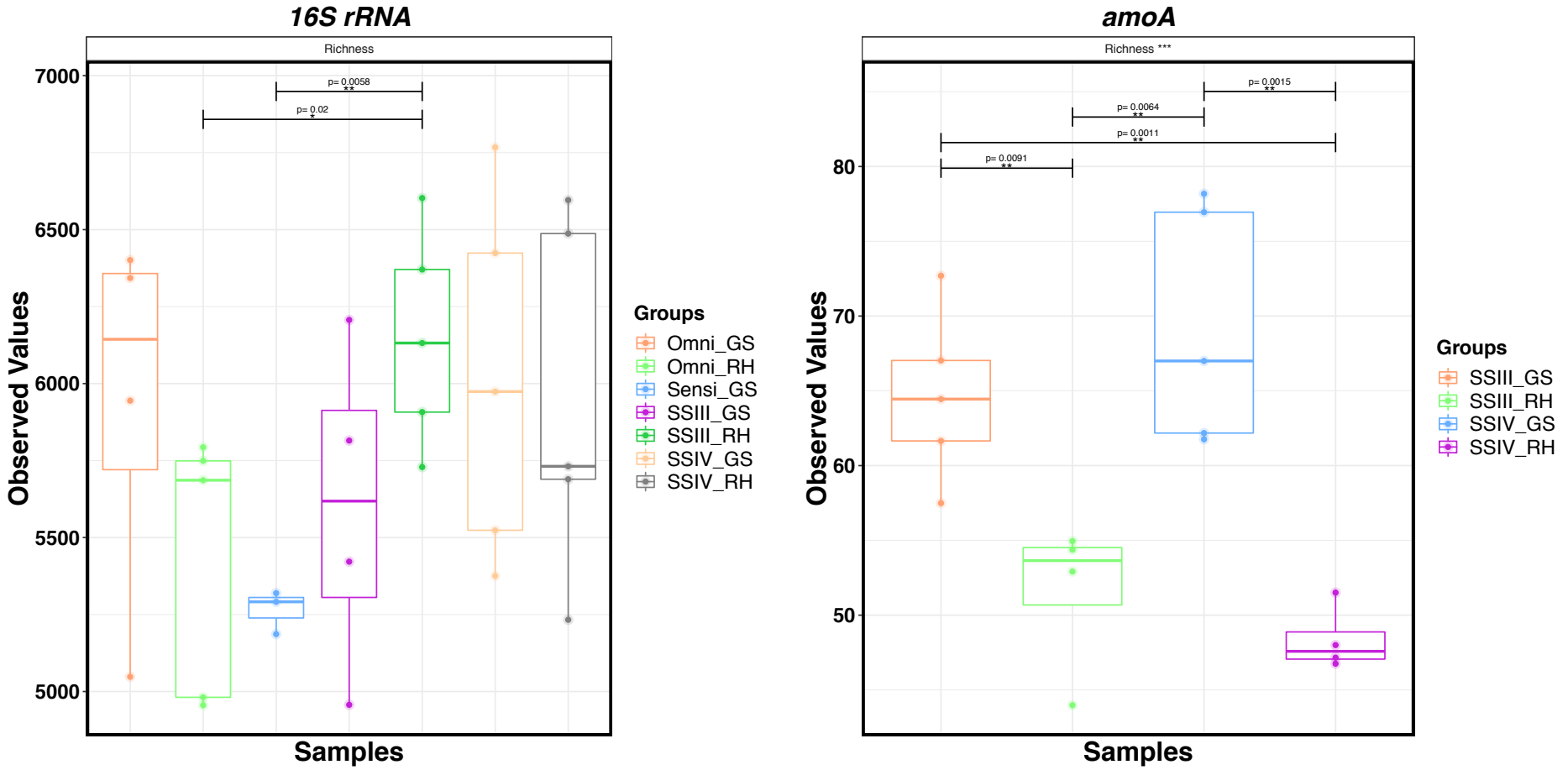


Figure 3.10 Effect of enzyme and priming on OTU richness. The number of OTUs detected for *16S rRNA* (left) and *amoA* (right) transcripts for the same sample using different RT systems was compared using two-way ANOVA. Results of the statistical tests are represented as lines on top of the plots. *: *p*.value<0.05; **: *p*.value<0.01; ***: *p*.value<0.001.

Table 3.9 Effect of enzyme and priming on *amoA* and *16S rRNA* community composition: summary of 2-ways PERMANOVA tests.

		Bray-Curtis				Unifrac				WUnifrac			
		Df	F.model	Pr(>F)	R^2	Df	F.model	Pr(>F)	R^2	Df	F.model	Pr(>F)	R^2
amoA	Enz	1	0.36	NSD	0.01	1	0.60	NSD	0.02	1	0.36	NSD	0.01
	Prim	1	21.58	**0.001**	0.60	1	10.52	**0.001**	0.41	1	21.58	**0.001**	0.60
	Enz:Prim	1	0.19	NSD	0.005	1	0.50	NSD	0.02	1	0.19	NSD	0.005
16S rRNA	Enz	3	2.42	**0.001**	0.20	3	1.42	**0.001**	0.14	3	2.47	**0.007**	0.20
	Prim	1	1.97	**0.03**	0.06	1	1.13	NSD	0.04	1	1.79	NSD	0.05
	Enz:Prim	2	1.22	NSD	0.07	2	1.08	NSD	0.07	2	1.64	NSD	0.09
16S rRNA (SSIII/SSIV)	Enz	1	0.87	NSD	0.05	1	1.05	NSD	0.06	1	0.95	NSD	0.05
	Prim	1	1.86	**0.047**	0.10	1	1.17	NSD	0.06	1	2.34	NSD	0.12
	Enz:Prim	1	0.82	NSD	0.04	1	0.97	NSD	0.05	1	0.86	NSD	0.05
16S rRNA (Sensi/Omni)	Enz	1	1.84	**0.04**	0.15	1	1.19	NSD	0.11	1	1.91	NSD	0.15
	Prim	1	1.71	**0.04**	0.14	1	1.16	NSD	0.10	1	1.88	NSD	0.15

NSD: Not Significantly Different; *: *p*.value<0.05; **: *p*.value<0.01; ***: *p*.value<0.001.

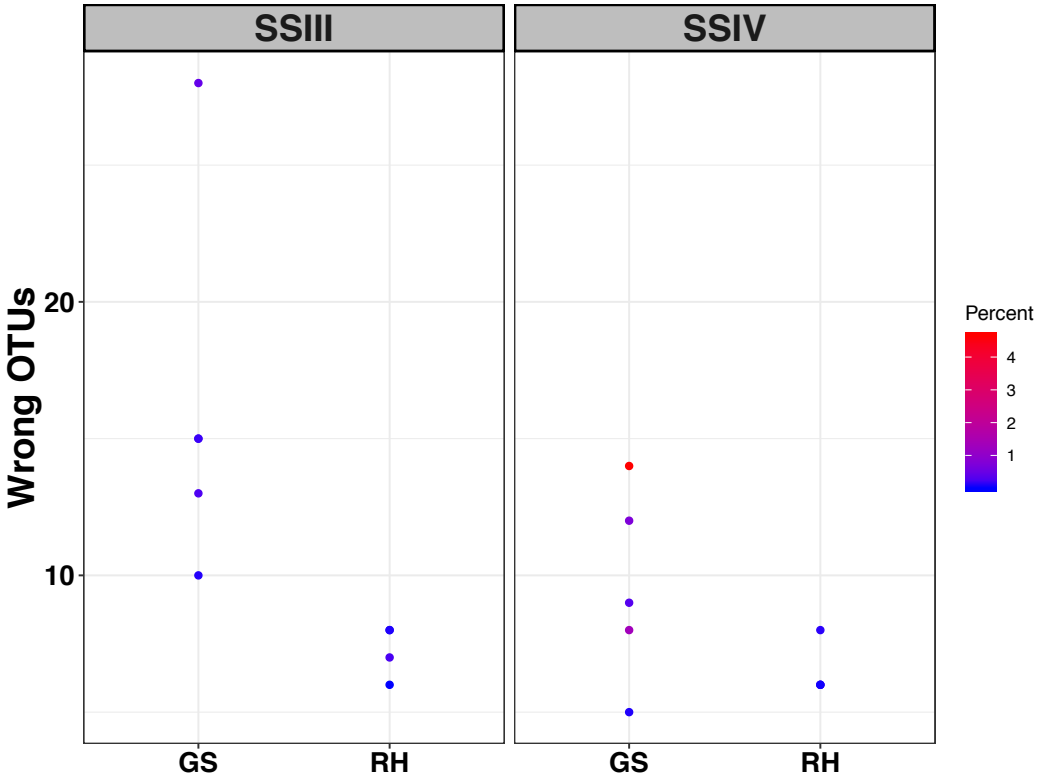


Figure 3.11 Wrong OTUs detected in the *amoA* libraries. The average number of OTUs with no assignment or wrong assignment when ran on BLASTx was plotted for each enzyme (SSIII = Superscript III; SSIV = Superscript IV) and each priming strategy (GS = Gene Specific; RH = Random Hexamer). The colour scale indicates the percentage of reads that these wrong OTUs represent compared to the total number of reads of each individual library.

Effect on amoA community composition. The outcome of *amoA* amplicon sequencing was more strongly influenced by the choice of priming (RH versus GS) than enzyme (SSIII versus SSIV)

(Figure 3.9, Table 3.9). In fact, the effect of enzyme on community composition was not significant. On the other hand, priming strategy resulted in a clear, statistically significant clustering of samples (Figure 3.9, Table 3.9). One sample, Env1, when prepared using RH priming for both SSIII and SSIV, failed to produce sufficient reads (more than 5000) to proceed and was removed from the analysis pipeline. In contrast, when GS priming was used, sufficient reads were produced to pass this quality step in the analysis pipeline. Indeed, GS priming always resulted in greater OTU richness (Figure 3.10) than RH (+13 and +21 OTUs on average for SSIII and SSIV respectively), indicating that this priming option was better at recovering the diversity of *amoA* transcripts in the samples. This observation supports the Q-PCR results where GS priming always resulted in lower Cts. To determine if the "missing OTUs" in the RH sequencing data sets were dominant or rare phylotypes, the mean abundance of OTUs present only in the GS data set was plotted for each individual OTU (Figure 3.12), revealing that most of the OTUs missing in the RH data set were low abundance OTUs. On the other hand, interestingly, a very small number of rare OTUs were only detected in the RH data set (Figure 3.12). Moreover, the choice of priming also affected the representation of OTUs present in both GS and RH datasets (Figure 3.12): with SSIII, 39 OTUs were significantly differentially expressed between GS and RH. With SSIV, it was found that 23 OTUs were significantly differentially expressed between GS and RH (Figure 3.12).

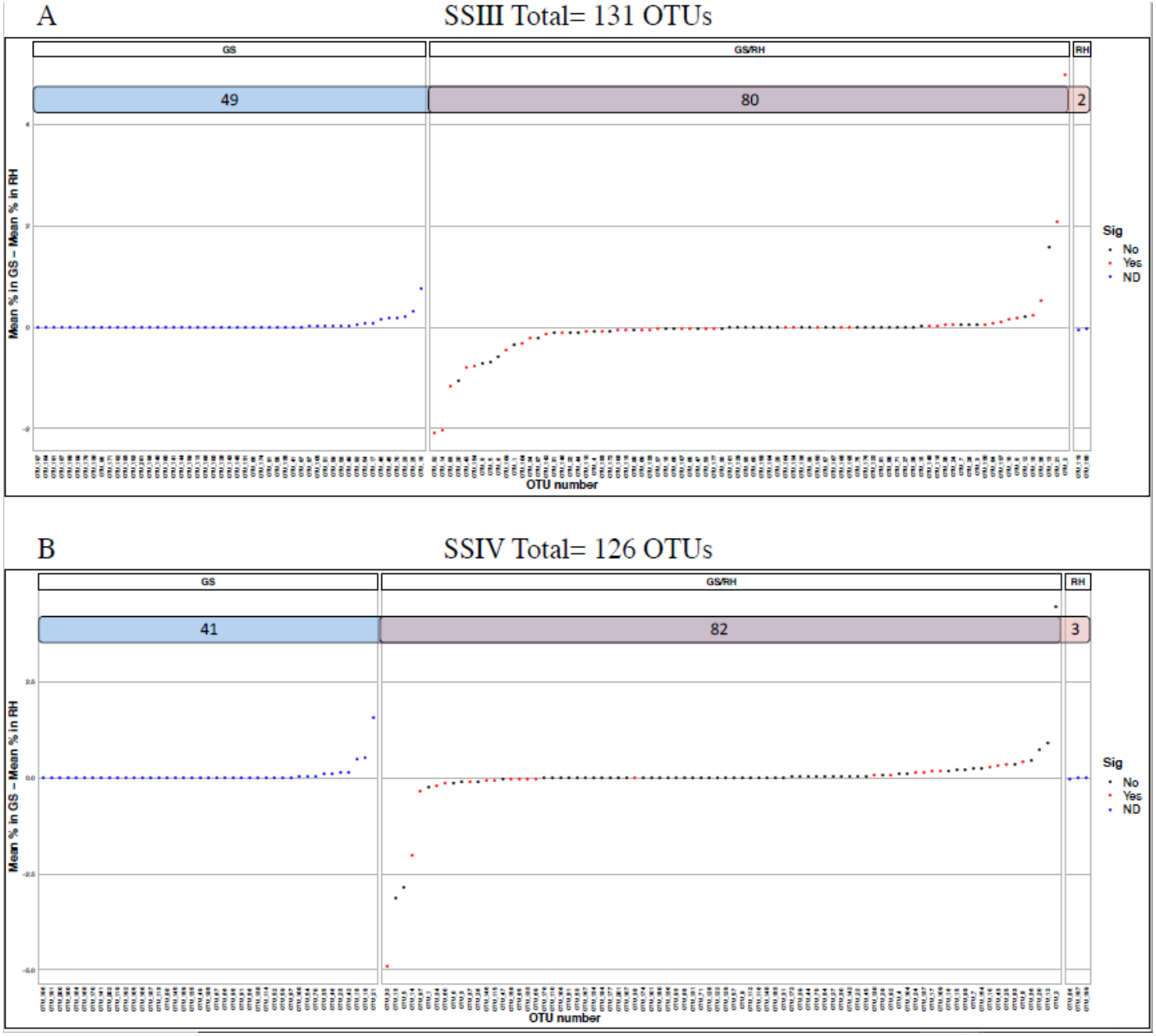


Figure 3.12 Differences in the expression and number of *amoA* OTUs detected by GS and RH priming. The Venn Diagrams on top of the plots show: the number of OTUs found in the GS data set only (blue), in both the GS and Rh data set (purple) and in the RH data set only (red). Results are presented as the average differences in proportions between GS and RH data sets (OTUs with positive values are overexpressed in the GS and inversely). When OTUs were found in only one data set, the results are presented as the average proportion of the OTU, with positive and negative value for GS and RH respectively. For the OTUs shared between GS and RH, the colour of the points indicates if the difference in expression is significant or not as explained in the legend (sig). ND= Not Determined.

Table 3.10 Effect of the RT system (Enzyme and Priming) on the community composition of the 16SrRNA libraries at different taxonomic levels. Enz = enzyme; Prim = priming; *p.*values < 0.05 in bold; NSD = Not Statistically Different

Taxonomic level		Bray-Curtis
OTU	Enz	**
	Prim	*
	Enz:Prim	NSD
Species	Enz	NSD
	Prim	NSD
	Enz:Prim	NSD
Genus	Enz	NSD
	Prim	NSD
	Enz:Prim	NSD
Family	Enz	**
	Prim	NSD
	Enz:Prim	NSD
Order	Enz	**
	Prim	NSD
	Enz:Prim	NSD
Class	Enz	**
	Prim	NSD
	Enz:Prim	NSD
Phylum	Enz	NSD
	Prim	NSD
	Enz:Prim	NSD
Kingdom	Enz	***
	Prim	*
	Enz:Prim	*

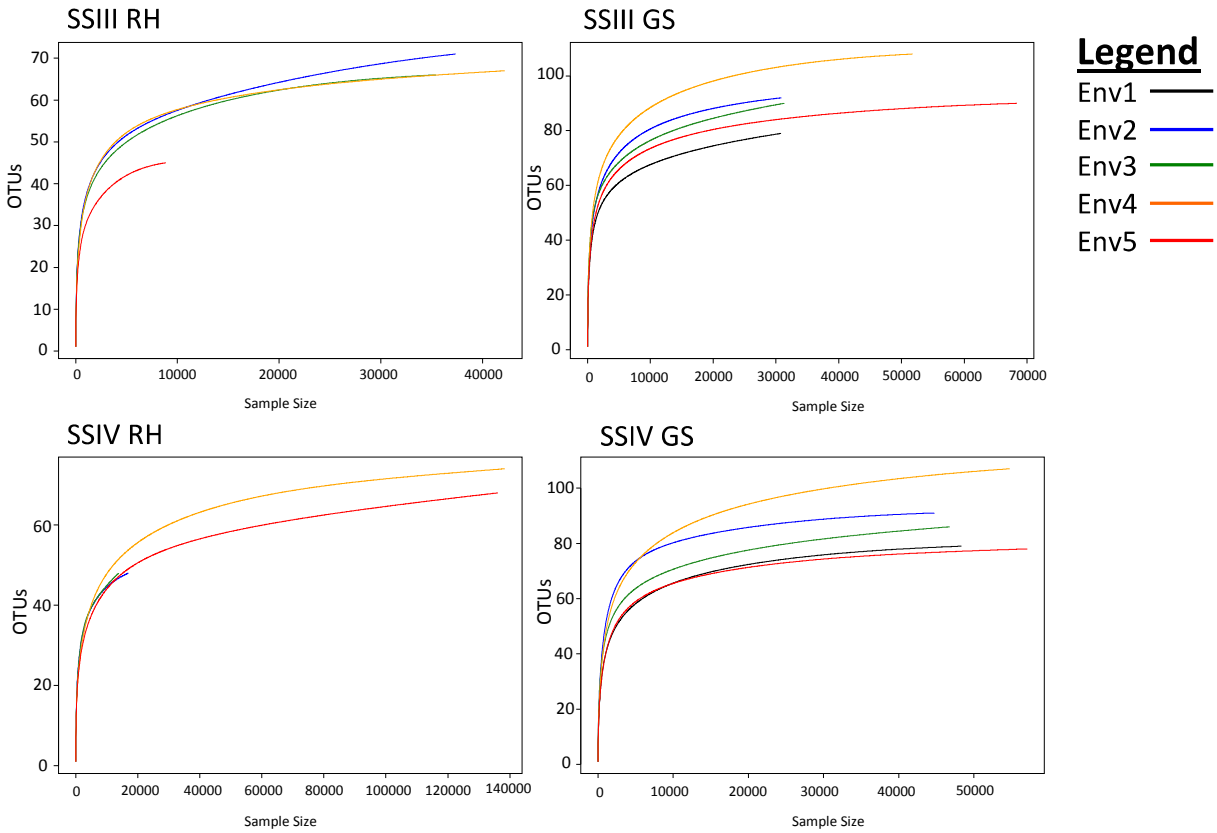


Figure 3.13 Rarefaction curves obtained from the sequencing of *amoA* transcripts from environmental samples using different RT protocols. The RT protocol used is indicated on the plots (SSIII= Superscript III; SSIV= Superscript IV/ RH= random hexamer; GS= gene specific). The rarefaction curves were drawn for each replicate as indicated on the legend.

3.4 Discussion

While RT-Q-PCR, and to a lesser extent RT-PCR amplicon sequencing, is widely used in environmental microbiology to quantify and determine the diversity of transcripts from environmental samples, the effectiveness and reproducibility of the reverse transcription step has not been evaluated. In particular, in complex environmental samples, to the best of our knowledge, there have not been any studies investigating the efficiency of the reverse transcriptase reaction to transcribe RNA to cDNA, despite this being a critical step informing the overall result. Furthermore, based on our own observations in the laboratory, we often noted the impact of different enzyme and primer choice on the same template. Therefore, we assessed the effect of the RT system (enzyme and priming strategy) on RT-Q-PCR and RT-PCR-amplicon sequencing and showed that the choice of enzyme and priming strategy can result in significant difference in both quantitative and qualitative results from the exact same sample. These methodological effects can bias and even alter final conclusions and interpretations of the underlying biological and ecological questions.

From the *sfGFP* spike experiments (Figures 3.3 and 3.4), we showed that the choice of enzyme and priming greatly affected the results of the RT-Q-PCR. When the *sfGFP* transcript was spiked into an environmental RNA background, it was found that the Superscript enzymes performed better than the Sensiscript and Omniscript enzymes. The Superscript enzymes systematically produced higher detected copy numbers, with values closer to the expected ones and, generally, differential expressions closer to the expected 5-fold difference. In a study by Levesque-Sergerie and co-workers (Levesque-Sergerie *et al.*, 2007) it was found that the Sensiscript and Omniscript enzymes had a dynamic range >50ng RNA versus >0.01ng RNA for Superscript III. Results obtained here are in accordance, with a better detection of the low concentration target by the Superscript enzymes compared to Sensiscript/Omniscript, especially when RH priming was used. Yet, the RT reactions for standard curves constructed using Sensiscript and Omniscript produced reliable Cts at target concentrations as low as 10^3 copies/µl, similar to that observed for SSIV and SSIII (except for Sensi-RH: lower limit at 10^4 copies/µl). This indicates that the lower performances observed for Sensiscript and Omniscript in the environmental spike experiment could be due to inhibition of the enzymes from co-extracted components in environmental RNA (Hata *et al.*, 2015) and/or the presence of background RNA. The later explanation contrasts with the results obtained by Levesque-Sergerie and co-workers (Levesque-Sergerie *et al.*, 2007) who observed a general increase in the recovered copies of a spike (*i.e.* lower Cts) as the concentration of background RNA from bovine tissue increased.

In this study, GS priming always performed better than RH for RT-Q-PCR, with higher copy numbers and values closer to the expected ones for the exogenous RNA spike. For the endogenous targets (*amoA* and *16S rRNA*) a similar trend was observed, with GS priming resulting in higher detected average copy numbers, except for SSIII, in the *amoA* assay and Omni in the *16S rRNA* assay. The differences between priming were particularly strong with the use of Sensiscript, where the combination of this enzyme and RH was clearly the least efficient RT strategy. Interestingly, small differences were observed between GS and RH when used with SSIV for the quantification of both the spiked *sfGFP* and the endogenous *amoA* showing that this enzyme reliably reverse-transcribed mRNAs. This was also supported by the differential expressions of the exogenous spiked *sfGFP* always being close to the expected 5-fold difference when using SSIV.

Differences in the performances of the RT enzymes and the priming strategies were similar between the *sfGFP* spike and the endogenous *amoA* mRNA with Superscript enzymes performing better than Sensiscript/Omniscript and GS generally performing better that RH. In contrast, for *16S rRNA* significant differences were detected only between Omniscript and the other three enzymes (no statistically significant differences between SSIII, SSIV and Sensiscript) and the effect of priming was very important for all four enzymes. For this assay, Omniscript in combination with RH priming yielded the highest copies/µg RNA. These differences could be a reflection of target concentration, *i.e.* highly abundant *16S rRNA* verses low abundance *amoA* transcripts, or indeed could be target dependant (*i.e.* ribosomal VS mRNA) reflecting for example the complex secondary structure of the RNA molecule.

Overall this study showed that the combination of Superscript IV with GS priming was the most accurate for the quantification of the exogenous *sfGFP* spike and showed the lowest variation in quantification when priming was changed to RH. SSIV GS was also the RT system yielding, on average, the highest copy number for the quantification of *amoA* mRNAs by RT-Q-PCR, coupled with the best precision (lowest standard deviation). This fits with our previous observations, where we would routinely achieve better results (*e.g.* detection verses no detection) and our subsequent choice of the Superscript enzyme with gene specific priming to quantify a range of N-cycle mRNA targets (Duff *et al*., 2017, Smith *et al.,* 2015, 2007).

Next, we investigated the effect of the RT strategy on cDNA sequencing. The result from cDNA sequencing demonstrated that the enzyme and priming strategy employed has an impact on cDNA amplicon diversity. As Sensiscript and Omniscript failed to reliably produce sufficient cDNA to produce PCR amplicons for *amoA*, they were not included, nor was the combination of Sensiscript and RH for the *16S rRNA* diversity study. We have shown that for *amoA* transcripts, priming is an important consideration (Figures 3.9, 3.12, 3.13; Table 3.9). Most notably, the use of RH for sample Env1, resulted in too few sequencing reads (<5000) for further analysis. We attribute this to the lower abundance of the *amoA* transcript in a high background of RNA. In this case, the choice of priming made the difference between the success or not of the amplicon sequencing of the transcript. This result was in line with observations from the RT-Q-PCR for *amoA* (Figure 3.6). Overall more OTUs (Figures 3.10) and better coverage of *amoA* transcript diversity (Figure 3.13) were obtained when GS priming was used. The differences in the number of OTUs detected was particularly important for low-abundance OTUs indicating that GS priming was better for the reverse transcription of rare members of

the *amoA* community (Figure 3.12). A possible explanation is that, for GS priming, all the RT resources (enzyme and dNTPs) are directed to the reverse transcription of the target transcript. On the other hand, when using RH priming, random priming may not be sufficient to prime rare mRNA target.

This observation is further supported by the results from the *16S rRNA* assay, where the choice of priming strategy was seen to be less important. In fact, here most of the differences observed were due to enzyme choice and not priming strategy. In contrast to the *amoA* results, for *16S rRNA* the use of GS priming did not necessarily result in a higher number of OTUs compared to RH even though GS priming resulted in a higher number of *16S rRNA* copies detected by Q-PCR. It may be that differences in RT performances are abundance or target molecule dependant (*i.e.* very abundant ribosomal RNA with complex secondary structures versus rare messenger RNA).

As the true representation of our transcripts in the environmental samples was unknown, we tested the RT systems against artificial defined RNA mock communities seeded into background environmental RNA. These artificial sequences were derived from target inserts with additional cloning vector sequence added, which allowed for their selective amplification from the background. To evaluate the bias introduced by the PCR/sequencing steps and separate them from the RT, a similar experiment was carried out using DNA mock communities. This experiment revealed that, for all RT systems, biases were introduced in both the RT and the subsequent PCR step of the reaction as the recovered proportions deviated from the expected ones (Figure 3.7; Figure 3.8). When testing a new approach for *16S rRNA* transcript sequencing based on ligation of an adapter to the end of the gene prior to RT with random hexamers, Yan *et al.*, 2017, found errors in the observed ratios of their RNA mock communities of up to 3-fold compared to the expected proportions. These results are comparable to those found in this study. Here, we found that the smallest amount of variation from the expected EM composition was observed with SSIV RH. In fact, surprisingly, RH priming always conserved the actual proportions better than GS priming in the seeded mock communities, as seen by lower standard deviations (Figure 3.8). Considering that the RNA template for the mock community construction went through both *in vitro*-transcription and a RT reaction prior to PCR, each of which could introduce errors, the standard deviation observed in the RNA mock communities (*i.e.* both GS and RH) was low and in fact, for RH, the same as the DNA mock (4.97 for SSIII GS; 3.31 for SSIII RH; 3.89 for SSIV GS; 3.15 for SSIV RH and 3.34 for DNA) (Figure 3.8).

As anticipated, errors were also seen in the UM resulting in observed regressions deviating from the expected. Interestingly, the errors were consistent between EM and UM (*i.e.* a sequence over-represented in the EM would also be over-represented in the UM and vice versa). As a result, when the UM proportions were corrected with the EM ones, the observed regressions were close to the expected y=x (Figure 3.7). Since the mock communities were constructed separately (Figure 3.2), this indicated that: 1) these errors are a reflection of sequence specific bias of the RT-PCR workflow and not attributed to user error such as pipetting; 2) Since artificial over/under representations is likely introduced by sequence specific bias, the relative abundance of transcripts within a sample (α diversity) might not always be absolute when small differences (*e.g.* ≈ 4-fold as in this study) in expression are observed; 3) However, as these biases are reproducible (UM reads corrected by EM reads), comparison between samples (*i.e.* β diversity) can be undertaken.

General Conclusion: Best Practice for Environmental RNA

The challenge when working with environmental samples will always be to retrieve RNA of a high enough quality and integrity. Here, we started with RNA extracted from marine sediments that had an average RIN of ≈7 and R_{amp} of ≈0.8. This is the best quality RNA we could produce with this bead-beating co-extraction method (Griffiths *et al*., 2000) and it already falls at the lower end of acceptable RIN for pure culture (Jahn *et al*., 2008). Therefore, methods to improve the initial quality of RNA extractions should also be a high priority, although this will be easier in some environments than others. Improvement of extraction methods is crucial as it can lead to important differences in the results. For example, Feike *et al* (Feike *et al*., 2012) showed that different sampling techniques influenced the relative abundance of transcripts retrieved from the suboxic zone of the Baltic Sea.

Another consideration raised by this work is in the very fact that the differential amplicon approach works. This shows that small cDNA amplicons can still be produced from highly degraded RNA samples whereas long amplicons tend to disappear quickly. When using RNA samples of poor quality, the comparison of expression levels between different targets might be irrelevant if the difference in length of the RT-Q-PCR targets between genes is large. In this case, it would be better to use only small amplicons, that are less sensitive to degradation (Antonov *et al*., 2005). An alternative, to deal with samples with different degradation status,

potentially could be to normalize RT-Q-PCR data to RNA integrity. A RIN based algorithm has been proposed by Ho-Pun-Cheung *et al* (Ho-Pun-Cheung *et al*., 2009) to reduce RT-Q-PCR errors due to RNA degradation in cancer biopsies. In our case, however, R_{amp} indexes correlated better than the RIN with *amoA* and 16S rRNA Cts, making them better potential candidates as normalization metrics. Therefore, we tested a normalization coefficient based on the R_{amp} (Table I.1, Figure I.2). As in Ho-Pun-Cheung *et al.,* 2009, we assumed a linear relationship between the integrity index and the changes in transcript Cts (i.e. change in Ct = α × change in R_{amp}). This assumption facilitated the calculation of a regression coefficient α that was used to normalize Cts as explained in Figure I.2. Although the use of such normalization reduced the errors attributable to RNA degradation (Figure I.2), several limitations remain: (i) the linear relationship between changes in Cts and R_{amp} might not always be true depending on the transcript tested, (ii) the regression coefficient α depends on the degradation technique (Table I.1), (iii) the regression coefficient α depends on the transcript tested (Table I.1) and (iv) the regression coefficient α may depend on the environment from which RNA was extracted.

Table I.1 Summary of the regression coefficients associated with the equation: change in Ct = f(change in R_{amp}). The coefficients are calculated assuming a linear relationship.

Gene	Degradation	R_{amp} 380/120		R_{amp} 380/170	
		Slope (α) (p-value)	Adjusted R-squared	Slope (α) (p-value)	Adjusted R-squared
amoA	Heat	12.75 (0.007)	0.98	14.31 (0.007)	0.98
	UV	12.07 (0.007)	0.98	13.56 (0.007)	0.98
	RNase	11.34 (0.002)	0.90	12.50 (0.0009)	0.94
16S rRNA	Heat	4.03 (0.0445)	0.87	4.52 (0.0441)	0.87
	UV	2.78 (0.0050)	0.99	3.12 (0.0045)	0.99
	RNase	6.50 (0.0007)	0.95	7.09 (0.0004)	0.96

In a recent review about the use of RT-Q-PCR, Bustin and Nolan (Bustin & Nolan, 2017) stated that *"the majority of published RT-Q-PCR data are likely to represent technical noise"*. The intrinsic variability of the RT step and the lack of information on protocols used were key points that lead them to this striking conclusion. This is likely to be similar, if not further amplified in complex environmental samples, from which ecosystem conclusions are drawn. In Chapter III we have shown that primer and RT system choice can range from no detection to a 600-fold difference in transcripts for the same template. In environmental studies, this is the difference between no gene expression to the presence of a highly active transcript – striking difference leading to opposite ecosystem conclusions. There is therefore an urgent need to ensure that the approaches we use are tested and recommendations as far as possible for best practice are made, followed and reported in future studies. Our studies show that the choice of correct enzyme and priming can improve the reliability and reproducibility of RT-Q-PCR and RT-sequencing data, facilitating insight into the transcriptionally active microbial communities directly from the environment. This, taken together with steps to monitor the purity and integrity of the extracted RNA prior to downstream analysis (Bustin and Nolan, 2017; Chapter II) and detailed documentation of the RT approach used should greatly improve the reliability and reproducibility of transcript based studies in environmental microbiology.

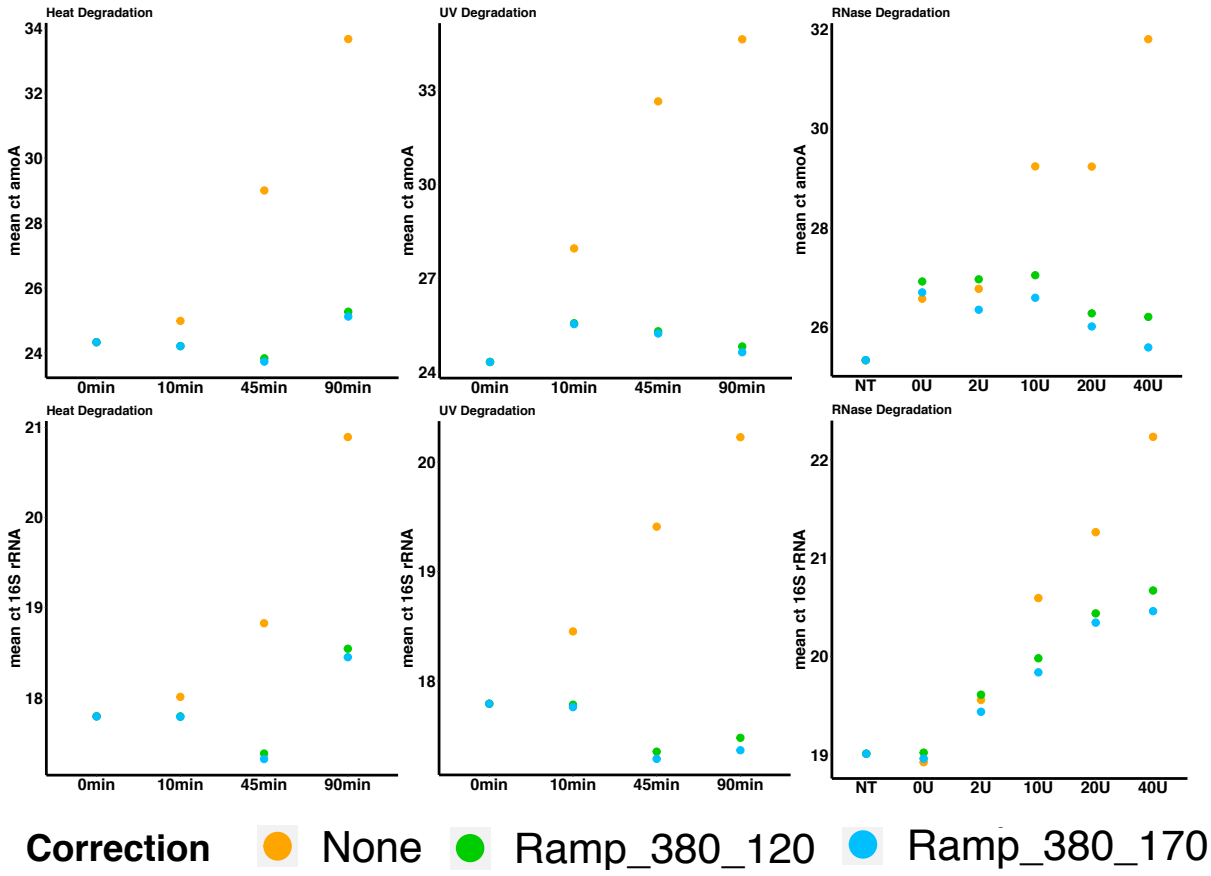


Figure I.2 Normalisation of *amoA* and *16s rRNA* RT-qPCR results to RNA integrity. The correction of the Cts was done by assuming linear relationship between the change in Cts and the change in R_{amp} index along the degradation gradient *i.e.* change in Ct = α x change in R_{amp}. Cts corrected for RNA integrity was then calculated as follows:

$$\text{corrected Ct} = Ct - (\alpha \times (R_{amp}t_X - R_{amp}t_0))$$

with $R_{amp}t_X$ corresponding to the R_{amp} at a degradation point X and $R_{amp}t_0$ corresponding to the R_{amp} at the initial point. Both R_{amp} 380/120 and R_{amp} 380/170 were used to calculate the correction coefficient α.

Based on the results obtained in chapter II and III, we propose a set of recommendations to improve transcriptomics results in microbial ecology:

1: Evaluate and report RNA quality and integrity

It was shown in chapter II that both quantitative (RT-Q-PCR) and qualitative (sequencing) results can be obtained, even from very degraded samples. Comparison of gene expression level between preparations with different degradation levels can therefore lead to false conclusions if integrity is not checked prior to analysis. Assessing RNA quality is therefore essential for

obtaining meaningful transcriptomic results. The current approach to monitor RNA integrity includes the RIN and RQI. These are useful techniques that are widely under-used (or reported) in microbial transcriptomics studies, to give an overview of total RNA quality based on a ratio between the 23S and 16S ribosomes. Since most transcriptomics studies are interested in the metabolic function and therefore mRNA, it is preferable to have an integrity index to target the mRNA. Furthermore, it is unknown if degradation of rRNA reflects mRNA degradation. We therefore recommend using the Ramp index is complementation to the RIN/RQI as it performed better at predicting the outcome of RT-Q-PCR of a functional gene (*amoA*).

2: RT-Q-PCR

i) Gene specific priming was more accurate, precise and sensitive than random hexamer priming for mRNA.

ii) Of the enzymes tested, Superscript IV was accurate, precise and sensitive, and therefore we recommend its use for the detection of transcripts in complex environmental RNA matrixes.

iii) The incorporation of an exogenous RNA target at known concentration into the environmental RNA being tested is an efficient way to validate RT-Q-PCR protocols.

iv) When converting Ct results into copy number, we advise the use of an RNA standard curve (*i.e.* serial dilution of the target RNA and individual RT-Q-PCR) rather than a cDNA standard curve (*i.e.* reverse transcription of a fixed concentration of RNA, dilution of the cDNA and Q-PCR).

v) Fully report the RT protocol used.

3: RT-amplicon sequencing

i) For RT-amplicon sequencing of mRNA targets, we recommend the use of gene specific priming as it resulted in better coverage and higher OTU richness of the bacterial *amoA* transcript. For *16S rRNA* RT-sequencing, the choice of priming is less important.

ii) The addition of RNA mock communities into environmental RNA (before reverse transcription) can aid interpret sequencing results: in our case, we deduced from our RNA mock communities that even though relative proportions of individual OTUs within a sample (α diversity) can be biased, the comparisons of changes in OTU composition between samples (β-diversity) are reliable.

Experimental Part II Application of the Optimised Workflow: Effect of Self-Organised Sedimentary Structures on Microbial Communities and Nitrogen-Cycle Related Activities in an Intertidal Mudflat

Contributions: Chapter IV will start by presenting some data that was not carried out within the context of this PhD thesis: Initial samples from the mudflats were collected by Agatha Lisik (AL) (National University of Ireland, Galway) in July 2016 for a MSc project, with the help of Philippe Pineau (PP), Hélène Agogué (HA) and Nicolas Lachaussée (NL) (LIENSs - La Rochelle Université). Nutrient Analysis was carried out by AL and PP. Potential rates measurement was carried out by AL with the help of Cindy Smith (CJ) (University of Glasgow). Chlorophyll measurement was carried out by AL with the help of Martine Bréret (MB) (LIENSs - La Rochelle Université). The data analysis from the generated data sets was carried out by Fabien Cholet (FC). In addition, all the molecular analysis on this data set was carried out by FC (University of Glasgow) as part of this PhD thesis. Sequencing data processing was carried out by FC with the help of Umer Ijaz (UI) (University of Glasgow). Statistical analyses and report writing were carried out by FC.

The 2019 sampling campaign and associated work presented later in chapter IV was entirely carried out within the context of this PhD thesis. FC carried out the sample collecting and processing in 2019 with the help of PP, HA, NL and CS. Nutrient Analysis was carried out by FC and PP. Potential Nitrification Rates were carried out by FC with the help of CS. Chlorophyll and pheophytin measurement were carried out by FC with the help of MB. DNA/RNA extractions and subsequent molecular work were carried out by FC. Sequencing data processing was carried out by FC with the help UI. Statistical analyses and report writing were carried out by FC.

Chapter IV Effect of the Ridge/Runnel Sedimentary Structure on Microbial Composition and Nitrifiers' Activity in the Marennes-Oléron Bay, France

Based on the preliminary results obtained in the 2016 sampling campaign and the molecular work carried out during this PhD thesis, we decided to further investigate the differences in ammonia oxidizers' activity between ridges and runnels. Indeed, as seen in this preliminary study, AOB abundances and PNR show negative correlations, with AOB more abundant in ridges while PNR was higher in runnels (Figure 4.1), thus indicating that gene quantification alone is not a valid proxy to estimate ammonia oxidation in this environment:

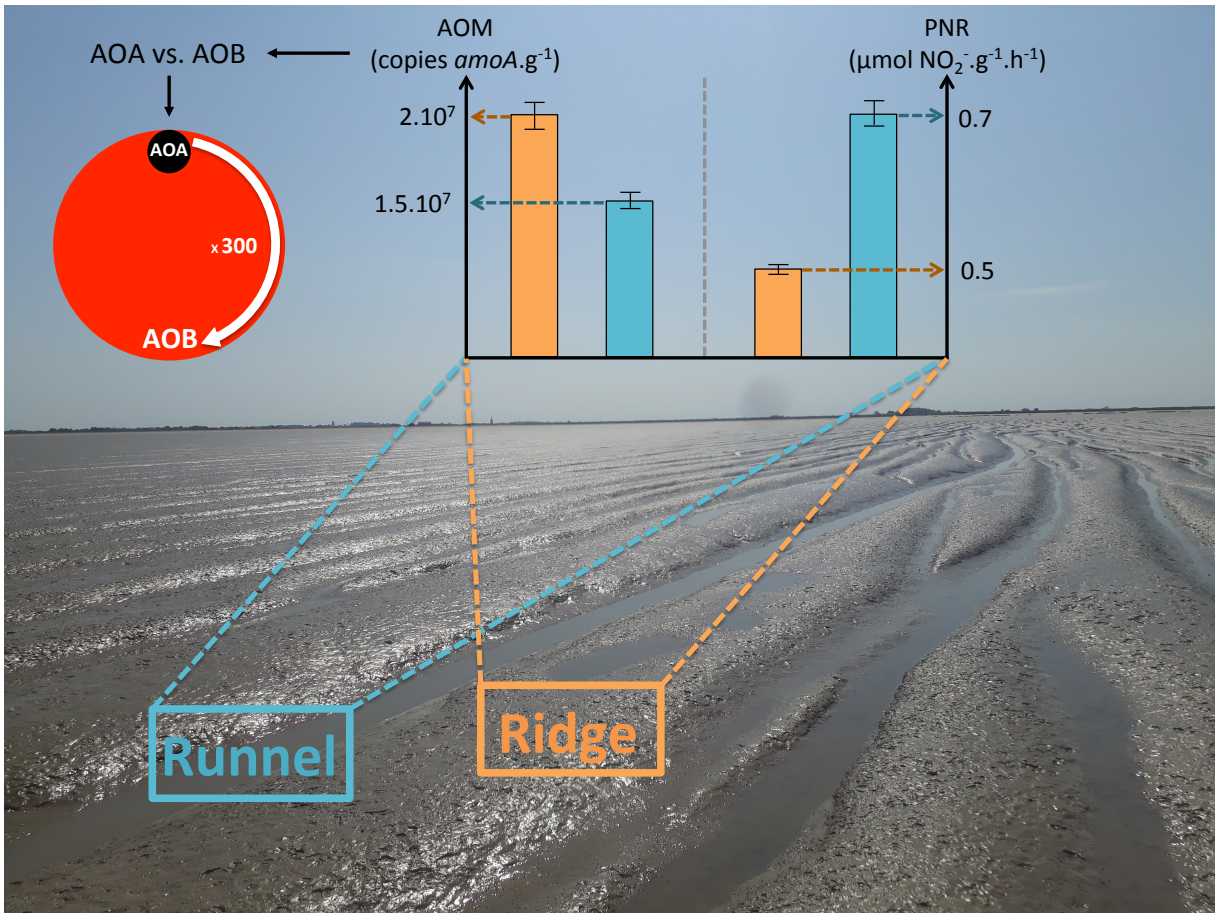


Figure 4.1: Summary of the findings of the 2016 sampling campaign and the subsequent molecular work carried out during this PhD thesis. The plot at the middle represents the significant differences in AOM abundances (left) and PNR (right) between ridges (orange) and runnels (blue). The circle on the left represents the difference in AOA and AOB abundances (AOB ≈ 300 times more abundant).

We therefore hypothesised that AOB, despite being less abundant in runnels, will be more active. To obtain samples suitable for transcriptomic analyses, a new sampling campaign was set-up in summer 2019. PNR and physiochemical parameters have also been measured. To increase the resolution of this study three different sites were surveyed and 10 biological replicates (5ridges /5 runnels) were collected at each site. We hypothesise that the observed differences in rates can be explained by differences in nitrogen cycle-related transcript quantity and/or composition between ridges and runnels.

4.1 General Introduction

Estuaries are transition zones where freshwater from the continent meets saline water from the ocean. These coastal ecosystems are among the most valuable environments as they provide both economical (Costanza *et al.*, 2014) and ecological services. They are used as feeding grounds for a wide variety of animals including invertebrates, fishes and migratory birds. They also provide nursery ground for economically relevant fishes and shellfishes (Foster et al., 2013; Seitz et al., 2014). On intertidal mudflats, the alternation of very different conditions due to the tidal and diurnal cycles combined with the concentration of a high amount of nutrients in a small area produces very diverse habitats displaying a high level of biological activity (McClain *et al.*, 2003). Estuaries are therefore important on a global scale as they control the budget of nutrients transferred from land to the sea and the atmosphere (Foster et al., 2013).

These ecosystems are also very impacted by human activities, as they are the recipients of most of the pollutants generated inland (Agardy *et al.*, 2005; Seitz *et al.*, 2014). Among those pollutants, which loading has greatly increased since the beginning of the industrial area, is nitrogen, especially since the invention of the Haber-Bosh process in 1906 that permitted the man-made conversion of gaseous nitrogen into reactive forms to generate vast quantities of inorganic fertilizer. Nitrogen is an essential element for life and supports primary production in the natural environment, but excess concentration leads to problems such as eutrophication and global acidification (Gruber and Galloway, 2008). In the environment, nitrogen transformation is largely mediated by microorganisms. The major steps of the microbial nitrogen cycle include: Fixation, the reduction of inorganic atmospheric nitrogen (N_2) into ammonia and the subsequent assimilation into organic matter; Nitrification, the chemo-litho-autotrophic process that results in the oxidation of ammonia to nitrate, which can be carried out as a one or two steps process (complete nitrification or nitritation/nitratation); Denitrification, the reduction of oxidized forms of nitrogen back to N_2 or intermediary forms such as NO or N_2O. Denitrification

is generally a chemo-organo-heterotrophic process carried out by microbes able to replace oxygen with nitrate or nitrite as the terminal electron acceptor. Alternative forms of oxidized-nitrogen utilisation for metabolism include the Dissimilatory Nitrate Reduction to Ammonia (DNRA) and the Anaerobic Ammonia Oxidation (ANAMMOX) pathways (Bianchi, 2007; Damashek and Francis, 2018). Each of these steps is controlled by different microbial groups that possess specific enzymatic machinery to catalyse a given process. For the two-step ammonia oxidation pathway, nitritation is carried out by both Archaea (Ammonia Oxidizing Archaea; AOA) and Bacteria (Ammonia Oxidizing Bacteria; AOB), possessing the *amo* (ammonia monooxygenase) and *hao* (hydroxylamine oxidoreductase) genes, whereas nitratation is carried out specifically by Bacteria (Nitrite Oxidizing Bacteria; NOB) possessing the *nxr* (nitrite oxydoreductase) genes (Arp *et al.*, 2002; Prosser, 2005; Bianchi, 2007; Prosser and Nicol, 2012; Ward, 2013). Denitrification and DNRA are carried out by a wide range of phylogenetically unrelated groups of microorganisms and require numerous genes, including the *nar/nap* (nitrate reductases), *nir* (nitrite reductase), *nor* (nitric oxide reductase), *nos* (nitrous oxide reductase) and *nrf* (*napA*-associated nitrite reductase) genes (Hayatsu *et al.*, 2008; Hallin *et al.*, 2009; Giblin *et al.*, 2013). ANAMMOX is restricted to five Candidatus genera affiliated with a monophyletic group in the phylum *Planctomycetes*. Several genes are involved in this process including a nitrite reductase *nir* and two genes specific to ANAMMOX bacteria, the *hzs* (hydrazine synthase) and *hzo* (hydrazine oxidoreductase) genes (Damashek and Francis, 2018).

Microbial denitrification can attenuate the excess load of nitrogen via the conversion of dissolved nitrite/nitrate to gaseous N_2. (Hou *et al.*, 2012; Damashek *et al.*, 2015). Nitrification, the conversion of ammonia to nitrate, therefore plays a central role as it controls the concentration of available substrate for denitrification (Seitzinger *et al.*, 2006). Identifying active nitrifiers in the environment and understanding what factors regulate their metabolisms is therefore of great importance. In estuarine sediment, the main environmental drivers of nitrifiers activity include salinity (Freitag *et al.*, 2006; Jane M Caffrey *et al.*, 2007; Zhang, Chen, Dai, Tian, *et al.*, 2015; Duff *et al.*, 2017; Santos *et al.*, 2018; Zhang *et al.*, 2018), ammonia concentration (Urakawa *et al.*, 2014; Damashek *et al.*, 2015; Duff *et al.*, 2017), oxygen availability (Abell *et al.*, 2011), temperature (Zheng, Hou, Liu, *et al.*, 2013), pH (Zhang *et al.*, 2018), light intensity (Merbt *et al.*, 2017), sulphide and iron concentration (Dollhopf *et al.*, 2005; Jane M Caffrey *et al.*, 2007) and the presence of phototrophs (Risgaard-Petersen *et al.*, 2004). On the Montportail-Brouage mudflat located in the Marennes-Oléron

bay, the physical arrangement of the sediment in ridges and runnels structures has been shown to significantly affect microbial nitrification with higher rates measured in runnels compared to ridges (Laima *et al.*, 1999, 2002). The authors hypothesised that these differences could be explained by a higher abundance of AOB in runnels, driven by higher ammonia availability.

The Marennes-Oléron bay covers a 150 km^2 area in the middle of the French Atlantic coast. This intertidal bay is characterized by a gentle slope (1:1,000) and receives input of mixed turbid water from the Charente River and seawater from the Atlantic ocean. The eastern part of the bay is composed of two large intertidal mudflats, Montportail-Brouage (north) and Mérignac (south). Together, they cover a 10 km region from north to south, about 3-4 km wide. On these mudflats, three zones can be distinguished: 1) the upper mudflat, extending ≈600m from the shore, that presents either flat or pseudo ridge/runnels structures depending on meteorological conditions and 2) the middle mudflat found between ≈600m to ≈3600m from the shore. The middle mudflat is composed of three parts, the uppermost, middle and lower parts, 700m, 1000m and 1300m long respectively. Ridges and runnels structures, with the same NW-SE direction, are found on all parts of the middle mudflat and get wider and deeper further from the shore. Channels are observed in this part of the mudflat and together with the ridge/runnel structures, constitute the major drainage systems in the Bay. Finally, 3) the lower mudflat, ≈3600m to ≈4500m from the shore characterized by reduced hydrodynamic energy and wider channels (Gouleau *et al.*, 2000). The Montportail-Brouage is an important site both from an economic (oyster and mussel farming) and ecological point of view. Indeed, the bare mudflat is a zone of high primary production, carried out by the microphytobenthos composed of diatoms. At low tide, the abundant diatoms and associated bacteria form a biofilm in the upper layers of the sediment, which constitute the basis of the benthic and pelagic food web. Indeed, this biofilm is consumed by the macrofauna, like the gastropod *Hydrobia ulvae*, the bivalves *Macoma balthica* and *Scrobicularia plana,* nematodes and copepods which then constitute the basis of the diet for shorebirds and fishes. The Montportail-Brouage is considered one of the major stopover site for migratory birds in France (Haubois *et al.*, 2005; Pascal *et al.*, 2009; Saint-Béat *et al.*, 2013, 2014).

The microbial communities in the Montportail-Brouage mudflat present a high degree of stratification and their distribution is strongly affected by salinity and nutrient and to a lesser extent, by the consumers (meiofauna) (Lavergne *et al.*, 2017). Interestingly, Lavergne et al (2018) also showed that the distribution of prokaryotes in the sediment layers is not affected by

diel and tidal rhythms. However, *16S rRNA* might not fully describe the changes in active microbial populations (Blazewicz *et al.*, 2013).

The aim of the next chapter is to further investigate the influence of the ridge/runnel structure on microbial composition and activity related to nitrification and nitrite oxidation. This chapter will focus on the main results and findings of the 2019 campaign that was carried out during the course of this thesis to test hypothesis drawn from the results obtained during the previous campaign. Here, we will focus on the effect of the ridge/runnel structures on ammonia oxidizers communities and activities. Specifically, we hypothesised that the same communities are present in both structures but are more active in the runnels, explaining the differences in nitrification rates.

4.2 Material and Methods

4.2.1 Sampling and field measurement

Sediment sampling was carried out at low tide on the 1st, 2nd and 3rd of July 2019 on the North station, VASIREMI station and south station, respectively (Figure 4.2). For each day, 10 biological replicates were collected from 5 ridges and 5 corresponding runnels structures (Figure 4.2 and 4.3) covering and area of approximately 100m^2. The sample labelling thereafter is as follows:

Dx(**B**/**S**)y

With D the sampling day (x=1, 2 or 3 for sampling days 1st, 2nd and 3rd of July respectively); B/S the type of sample (B for ridges and S for runnels) and y the biological replicate (“a” to “e” for biological replicate 1 to 5 respectively).

For each biological replicate, approximately 6 cores were taken along a distance of approximately 2-3 meters. Only the top 1cm of the sediment cores were collected and brought back to the beach in boxes at *in situ* temperature (Figure 4.4 and 4.5). Sediment temperature was determined on site. For every biological replicate, sediment from the different cores was homogenised by mixing, then:

For DNA/RNA extraction: approximately 2g of sediments were flash frozen in syringes using liquid nitrogen. They were transported from the sampling site to the laboratory using a dry shipper, then stored at -80°C.

For Flow cytometry: approximately 2g of sediment was mixed with 2ml of 2% glutaraldehyde. They were transported from the sampling site to the laboratory in cold boxes (≈0°C), then stored at -20°C.

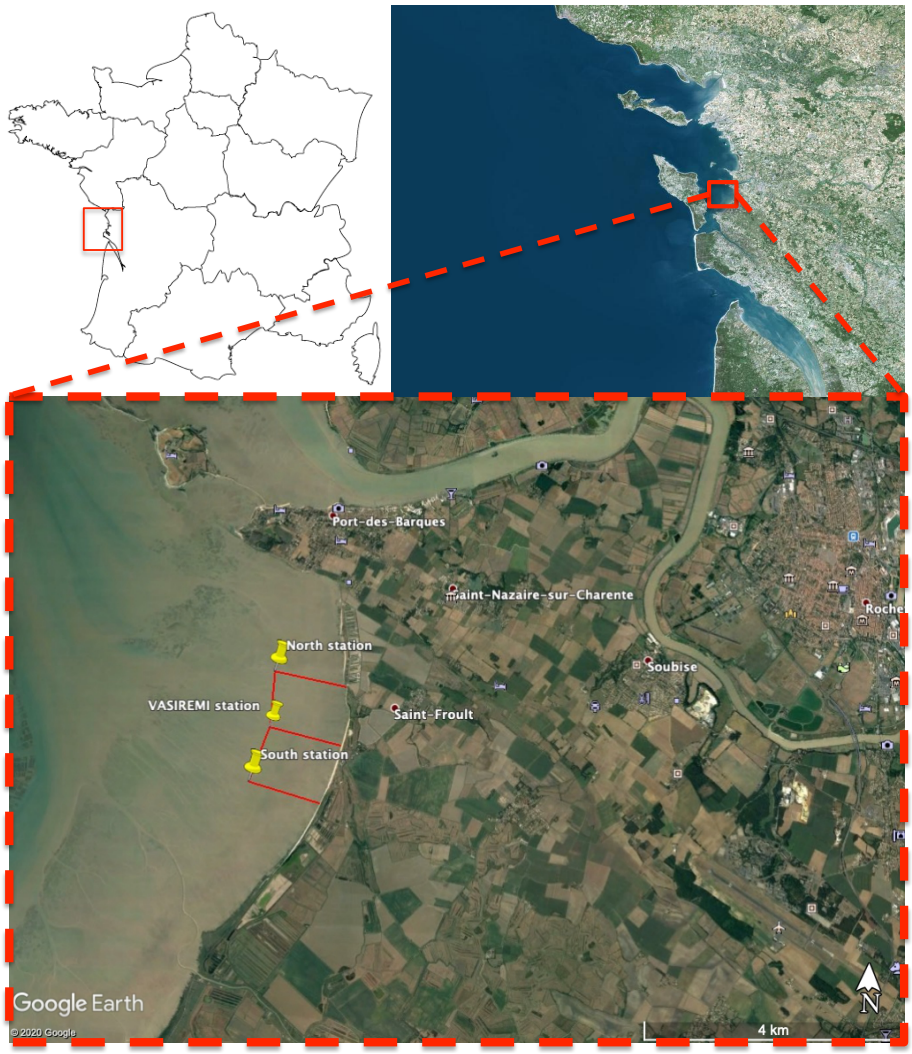


Figure 4.2 Map of the Montportail-Brouage mudflat along the French Atlantic coast.

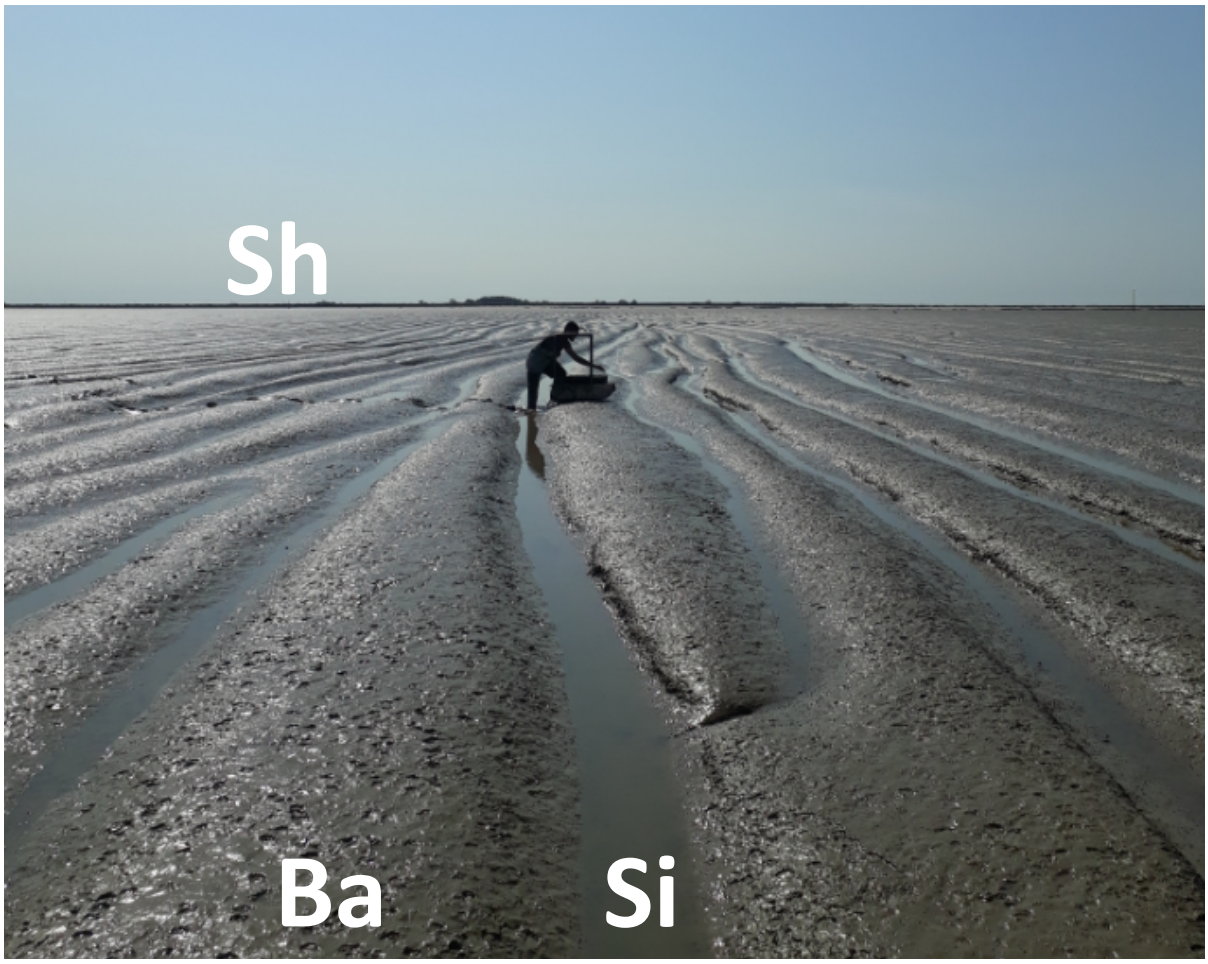


Figure 4.3 Photography of ridge/ runnel structures in the Brouage Mudflat. The photography shows the succession of ridges (Ba) and runnels (Si) perpendicularly to the shore line (Sh). Photography taken on July 2019 (credit: Philippe Pineau, LIENSs, Université de LaRochelle).

For PNR and other bio-physico-chemical analysis: Sediment samples were transported to the laboratory in a cold box (≈4°C) temperature. Samples were stored at -20°C if not used immediately except for the granulometry analysis (samples stored at room temperature).

4.2.2 Biological and physio-chemical analysis

Nutrient analysis

Nutrient content was determined on pore water extracted by centrifugation of the sediment samples and filtration through 0.2µm cellulose nitrate filters (samples diluted 1/10 for ammonia): ammonium, nitrite, nitrate, phosphate and silicates concentrations were determined using an autoanalyzer (Seal Analytical, GmbH Nordertedt, Germany) equipped with an XY-2 sampler. Salinity and pH were measured directly on unfiltered pore water.

Density and Water content

Density was determined by weighting a known volume of sample. Water content was determined by weight loss before and after incubation at 60°C for 48h.

TOC, DOC, DON and TN

Sediment Total Organic Carbon (TOC) was determined by weight loss of dried samples incubated at 450°C for 12h. Pore water TOC, Dissolved Organic Carbon (DOC) and Total Nitrogen (TN) were measured on a Shimadzu TOC-L Analyzer. TOC was measured on unfiltered and DOC on filtered samples after acidification with HCl to pH below 2 and flushing with air for 5 minutes to remove inorganic carbon. Pore water total carbon/ total nitrogen ratio was calculated as (pore water TOC + pore water DOC)/ (pore water total nitrogen).

Granulometry

Sediment granulometry was measured on a Mastersizer laser granulometer 2000 equipped with a Hydro MU sampler. Clay was defined as particles with diameter smaller than 2µm, silt as particles with diameter between 2µm and 63µm and sand as particles with diameter higher than 63µm.

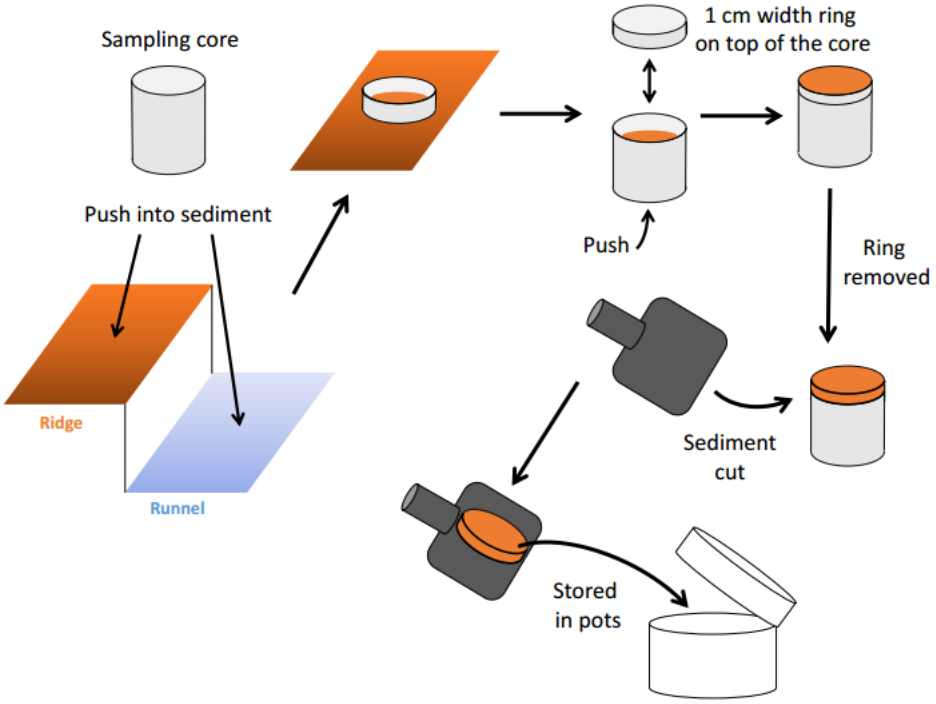


Figure 4.4 Schematic representation of the procedure for sediment sampling.

Figure 4.5 Photograph of the sampling procedure. 1: Sample in 2: collection pot; 3: platform to push the sediment core; 4: sampling core (in sediment); 5: sediment core after sampling; 6: shovel used to cut the top 1st cm of sediment; 7: 1cm thick ring; 8: slay used to bring the material back to shore.

Chlorophyll and pheophytin content

Chlorophyll and pheophytin pigments were extracted from lyophilised and homogenised sediments using acetone (8ml 90% acetone for 50mg sample). Extraction was carried out for 12h in the dark at 4°C, shaking. After centrifugation at 2000rmp for 10min, top aqueous phase was collected and the absorbance at 665nm was measured before and after acidification (80µl HCL 1M) of the samples. The concentration of chlorophyll and pheophytin were the calculated as:

$$[\text{Chlorophyll}]\ (\mu g/g\ \text{sample}) = \frac{F - Fa}{K - Ka} \times \frac{V}{m}$$

$$[\text{Pheophytin}]\ (\mu g/g\ \text{sample}) = \frac{(K \times Fa) - (Ka \times F)}{(K \times Ka) - (Ka \times Ka)} \times \frac{V}{m}$$

With V the volume of acetone used for extraction of a mass m of dried sample. F and Fa the absorbance read before and after acidification respectively; K and Ka the calibration values of the fluorometer measured before and after acidification respectively.

Potential Nitrification Rates

Potential Nitrification Rates (PNR) were measured using 5g of fresh sediments incubated in 30ml of artificial seawater (in $g.L^{-1}$: Nacl-24.6; KCl-0.67; $CaCl_2.2H_2O$-1.36; $MgSO_4.7H_2O$-6.29; $MgCl_2.6H_2O$-4.66; NaH_2CO_3-0.18→ 37.76ppm salts) supplemented with 24µM NaN_3 and 0µM or 250µM of $(NH_4)_2SO_4$ in 100ml glass bottles. After 24h in the dark at room temperature (≈20°C), shaking 100rpm, the incubation was stopped by addition of 30ml KCl (2M). After another 1h shaking (100rpm), the samples were filtered with Whatman paper and the concentration of nitrite was measured using an autoanalyzer (see *Nutrient analysis*). The accumulation of nitrite was determined by subtraction the initial nitrite concentration (see *Nutrient analysis*). Effect of salinity was tested by also incubating the sediment from day 1 in artificial seawater with reduced salts concentration (22.6ppm)

Flow cytometry

Prokaryote enumeration in the sediment samples was carried out using a 2-extractions protocol (Lavergne *et al.*, 2014).

1st extraction: Samples fixed in glutaraldehyde (see *Sampling and field measurement*) were diluted 1/1000 in a solution of sodium pyrophosphate (0.01M) + Tween 80 (0.1%), resulting in a final dilution of 1/2000 of the original sediment. The 1/2000 dilution were incubated at 4°C in the dark shaking. A separation step was then performed by sonication (60W for 30sec) and a 500µl aliquot of the sample was stained with 10µl of a solution of SYBR (1:100). Samples were then incubated for 15min in the dark before flow cytometry analysis.

2nd extraction: The remaining sample was centrifuged at 100rpm for 1min at 4°C. The pellet was resuspended in a volume of Tween pyrophosphate equal to the volume of sample before centrifugation, incubated 30min in the dark at 4°C shaking and sonicated (60W for 30sec). A 200µl aliquot of the sample was stained with 4µl of a solution of SYBR (1:100) and incubated for 15min in the dark before flow cytometry analysis.

4.2.3 DNA/RNA extraction

All surfaces and equipment were cleaned with 70% ethanol and RNase Zap (Ambion) before sample processing. Consumables used, including tubes and pipette tips were RNase free. DNA and RNA extraction were carried out on ≈2g frozen samples using the RNeasy PowerSoil RNA kit together with the RNeasy PowerSoil DNA elution kit (Qiagen). DNA/RNA preparations were stored at -80°C if not used immediately. The RNA extractions were further DNase treated using the Turbo DNase Kit (Ambion) to ensure the absence of DNA carryover. To do so, an extended protocol was used: half the recommended DNase volume is added to the sample and incubated for 30 min at 37°C, after which the second half of DNase is added, and the sample is re-incubated at 37°C for 1 hour. Success of the DNase treatment was checked by no PCR amplification of the V1-V3 Bacterial *16S rRNA* gene (Smith *et al.*, 2006)

4.2.4 Nucleic acid extraction quality check

Nucleic acids quantification: The quantity of extracted environmental nucleic acids was determined as follows: Total RNA was quantified using the Bioanalyser 2100 RNA Nano kit (Agilent Technologies); Total DNA was quantified using the Qubit High Sensitivity Kit (Life Technologies).

Nucleic acids quality check: RNA integrity was determined using two different approaches: microfluidics (RIN: Bioanalyser 2100 RNA Nano; Agilent Technologies) according to the manufacturer's instructions and the R_{amp} differential amplicons approach (Chapter II; Cholet *et al.*, 2019). DNA and RNA purity were evaluated with a NanoDrop spectrophotometer (Life Technologies).

4.2.5 Reverse-Transcription

DNA-free RNA was used for cDNA synthesis using Superscript IV kit (Invitrogen) and gene specific priming. The initial RT mixture containing 6µl water, 1µl reverse primer (10µM), 1µl dNTP's (10mM each) and 5µl template was incubated at 65°C for 5 min and quickly transferred on ice for 1 min. A second mix composed of 4 µl 5X first-strand buffer, 1 µl 0.1 mM dithiothreitol (DTT), and 1µl SuperScript IV (200 units/µl) was added and the resulting mixture was incubated at 55°C for 10 min and then at 80°C for 10 min.

4.2.6 Quantitative PCR

Construction of standard curves

DNA standard curves were constructed by amplifying the gene of interest and cloning the amplicons using pGem-T-Easy Vector Systems (Promega). The resulting ligation was transformed into the competent cell *E.coli* JM109 by heat shock according to manufacturer's instruction. White colonies were then PCR screened using T7 (5'-TAATACGACTCACTATAGGG–3') forward and gene specific reverse primer. Colonies that gave positive results were re-amplified using M13F (5'-GTAAAACGACGGCCAGT-3') and M13R (5'-CAGGAAACAGCTATGAC-3') and the resulting PCR products were purified using the SureClean Plus DNA purification kit (Bioline) then quantified using Qubit DNA High Sensitivity (ThermoFisher Scientific). The corresponding number of DNA amplicons was determined using the EndMemo RNA copy number Calculator (http://endmemo.com/bio/dnacopynum.php). An eight points serial dilution (5^{10} → 5^{3} copies.μl^{-1}) was prepared by successive 1/5 dilutions for each gene and was used as standard curve for Q-PCR. To ensure the cloned DNA sequence was the correct target gene, inserts were sequenced by Sanger sequencing and the identity was confirmed by a BLAST search (Altschul *et al.*, 1990).

Q-PCR

Phylogenetic markers (*18S rRNA* and *16S rRNA*) were amplified from extracted DNA diluted 1/10 using iTaq Universal Probes Supermix (Bio-Rad) in a 20µl final volume composed of 10µl iTaq Universal Probes Supermix buffer, 1.8µl each primers (10µM), 0.4µl probe (10µM), 5µl PCR grade water and 1µl template.

Functional genes and transcripts were quantified from the DNA and cDNA preparations diluted 1/10, respectively by quantitative PCR using the QuantiTech SYBR Green kit (Qiagen) as detailed in Table 4.1.

4.2.7 Illumina MiSeq amplicon libraries preparation

Libraries for Illumina sequencing were prepared by PCR amplification with primers listed in Table 4.1 using the same protocol as detailed in section 2.2.8.

4.2.8 Processing of amplicon sequences

To avoid the construction of operational taxonomic units (OTUs) based on an arbitrary similarity threshold, the amplicon sequence variants (ASVs) method was employed to identify single nucleotide differences (Callahan *et al.*, 2017) in order to construct abundance tables. To do so, the DADA2 pipeline (Callahan *et al.*, 2016a) was used on Qiime2, and the full pipeline can be found at https://github.com/umerijaz/tutorials/blob/master/qiime2_tutorial.md. Briefly, reads were demultiplexed using qiime demux emp-paired and denoised/quality trimmed using qiime dada2 denoise-paired. Forward reads were trimmed at 240bp and reverse reads at 200bp. ASVs were then constructed by merging the forward and reverse ASVs together and dereplicated to generate abundance files (ASVs counts in each samples) generated. To find the phylogenetic distances between ASVs, they were aligned using Maftt (Katoh *et al.*, 2009) and the phylogenetic tree was constructed using FastTree (Price *et al.*, 2010).

For the Bacterial *amoA* and *nxrB* and *nirK*, datasets forward and reverse reads did not merge properly when using the previous method. Therefore, demultiplexed reads were processed in R using R's DADA2 package (Callahan *et al.*, 2016b). First, quality trimming was done using the filterAndTrim function. Forward reads were trimmed at 275bp and 230bp, 213bp and 215bp for the reverse reads of *amoA* and *nxrB* and *nirK* respectively, allowing for a maximum of 2 errors in the forward and reverse reads to merge. Error models were generated using the learnErrors function. Reads were dereplicated using the derepFastq function and ASVs were inferred using the dada function. Forward and reverse reads were merged using mergePairs function allowing for a minimum overlap of 10bp. A sequence table was generated using the makeSequenceTable function and chimeras were removed using the removeBimeraDenovo function. A count table was then generated and distances between the representative ASVs were inferred by aligning the sequences using Maftt (Katoh *et al.*, 2009) and constructing a phylogenetic tree using FastTree (Price *et al.*, 2010).

For AOA and AOB *amoA* amplicon sequences, further steps were undertaken to ensure the reliability of the data. First, sequences were filtered based on their length in R using the seqinr package (Charif and Lobry, 2007). For AOA and AOB, sequences that were not 256bp and 491bp respectively were deleted from the data set. The sequences that passed this step were then translated to protein using MEGA7 (Kumar *et al.*, 2016). Resulting sequences were checked using BLASTp to ensure they were translated in the correct dataframe. ASVs that translated to the same protein sequence were considered the same and their abundances were

added together. Finally, a phylogenetic tree was constructed using only those ASVs with the correct length that translated to unique proteins: the sequences were aligned using Maft and the phylogenetic tree drawn using FastTree. The tree was visualised and coloured in R using the ggtree package (Yu *et al.*, 2017).

Taxonomic assignation

To assign the taxonomy to the representative ASVs, two different approaches were used: first, representative ASVs were taxonomically classified using qiime feature-classifier classify-sklearn against the reference databases constructed as follows: for the functional genes (Bacterial *amoA*, Archaeal *amoA*, *nxrB*, *nirS*, *nirK* and *nrfA*), the sequences available on the Fungene database (http://fungene.cme.msu.edu/) were downloaded as fasta files. The annotations from the fasta headers were used to get the taxonomic details for each sequence using R's rentrez package (Winter, 2017) and used to generate a taxonomy file. The FASTA file and the corresponding taxonomy file were then formatted to work with Qiime. For *16S rRNA* and *18S rRNA*, the SILVA SSU Ref NR database release v123 was used. Biom files were then generated (containing the abundance and taxonomy tables) to be used in R with the plyloseq package (McMurdie and Holmes, 2013).

For the functional genes, the Bayesian Lowest Common Ancestor (BLCA) approach was also used (Gao *et al.*, 2017) as described here https://github.com/qunfengdong/BLCA. First, curated databases were constructed, including only the sequences from Fungene for which the taxonomy was known to species level. For AOA and AOB, additional sequences from previously defined marine sediment clusters (Duff *et al.*, 2017; Zhang *et al.*, 2018) were also added. The names of these clusters were used instead of the species name in the reference taxonomy. BLAST formatted databases were then generated using the makeblastdb function (python3) and the curated databases as input. The output of the makeblastdb function was formatted to work with the BCLA software using a custom awk script. The BCLA software was then ran using the 2.blca_main.py function, with the representative ASVs and the databases (reference taxonomy and sequences) as input. The output taxonomy file was then used with the abundance tables to generate the biom files.

4.2.9 Statistical analysis

To test if the conditions in ridges and runnels were significant drivers of the communities retrieved from the sequencing of functional and marker genes, a canonical correspondence

analysis (CCA) was carried out in R. First, the abundance tables (*i.e.* the ASVs counts for each targets in each samples) were normalized using the Hellinger transformation (Legendre and Gallagher, 2001). Secondly, the parameter table (*i.e.* the table containing the bio-physico-chemical parameters for each samples) was normalized by centring and reduction. The CCA was then computed using the cca function from R's vegan package (Oksanen *et al.*, 2005) with the standardized parameter table as the explanatory table and the Hellinger-transformed ASV abundance table as the response table. Variable selection was then carried out using the ordistep function (vegan package) with the option direction="both", allowing for a simultaneous backward and forward selection to find significant drivers for each target genes.

Table 4.1 List of primers used in this study.

Primer	Sequence (5'→3')	Orientation	Target (length)	Experiment condition	Q-PCR			Reference (for primer sequence)
					Slope/ Efficiency	Intercept	R^2	
BacamoA-1F	GGGGHTTYTACTGGTGGT	Forward	Bacterial *amoA* gene (491bp)	**PCR:** 95°C-15min; [94°C-30sec; 47°C(DNA) or 55°C(cDNA) -30sec; 72°C-30sec]x32; 72°C-10min ----- **Q-PCR:** 95°C-15 min; [95°C-30sec; 47°C(DNA) or 55°C(cDNA)-30 sec; 72°C-1min; 79.5°C-2sec → plate read]x40; melt curve 65°→95° 0.5° increment/ 5sec	cDNA: -3.94/ 79.39% DNA: -4.07/ 76.08%	cDNA: 41.02 DNA: 41.8	cDNA: 0.995 DNA: 0.999	(Hornek *et al.*, 2006)
BacamoA-2R	CCCCTCBGSAAAVCCTTCTTC	Reverse						
Arch-amoWAF	CTGAYTGGGCYTGGACATC	Forward	Archaeal *amoA* gene (256bp)	**PCR:** 95°C-15min; [95°C-30sec; 58°C-40sec, 72°C-1min]x35; 72°C-10min ----- **Q-PCR:** 95°C-15min; [95°C-30sec; 58°C-40sec, 72°C-10sec → plate read]x40; melt curve 65°→95° 0.5° increment/ 5sec	-4.46/ 67.58%	43.97s	0.999	(Wuchter *et al.*, 2006)
Arch-amoWAR	TTCTTCTTTGTTGCCCAGTA	Reverse						
nxrBqF	TGTGGTGGAACAACGTGGAA	Forward	*Nitrospira nxrB* (180bp)	**Q-PCR:** 95°C-15min; [95°C-20sec; 56°C-20sec; 72°C-30sec → plate read]x40; melt curve 65°→95° 0.5° increment/ 5sec	cDNA: -3.9/ 80.47% DNA: -3.74/ 85.09%	cDNA: 38.16 DNA: 37.72	cDNA: 0.996 DNA: 0.999	(Feng *et al.*, 2016)
nxrbqR	CCCGGCATCGAAAATGGTCA	Reverse						
515F	GTGYCAGCMGCCGCGGTAA	Forward	Bacterial/Archeal *16S rRNA* (V4-V5; 411bp)	**PCR:** 95°C-15min; [94°C-45sec, 50°C-30sec, 72°C-40sec]x25; 72°C-10min		NA		(Walters *et al.*, 2015)
926R	CCGYCAATTYMTTTRAGTTT	Reverse						
1369F	CGGTGAATACGTTCYCGG	Forward	Bacterial/Archeal *16S rRNA* (V9; 123bp)	**Q-PCR:** 95°C-10min; [95°C-10sec, 60°C-30sec]x40; 40°C-10min	-3.25/ 103.09%	39.87	0.989	(Suzuki *et al.*, 2000)
1492R	GGWTACCTTGTTACGACTT	Reverse						
1389P	CTTGTACACACCGCCCGTC	Probe						

4.3 Results

4.3.1 Bio-physico-chemical parameters

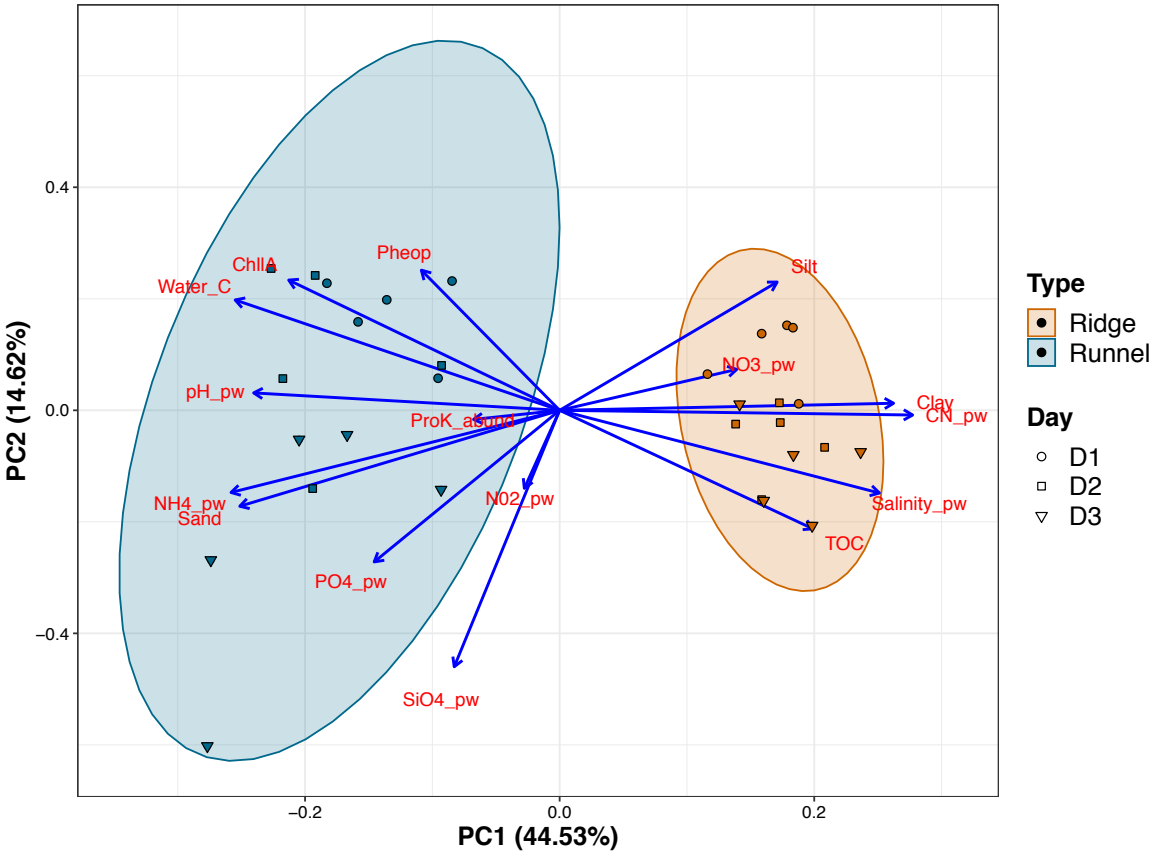


Figure 4.6 Principal Component Analysis of the physio-chemical and biological parameters measured in the ridge/runnels structures. "pw": parameter measured on pore water; Water.C: water content; TOC: Total Organic Carbon; ChlA: Chlorophyll A; Pheop: Pheophytin; CN.pw: Total Carbon/ Total Nitrogen ratio in pore water. ProK_abund: Prokarote abundance measured by flow cytometry.

Conditions in ridges and runnels were strongly different, based on the bio-physico-chemical parameters measured in this study (*p*.value PERMANOVA<0.001) as observed by the clustering of the two different sediment structures (Figure 4.6). Indeed, 44.5% of the variance displayed along PC1, separated ridge and runnel samples. A further 14.6% of the variance displayed along PC2, separated samples from Day1 (more positive values along PC2) and sample from Day2 and 3 (more negative values along PC2) in ridges whereas in runnels, samples from Day1 and 2 were more similar between each other and samples from day 3 more scattered. The effect of the sampling day on the parameters measured was also significant (*p*.value PERMANOVA<0.01). Overall, sediment in ridges was finer (more clay, less sand), dryer (lower water content) and slightly more acidic. Salinity was found higher in ridges, which could be linked to more evaporation of the seawater explaining the lower water content as well. Ridges sediment was less rich in nutrient, with lower ammonia (≈+5fold in runnels) and

phosphate (≈+2fold in runnels). Only nitrate was found higher in ridges. Nitrite concentrations were also determined but no statistical differences were observed between ridges and runnels (≈0.07µM in average in both). Results on pore water were in accordance with measures of the sediment, with an observed higher CN ratio in pore water in ridges than in runnels. Interestingly, photosynthetic pigments (chlorophyll A and pheophytin) were found in higher concentration in runnels (Table 4.2).

Table 4.2 Differences in bio-physico-chemical parameters between Ridges and Runnels

Parameter	Mean Ridges (sd)	Mean Runnels (sd)	Type	Day
Salinity.pw (‰)	36.47 (1.12)	33.81 (0.66)	***	***
pH.pw	7.02 (0.08)	7.21 (0.06)	***	***
Nitrate.pw (µM)	1.56 (1.2)	0.7 (0.18)	**	*
Nitrite.pw (µM)	0.07 (0.03)	0.07 (0.02)	-	-
Ammonia.pw (µM)	44.1 (15.49)	203.89 (85.87)	***	-
Phosphate.pw (µM)	0.07 (0.05)	0.13 (0.09)	**	***
Silicate.pw (µM)	163.27 (21.36)	176.19 (37.37)	-	***
Water.C (%)	43.36 (1.64)	50.62 (2.23)	***	-
TOC (%)	14.79 (2.62)	10.89 (2.43)	***	-
ChllA (µg/g)	10.32 (3.12)	20.13 (6.24)	***	-
Pheop (µg/g)	15.39 (1.6)	20.72 (9.3)	**	***
ProK abundance	$6.6x10^6$ ($8.5x10^5$)	$7.0x10^6$ ($1.3x10^6$)	-	-
CN.pw	5.9 (0.53)	2.96 (0.96)	***	-
Sand (%)	7.82 (1.03)	14.51 (3.95)	***	-
Silt (%)	81.14 (1.76)	78.22 (3.67)	*	-
Clay (%)	11.04 (1.19)	7.27 (0.81)	***	-

Mean values (standard deviation) are presented. Effect of sediment type (Effect.Type) and sampling Day (Effect.Day) are reported. -: No Statistical Differences; .: 0.05< *p*.value<0.1; *: *p*.value<0.05; **: *p*.value<0.01; ***: *p*.value<0.001. "pw": parameter measured on pore water; Water.C: water content; TOC: Total Organic Carbon; ChlA: Chlorophyll A; Pheop: Pheophytin; CN.pw: Total Carbon/ Total Nitrogen ration in pore water.

4.3.2 Potential Nitrification Rates

PNR was found to be significantly higher in runnels than ridges for all three sampling days, with PNR being ≈22, ≈40 and ≈17 fold higher in runnels at day 1, 2 and 3, respectively. These differences were still significant when ammonia was added prior to the incubation but the observed differences were smaller (≈+1.8fold and ≈+2.5fold in runnel VS ridges at day 1 and day 2 respectively), and actually became non-significant for the incubation at day 3. Relative to the PNR measured at *in situ* ammonia concentration (+0µM ammonia), the addition of ammonia induced a stronger response in ridges (+15.7fold day1; +26.8fold day2; +15.1fold day3) than in runnels (+1.3fold day1; +1.7fold day2; no difference for day3). However, even

when comparing the PNR from runnels incubated at *in situ* ammonia concentration with the ridges incubated with 500µM ammonia, the PNR in runnels was still higher (Figure 4.7).

Changes in salinity had little effect on the PNR (Figure 4.8) The only significant effect was detected for the ridge sediment without addition of ammonia, with an decrease of 1.7fold in PNR when salinity was increased from 22.6‰ to 37.76‰ (*p*.value<0.05).

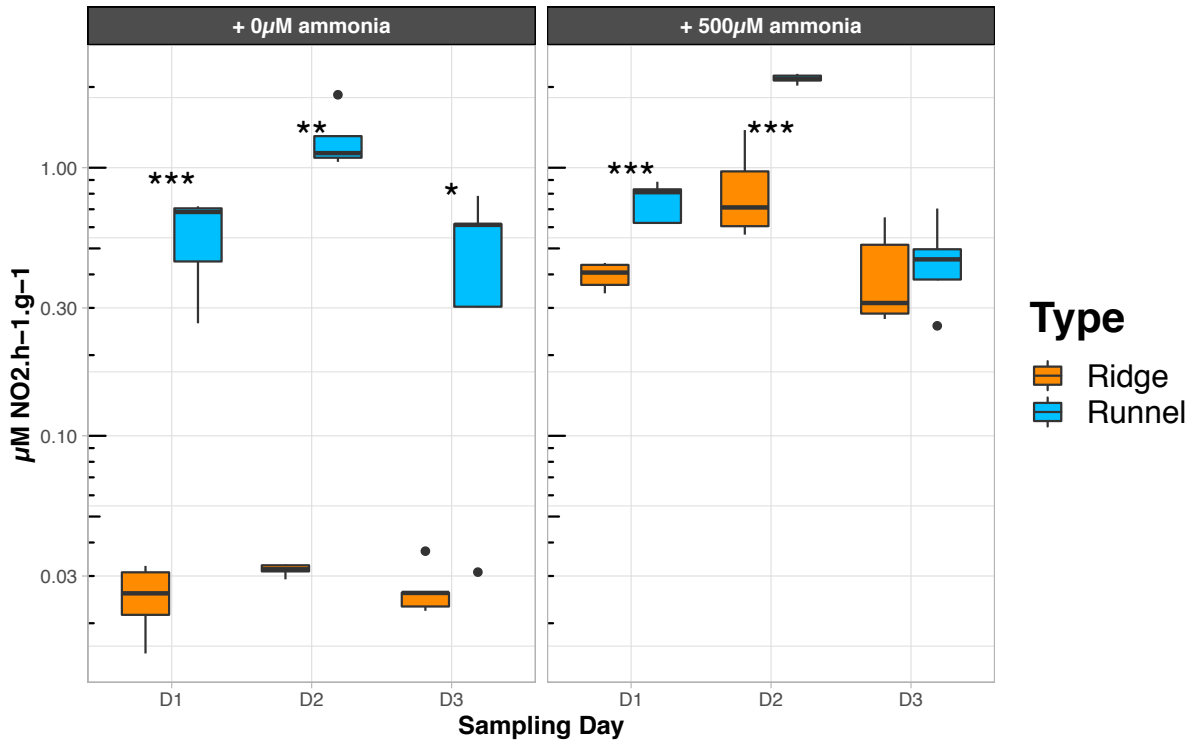


Figure 4.7 Potential Nitrification Rates (PNR) in ridges and runnel. PNR was calculated by the accumulation of nitrite, with (right) and without (left) addition of 500µM ammonia in the sediment. D1, D2 and D3 refer to the sampling Day 1, Day 2 and Day 3 respectively. Results are reported in nmol nitrite per g sample per day. *: *p*.value<0.05; **: *p*.value<0.01; ***: *p*.value<0.001.

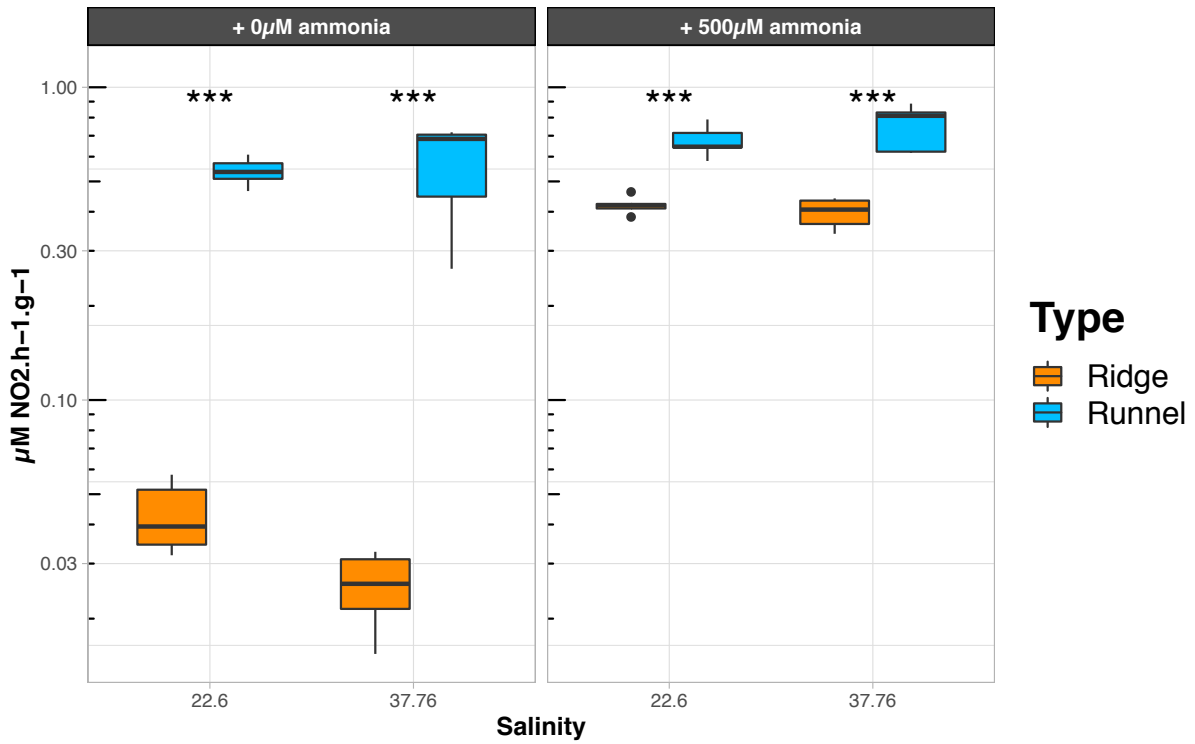


Figure 4.8 Effect of Salinity on PNR in ridges and runnels. PNR was calculated by the accumulation of nitrite, with (right) and without (left) addition of 500µM ammonia in the sediment from day 1 incubated at two different salinities (22.6 ‰ and 37.76 ‰).

4.3.3 RNA integrity check

Table 4.3 RNA integrity check.

Sample:	**D1Ba**	**D1Bb**	**D1Bc**	**D1Bd**	**D1Be**	**D1Sa**	**D1Sb**	**D1Sc**	**D1Sd**	**D1Se**
RIN	7	6.4	6.9	ND	6.8	7.1	4.3	6.4	6.4	NA
R_{amp}	0.72	0.66	0.74	ND	0.77	0.60	0.56	0.64	0.65	0.64
Sample:	**D2Ba**	**D2Bb**	**D2Bc**	**D2Bd**	**D2Be**	**D2Sa**	**D2Sb**	**D2Sc**	**D2Sd**	**D2Se**
RIN	NA	8.1	7.8	7.3	7.3	6.7	7	6.7	7.2	7.1
R_{amp}	0.64	0.69	0.75	0.72	0.62	0.64	0.69	0.77	0.78	0.79
Sample:	**D3Ba**	**D3Bb**	**D3Bc**	**D3Bd**	**D3Be**	**D3Sa**	**D3Sb**	**D3Sc**	**D3Sd**	**D3Se**
RIN	7.3	7	7.2	7	7.8	7.2	8	7.3	7.1	NA
R_{amp}	0.75	0.81	0.75	0.78	0.73	0.65	0.63	0.64	0.82	0.80

RIN: RNA Integrity Number (Bioanalyser; Agilent Technologies); R_{amp}: Ratio Amplicon (Cholet *et al.*, 2019). NA: No value generated; ND: Not determined

Except for D1Bd and D1Sb, all samples showed good RNA integrity with RIN values ranging from 6.4 to 8.1 and Ramp values ranging from 0.63 to 0.81 (Table 4.3). Sample D1Bd was lost during processing and therefore no integrity values were calculated. Sample D1Sb showed

lower integrity values for both the RIN and R_{amp} indexes and was therefore not used for RNA analysis.

4.3.4 Differences in gene and transcript abundance

16S rRNA: A high number of 16S rRNA genes was detected, sometimes up to 10^{10} copies/g samples, with overall, a higher number found in runnels than in ridges. This difference was however not statistically significant (*p*.value = 0.052). The abundance at day 3 was lower compared to day 1 and 2 for both ridges and runnels (≈5fold lower) and indeed, the effect of the sampling day on the abundance of 16S rRNA genes was significant (*p*.value>0.001) and the abundance in *16S rRNA* genes in the samples from ridges day 3 was statistically different from all other samples except for runnel day 3 (Figure 4.9; Table 4.4).

Archaeal *amoA*: The number of AOA *amoA* genes copies was higher in ridges than in runnels at day 1 and 2 but the differences were not statistically significant. For AOA *amoA* transcripts, the detected signal on Q-PCR was too weak to be clearly distinguished from the no template control and the results are therefore not reported (Figure 4.9; Table 4.4).

Bacterial *amoA*: The effect of the sediment type on AOB *amoA* gene copy number was important (*p*.value<0.001) with AOB *amoA* genes more abundant in ridges overall. When considering day-by-day comparison, significant differences were only observed at day 2 (*p*.value<0.01), but the differences at Day 1 were still important (*p*.value=0.06). The ridge samples from day 2 were the ones with the highest copy number of AOB *amoA* genes (statistical differences with all runnel samples; Table 4.4). Results for AOB *amoA* transcripts Q-PCR strongly contrasted with the DNA ones: the effect of both the sediment type and sampling day were significant (*p*.value<0.05) but the AOB *amoA* transcripts were more abundant in runnels than in ridges, except for day 3. Differences between the transcript copy number in the runnels at day 1 and all other samples were significant (Figure 4.9; Table 4.4).
AOB/AOA ratio (DNA): In all samples, AOB were dominant over AOA but the ratio of AOB/AOA gene copy number varied between ridges and runnels: ≈50 in ridges at day 1 and 2 (*i.e.* 50 times more AOB *amoA* gene than AOA *amoA*) and ≈40 at day 3. The ratio decreased to ≈30 in runnels at all sampling days and these differences were statistically significant (*p*.value<0.001) (Figure 4.10).

nxrB: A similar trend as seen for AOB *amoA* was observed, with the gene copy numbers higher in ridges whereas transcripts were more abundant in the runnels. Statistically significant differences were observed only between the gene copy number in the ridges at day 2 and the runnel samples (Figure 4.9; Table 4.4).

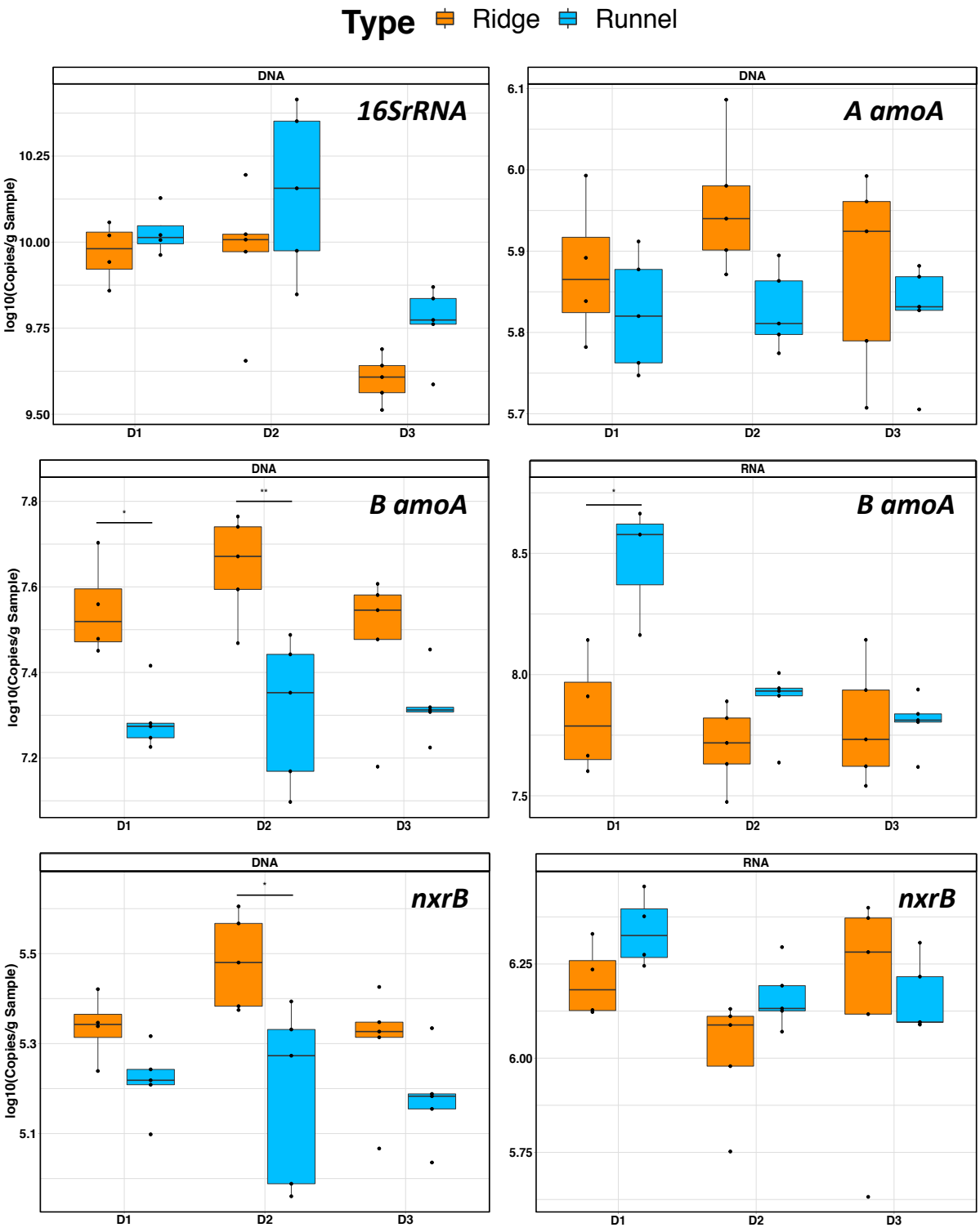


Figure 4.9 Differences in gene and transcript abundance in ridges and runnels. The target gene (DNA)/transcript (RNA) is indicated on top of each plot. Only statistical differences between sediment types from the same sampling day are reported. .: 0.05< *p*.value<0.1; *: *p*.value<0.05; **: *p*.value<0.01; ***: *p*.value<0.001.

Table 4.4 Summary of the TukeyHSD multiple comparison tests for Q-PCR quantification of target genes and transcripts.

	Comparison	*p*.value
16S rRNA	D1 ridge - D3 ridge	*
	D1 runnel - D3 ridge	**
	D3 ridge - D2 ridge	**
	D3 ridge - D2 runnel	***
	D3 runnel - D2 runnel	**
AOB *amoA* DNA	D2 ridge - D1 runnel	**
	D2 ridge - D2 runnel	**
	D3 runnel - D2 ridge	**
AOB *amoA* RNA	D1 ridge - D1 runnel	**
	D1 runnel - D2 ridge	***
	D1 runnel - D2 runnel	**
	D1 runnel - D3 ridge	**
	D1 runnel - D3 runnel	**
nxrB DNA	D2 ridge - D1 runnel	*
	D2 ridge - D2 runnel	*
	D3 runnel - D2 ridge	*

Only the comparisons for which *p*.value<0.05 are shown. *: *p*.value<0.05; **: *p*.value<0.01; ***.

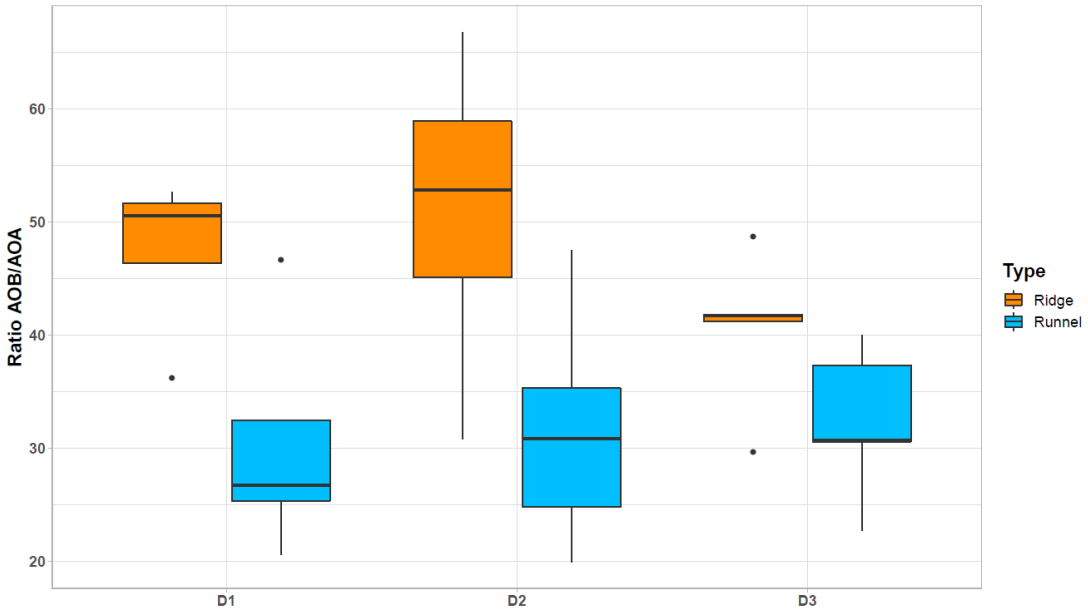


Figure 4.10 Ratio of AOB/AOA amoA gene copy number in ridges and runnels.

4.3.5 Differences in transcription ratio

For AOB *amoA* and *nxrB*, the transcript ratios were calculated as the ratio between the gene and transcript copy numbers. For both genes, the ratios were always >1 and sometimes up to ≈25 (i.e. 25 transcripts per gene copy), indicating active transcription of bacterial nitrification genes in these samples. Overall, for both AOB *amoA* and *nxrB*, the transcription ratios were higher in runnels than in ridges, and indeed, the effect of the sediment type was significant (*p*.value<0.001 for AOB *amoA* and *p*.value<0.01 for *nxrB*). When comparing ridges and runnels day by day, significant differences were observed only at day 1 for AOB *amoA*. The lowest differences between ridges and runnels were observed for day 3 (Figure 4.11).

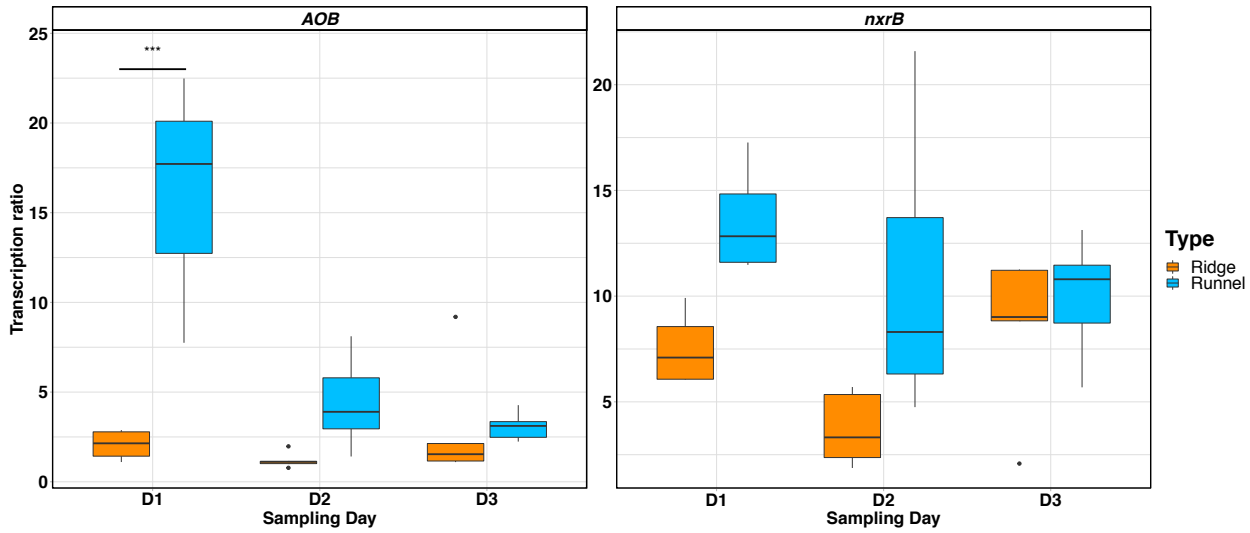


Figure 4.11 Transcription ratio of the Bacterial *amoA* (top) and *nxrB* (bottom) transcripts in ridges and runnels. The ratios were calculated as copy number mRNA/ copy number DNA. .: 0.05< *p*.value<0.1; *: *p*.value<0.05; **: *p*.value<0.01; ***: *p*.value<0.001.

4.3.6 Potential nitrification rates per AOB cells

The number of AOB cells per gram of sample was estimated by dividing the *amoA* gene abundance (per gram of sample) by 2.5 (the average number of *amoA* gene per AOB cells). PNR values reported in Figure 4.7 were then divided by the estimated number of cells per gram of sediment to calculate the PNR per AOB cells. The estimated PNR per AOB cells was much higher in runnels than in ridges when no ammonia was added prior to incubation (40, 105 and 20 fold at day 1,2 and 3, respectively). When ammonia was added, the PNR per AOB cells increased slightly in runnels (1.3 and 1.5 fold at day 1 and 2, respectively; no change at day 3) and increased more in ridges (17, 26 and 15 fold at day 1,2 and 3, respectively). However, the values in runnels were still higher than in ridges, except at day 3. Interestingly, the PNR per

cells in ridges with ammonia added was still lower than in runnels without ammonia added (Figure 4.12).

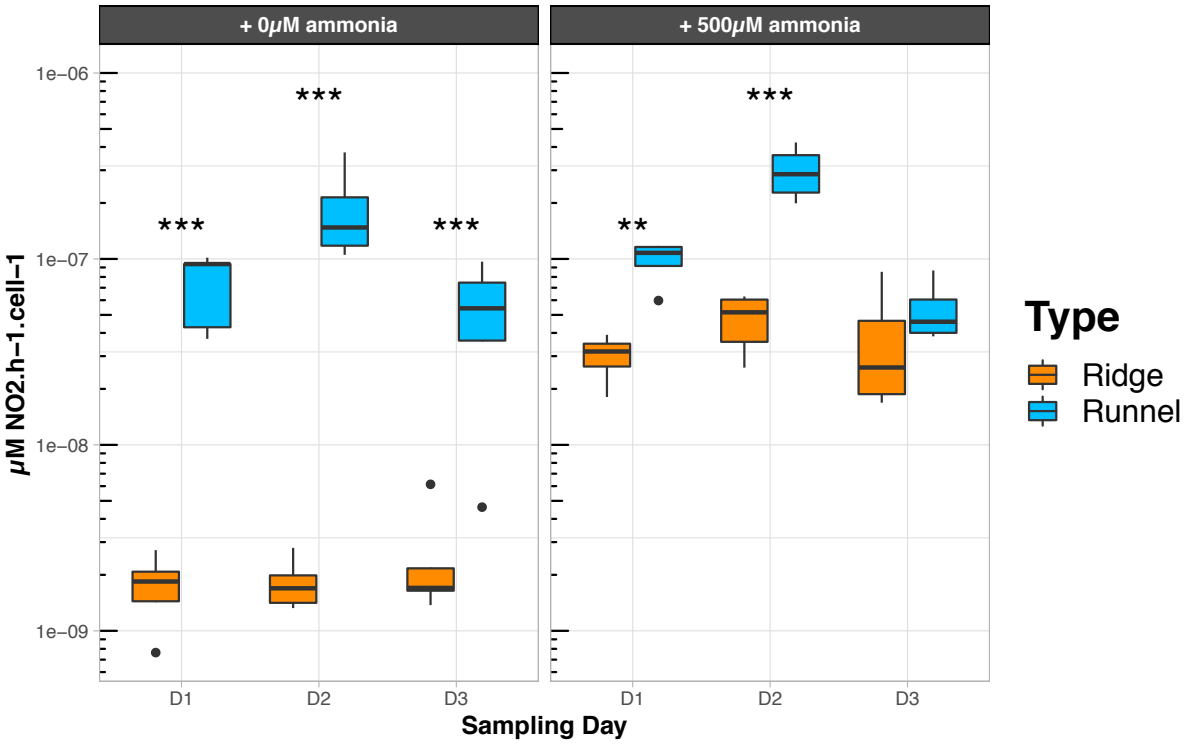


Figure 4.12 PNR per AOB cells in ridges and runnels. PNR values calculated previously were divided by the estimated number of AOB cells per gram of sediments. Only statistical differences between sediment types from the same sampling day are reported. .: 0.05< *p*.value<0.1; *: *p*.value<0.05; **: *p*.value<0.01; ***: *p*.value<0.001.

4.3.7 Differences in microbial communities between ridges and runnels

The effect of the sediment type (ridge/runnel) and sampling day were tested on the overall prokaryote structure (*16S rRNA*) and the ammonia oxidizer communities (Bacterial and Archaeal *amoA*) by two-ways PERMANOVA test on the Bray-Curtis dissimilatory and Unifrac/WUnifrac distance matrixes. To determine which bio-physiochemical parameters were significant drivers of the communities, a canonical correspondence analysis (CCA) was carried out using the ASVs abundance table as response table and the bio-physiochemical table as explanatory table. Variable selection was done using a combined forward and backward selection.

As shown in Figure 4.13 the sediment type significantly impacted the overall microbial community (p.value PERMANOVA<0.01). Clustering between ridges and runnels could be observed on the PCoA plots for all three β diversity indexes tested. When considering WUnifrac distances, the sediment type explained 34% of the total variance indicating a significant differentiation of the microbial communities between sediment types. On the other hand, there was no significant effect of the sampling day on the microbial communities recovered, which accounted for only 6% to 8% of the total variance in the β diversity distances matrices. Result from the CCA indicated that the *16S rRNA* ASV composition was significantly driven by pH, TOC (in the pore water) and the sediment grain size (% of silt+clay) (Table 4.5). Next, the effect of the sediment type and sampling day was tested on the overall prokaryote structure after grouping the ASVs abundances at relevant taxonomic level (Family, Order, Class). At all taxonomic level tested, the sediment structure had a strong effect on the community structure (Bray-Curtis) and explained 38%, 43% and 50% of the total variance in microbial Families, Orders and Classes respectively between samples. Again, the sampling day did not have a significant impact. In both ridges and runnels, the microbial community was dominated by the Proteobacteria phylum, with gamma-, delta- and alpha- proteobacteria being the first, third and fourth most abundant Classes, respectively, in both ridges and runnels; Bacteroidia was the second most abundant Class. Interestingly, families Nitrosomonadaceae, Nitrospiraceae and Nitrosococcaceae were detected among the top most abundant 30 families (Figure 4.14).

Table 4.5 Drivers of the microbial communities at DNA level as determined by variable selection of CCA models. For all targets, the p.value of the variables retained and the overall model are reported; **.**: 0.05< *p*.value<0.1; *: *p*.value<0.05; **: *p*.value<0.01; ***: *p*.value<0.001. Numbers in italic below the variables indicate their VIF. SGS: sediment grain size (% silt+clay)

	pH	NO_3^-	NO_2^-	NH_4^+	PO_4^{2-}	TOC	SGS				**Overall**
	2.38	*1.41*	*1.24*	*2.72*	*1.76*	*1.34*	*2.75*				
16S rRNA	**					*	**				***
	pH	NO_3^-	NO_2^-	NH_4^+	PO_4^{2-}	TOC	SGS	ChllA	Pheop	Cyano	
	3.73	*1.46*	*1.34*	*2.98*	*2.35*	*1.54*	*3.06*	*2.02*	*1.58*	*1.35*	
B *amoA*				**					*		***
A *amoA*	.			.			.				*

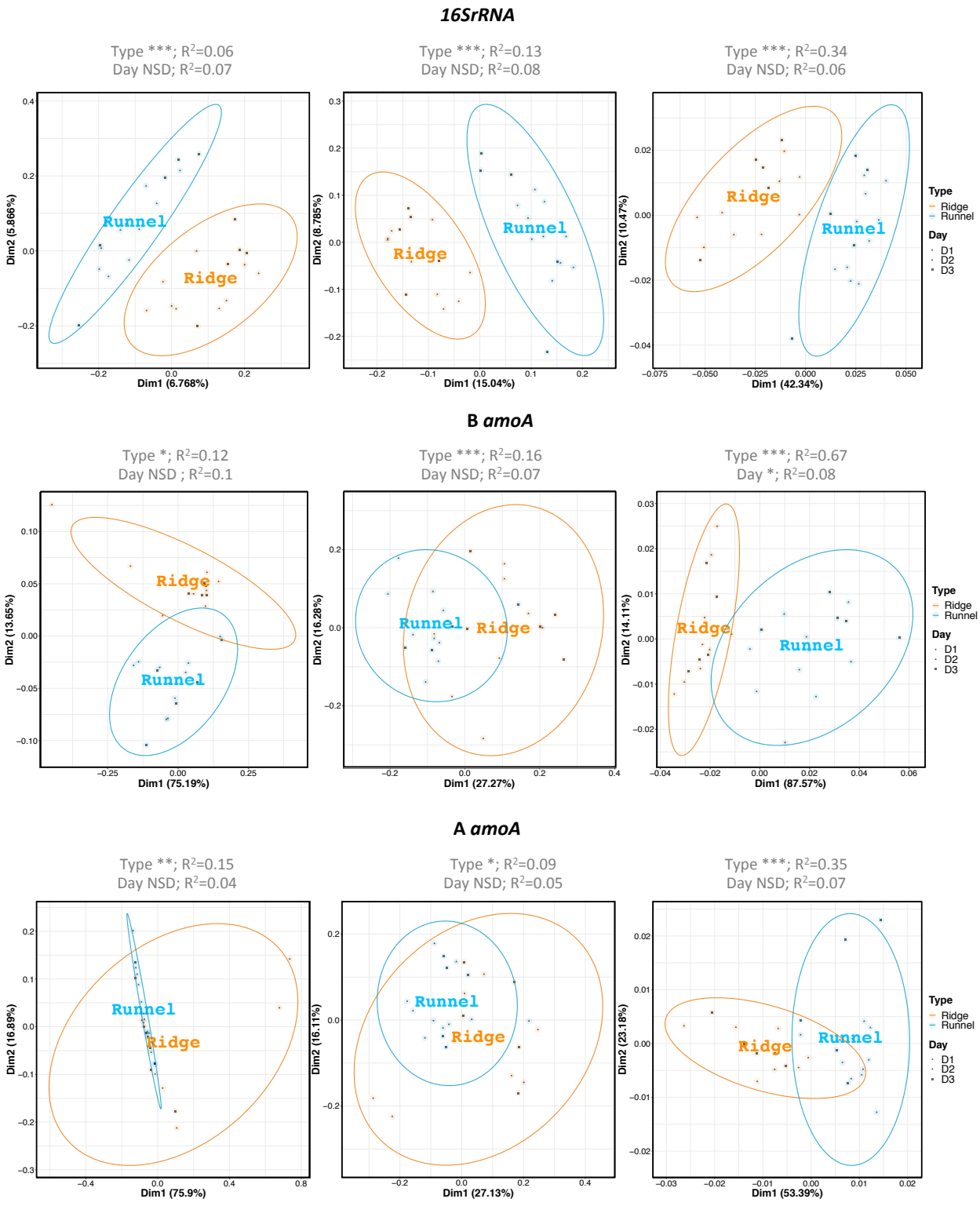


Figure 4.13 Principal Component Analysis (PCoA) showing the effect of the Ridge/Runnel structure on the overall microbial community and ammonia oxidizers at the ASV level. Result of the PERMANOVA test for all three beta diversity indexes is shown on top of each plot. NSD: No Significant Differences; *: *p*.value<0.05; **: *p*.value<0.01; ***: *p*.value<0.001. R^2: proportion of variance explained.

The sediment structure also had a significant effect on the ammonia oxidiser communities, for all three β diversity indexes measured (Figure 4.13). The effect of the sediment structure was always stronger when considering WUnifrac distances than Unifrac or Bray-Curtis as shown by more significant *p*.values, higher R^2 values and clearer clustering on the PCoA plots. Interestingly, the sampling day had a significant effect on the AOB community (Bray-Curtis and WUnifrac) but not on the AOA community. Result from the CCA indicated that these differences in AOA and AOB communities were driven in part by the ammonia concentration, with a stronger effect on AOB than AOA. Interestingly, pH was found to be a driver for the AOA community but not for AOB. On the other hand, AOB seemed to be influenced by the phototrophs present in the samples, with the pheophytin content as drivers of the AOB community (Table 4.5).

Table 4.6 ASVs in the DNA libraries not shared between ridges and runnels.

	Total	**Unique ASVs Ridges**			**Unique ASVs Runnels**		
		Count	Abund Ridges	Abund Total	Count	Abund Runnels	Abund Total
16S rRNA	14408	5223	35%	16.1%	6266	37.2%	20.3%
AOB	112	21	1.8%	0.9%	33	3.1%	1.6%
AOA	107	32	28.1%	11%	34	16%	9.8%

For each target, the total number of ASVs detected in the DNA libraries (ridges and runnels) is reported (Total) along with the number of ASVs that are unique to ridges and runnels respectively. The proportions that those unique ASVs represent relative to the abundance in ridges or runnels and relative to the total community are also reported.

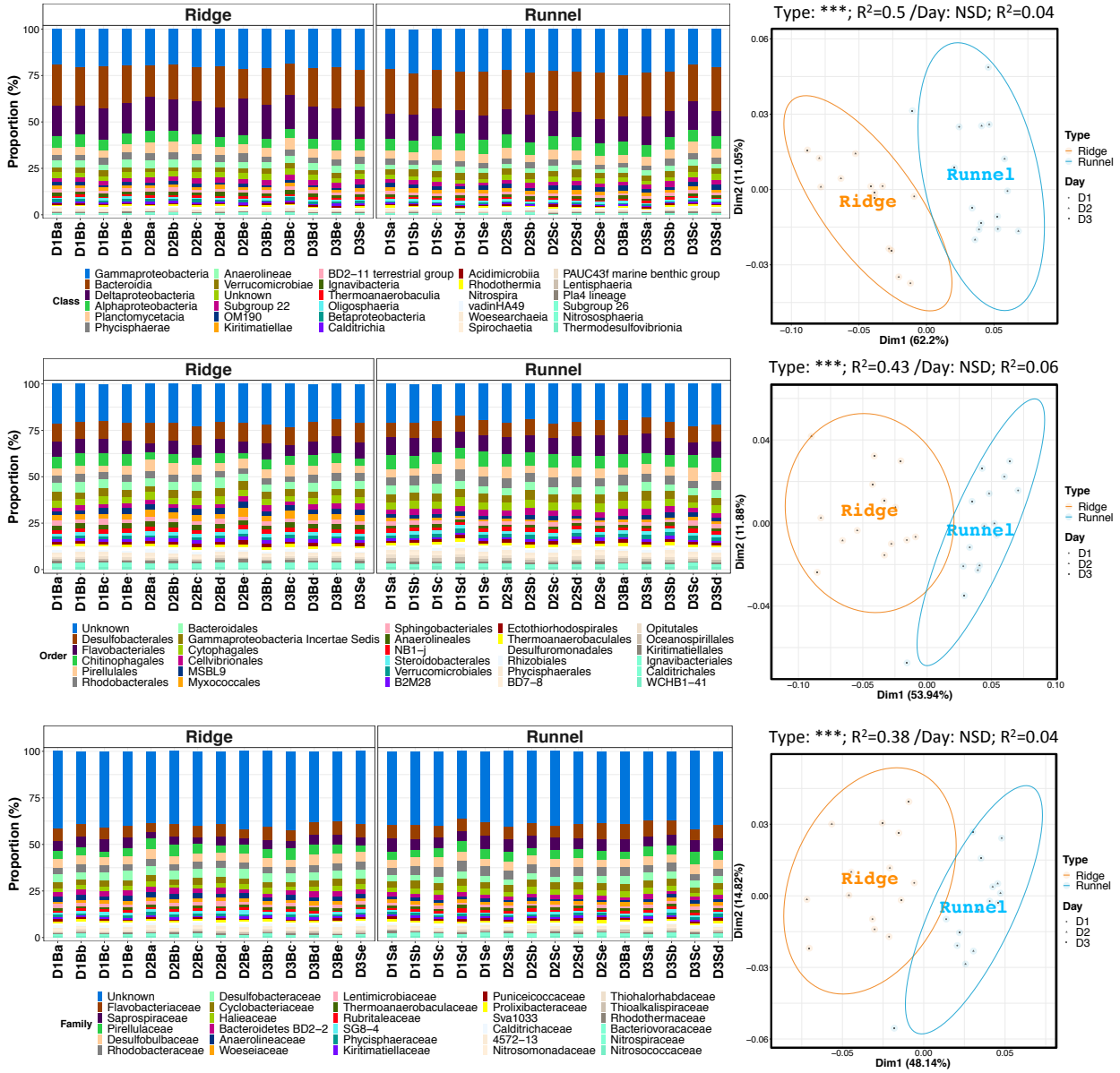


Figure 4.14 Taxabars of the microbial communities in ridges and runnels. The proportions are shown as the average over the three sampling days, at three different taxonomic levels (Family, Order and Class). The PCoA on the right of each taxabars shows the clustering between ridges and runnels. Results of the PERMANOVA test on the Bray-Curtis dissimilatory matrices calculated at each taxonomic level are shown on top of the plots. NSD: No Significant Differences; *: *p*.value<0.05; **: *p*.value<0.01; ***: *p*.value<0.001. R^2: proportion of variance explained.

4.3.8 Differences in total and active ammonia oxidizers in ridges and runnels

4.3.8.a AOB

Total AOB: To identify the AOB present in ridges and runnels, a phylogenetic tree of ASVs detected at DNA level was drawn and their relative abundances (log2 transformed) was calculated in both ridges and runnels (Figure 4.15). Three clusters could be distinguished on

the AOB phylogenetic tree: cluster C1, mainly containing ASVs that could be assigned to *Nitrosomonas* Cluster Group A and B and some unknown ASVs; cluster C2: mainly containing unknown ASVs one representative of *Nitrosomonas oligotropha*, *Nitrosomonas aestuarii* and *Nitrosomonas eutropha*; cluster C3: containing unknown ASVs and representative of *Nitrosospira*.

When considering abundances, both ridges and runnels were dominated by ASV_2 (identified as *Nitrosomonas* group A) representing 55.5% and 52.2% of the total abundance in ridges and runnels, respectively. The next most abundant member was ASV_1 (also identified as *Nitrosomonas* group A), representing 27.3% and 23% of the total abundance in ridges and runnels, respectively; both belonged to cluster C1. The next three most abundant ASVs (ASV_8, ASV_12 and ASV_18) represented a total of 12.3% and 12.9% in ridges and runnels, respectively and were also identified as *Nitrosomonas* group A and belonged to cluster C1. To summarize, the overall community present in ridges and runnels were similar and dominated by *Nitrosomonas* Cluster group A, with the majority of ASVs (in terms of number of individual ASVs and relative abundance) shared between ridges and runnels (Figure 4.15 and Table 4.6).

Active AOB: To estimate the activity of individual AOB members, the ratio of abundance between the RNA and DNA libraries was calculated for each individual ASV (a positive value for the RNA/DNA ration indicates high activity while a negative ratio indicates low activity). In both ridges and runnels, two clusters could be distinguished: a low activity cluster, containing members of the C1 cluster (as defined in Figure 4.15) and a high activity cluster, containing members of the C2 cluster. The activity (abundance in RNA library) of the highly active cluster generally correlated positively with PNR measured in individual structure (ridges or runnels). It is interesting to note that the ASVs dominant at DNA level (ASV_1, ASV_2, ASV_8, ASV_12 and ASV_18) were found in the low activity cluster in both ridges and runnels (Figure 4.16).

When comparing ridges and runnels, there were more highly active ASVs in runnels (53 ASVs) compared to ridges (18 ASVs) and all but 3 ASVs (ASV_220, ASV_242 and ASV_297) highly active in ridges were also active in runnels (Figure 4.16). Also, differences in activity of individual ASVs (*i.e.* differences in abundances at RNA level) revealed that out of the 40 ASVs whose activities were significantly different between ridges and runnels, 35 were more active

in runnels; 30 out of these 35 ASVs had activities that positively correlated with overall PNR (Figure 4.17).

As seen in Figure 4.18, 32 out of the 35 ASVs that were significantly more active in runnels compared to ridges were also identified as highly active ASVs (significantly more abundant in the RNA library compared to DNA). Interestingly, these ASVs represented a small proportion of the community at DNA level in both ridges (1.21%) and runnels (4.19%) and their taxonomy could not be assigned. A further 21 ASVs, identified as highly active, were not significantly more active in runnels and also represented a small proportion of the community at DNA level (0.06% in ridges and 0.19% runnels). Finally, 3 ASVs were found to be more active in runnels compared to ridges but did not belong to the highly active group. Surprisingly, they were more abundant than the ASVs from the highly active group and represented a combined 6.30% and 7.74% of the community at DNA level in ridges and runnels, respectively (Figure 4.18).

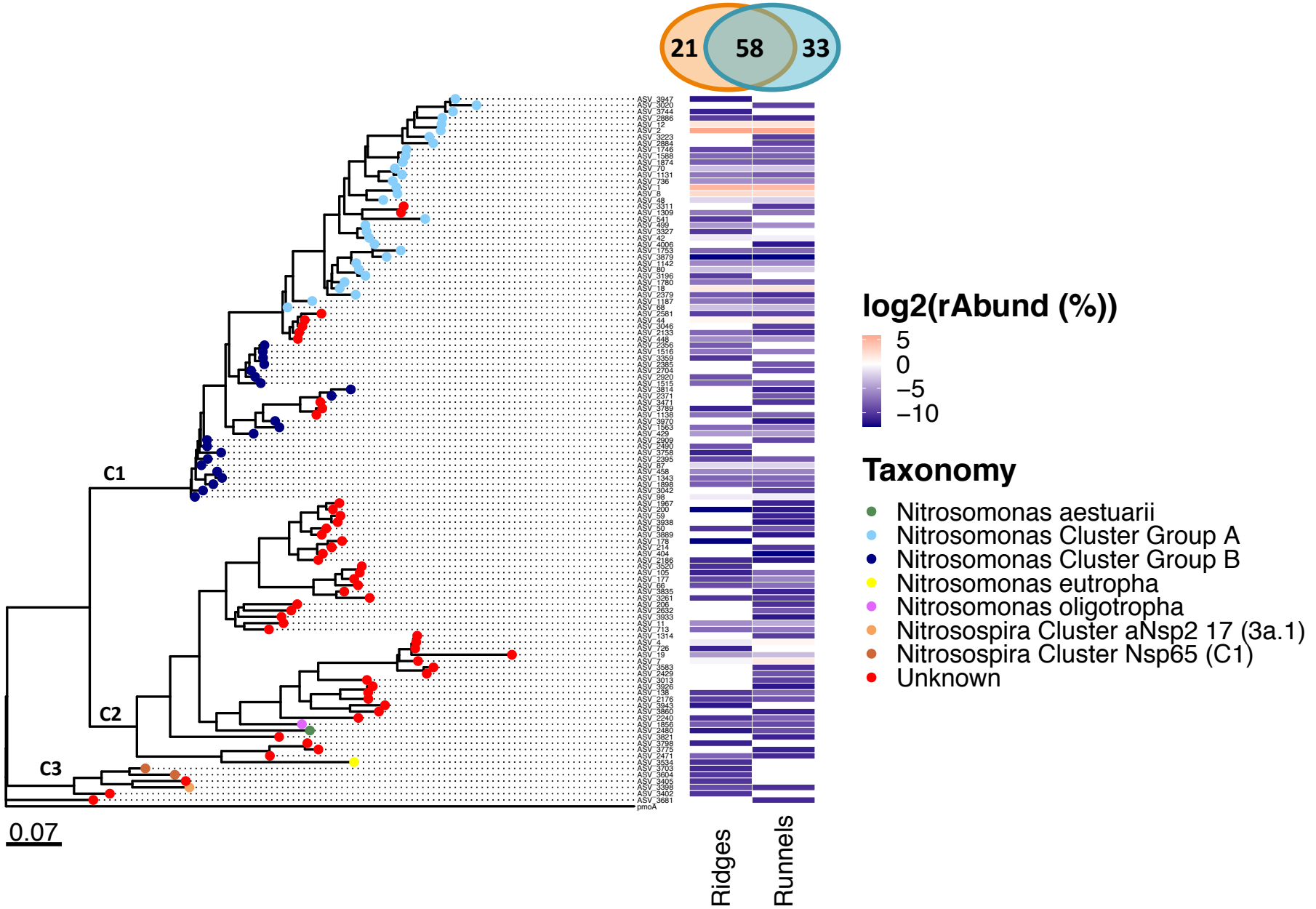


Figure 4.15 AOB community in ridges and runnels. Colour of tips represents the taxonomy attributed to ASVs using the BLCA method. The heatmap represents the relative abundance of each ASV in % in ridges and runnels (log2 transformed: red indicates relative abundance >1% and darker red indicates higher abundance;

blue indicates relative abundance <1% and darker blue indicates lower abundance; white indicates that the ASV is absent). Venn diagram on top of the heatmap represents the number of ASVs present in ridges only (orange oval), present in runnels only (blue oval) and shared (intercept). The three AOB clusters discussed in the text are indicated by C1, C2 and C3. Phylogenetic tree was drawn using FastTree, from Maftt-aligned ASVs sequences and rooted using the particulate methane monooxygenase (*pmoA*) gene. Tree colouring and heatmap was done in R using ggtree.

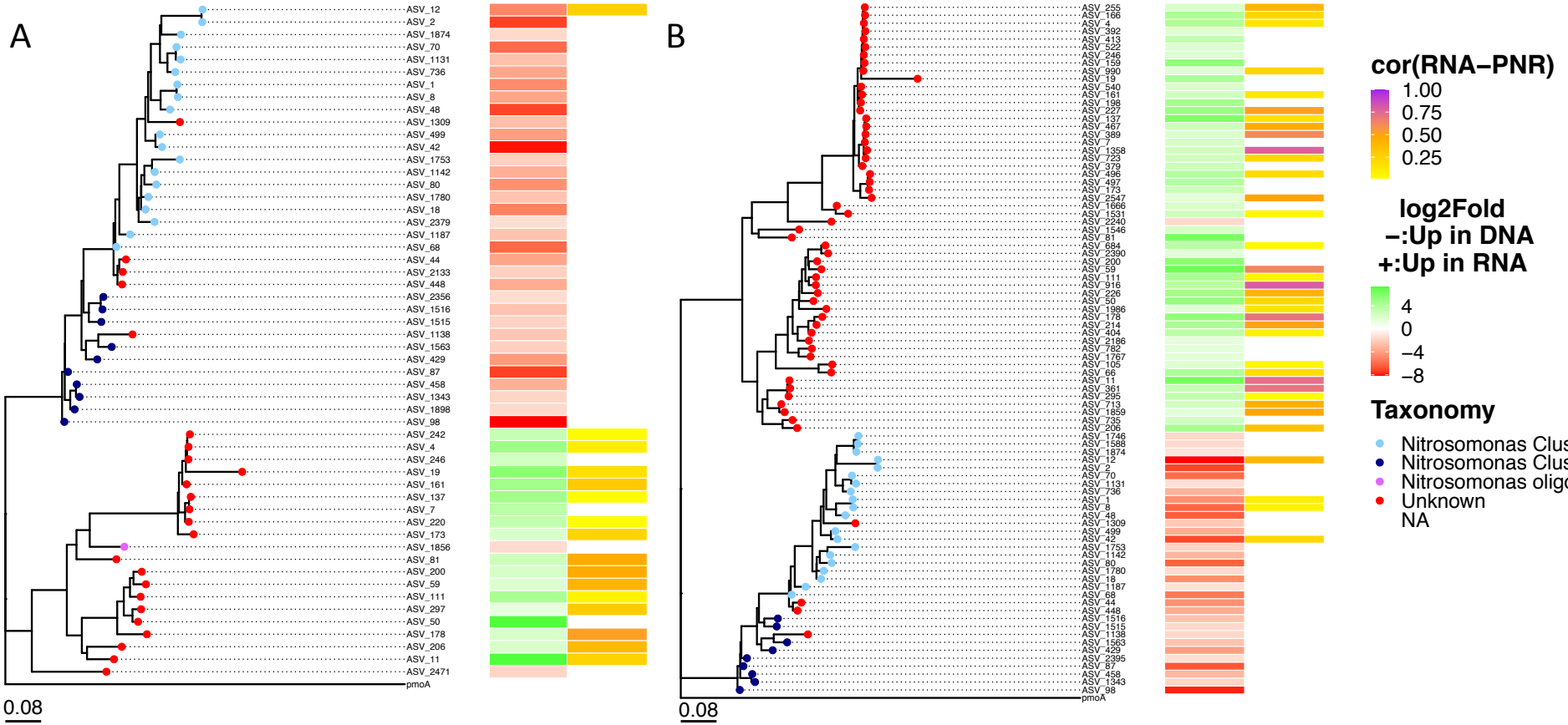


Figure 4.16 Identification of highly active AOB in ridges (A) and runnels (B). Colour of tips represents the taxonomy attributed to ASVs using the BLCA method. The first heatmap represents the log2 fold change (*i.e.* fold change, log2 transformed) in abundance of individual ASV between the DNA and RNA libraries (only ASVs with significant differences in abundance between DNA and RNA libraries are represented; red indicates higher abundance in DNA and green indicates higher abundance in RNA). The second heatmap represents the correlation between activity of individual ASV (abundance in the RNA library) and PNR in each separate structure (PNR +0µM ammonia; only positive values are represented). Phylogenetic tree was drawn using FastTree, from Maftt-aligned ASVs sequences and rooted using the particulate methane monooxygenase (*pmoA*) gene. Tree colouring and heatmap was done in R using ggtree.

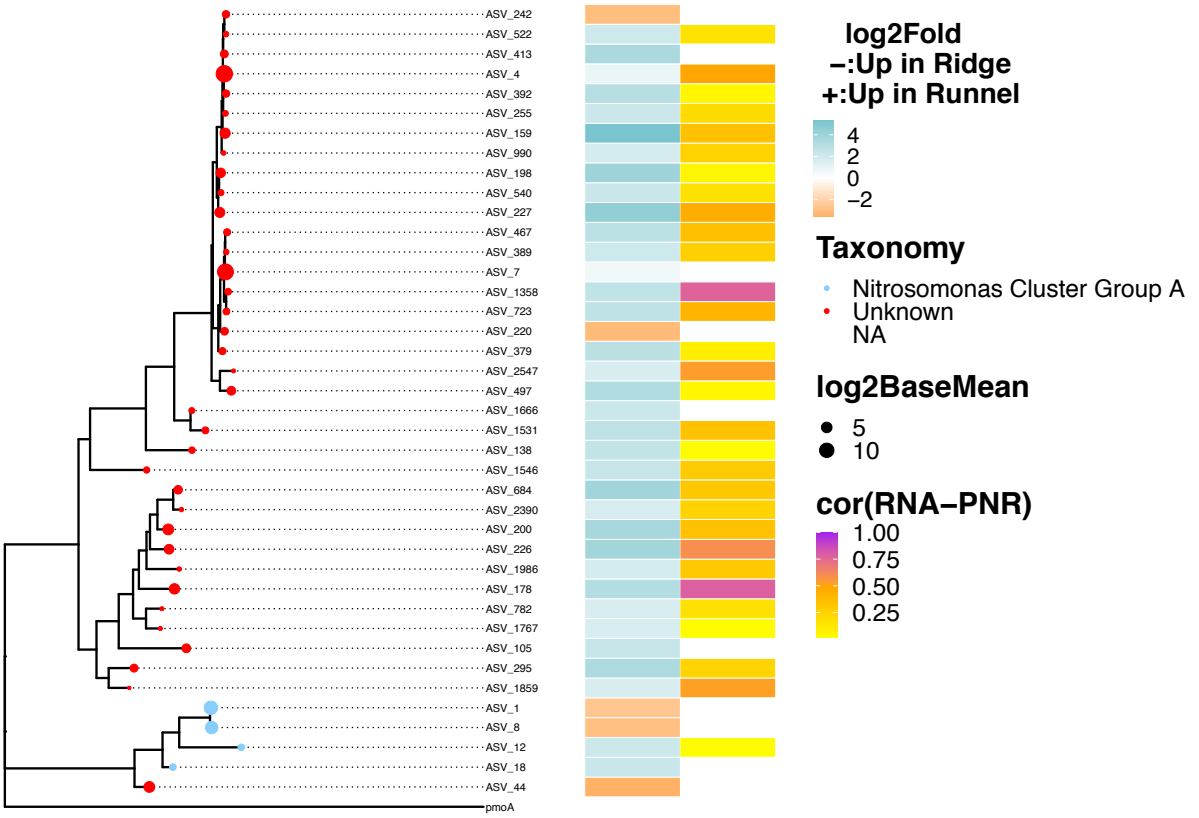


Figure 4.17 AOB ASVs differentially active between ridges and runnels. Colour of tips represents the taxonomy attributed to ASVs using the BLCA method. The first heatmap represents the log2 fold change (*i.e.* fold change, log2 transformed) in activity (abundance in the RNA library) of individual ASV between ridges and runnels (only ASVs with significant differences are represented; blue indicates higher activity in runnels and orange indicates higher activity in ridges). The second heatmap represents the correlation between activity of individual ASV and PNR (PNR +0µM ammonia; only positive values are represented). Size of tips represents the mean activity (log2 transformed) between ridges and runnels. Phylogenetic tree was drawn using FastTree, from Mafft-aligned ASVs sequences and rooted using the particulate methane monooxygenase (*pmoA*) gene. Tree colouring and heatmap was done in R using ggtree.

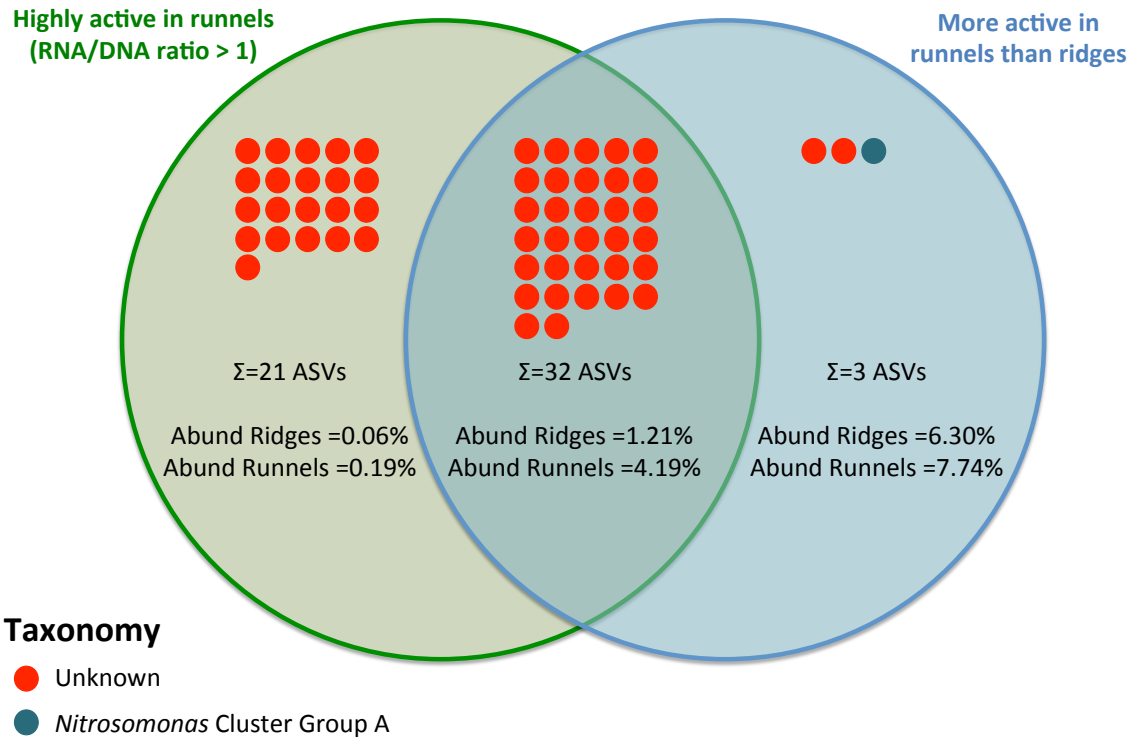


Figure 4.18 Identification AOB ASVs driving differences in nitrification between ridges and runnels. The Venn diagram shows AOB ASVs that are more active in runnels compared to ridges (blue circle) and those more identified as highly active (more abundant in the RNA library compared to the DNA one; green circle). Those at the intersection of the two circles are therefore likely drivers of the nitrification process. Taxonomy of the ASVs is indicated by the colour of the points. Percentages reported are the combined abundances of the ASVs in the DNA library in ridges (Abund_ridges) and in the runnels (Abund_runnels).

4.3.8.b AOA

Total AOA: AOA communities were dominated by ASV_1, representing 64.7% and 66.7% of the total community in ridges and runnels, respectively. Interestingly, the taxonomy could not be resolved for this sequence. However, based on the phylogenetic tree, ASV_1 seemed to be closely related to other *Nitrosopumilus* ASVs. The second most abundant ASV was ASV_5 (*Nitrosopumilus*) representing 13.2% and 13.5% of the total abundance in ridges and runnels, respectively. The next five most abundant were ASV_9 (*Nitrosopumilus oxyclinae*), ASV_8 (*Nitrosopumilus*), ASV_16 (unknown), ASV_26 (unknown) and ASV_37 (*Nitrosopumilus salaria*) represented a combined 19% and 17.2% of the total community in ridges and runnels, respectively. *Nitrosopumilus* therefore largely dominated the AOA community in both sediment structures. Two representatives of *Nitrosomarinus* (*Nitrosomarinus catalina*) were also detected in rides and runnels albeit at a very low level. Interestingly, more unique ASVs (not shared between ridges and runnels) were detected compared to AOB and these unique ASVs represented a higher percentage of the total community (Figure 4.19 and Table 4.6).

Active AOA: Like for AOB, the activity of individual AOA ASVs was estimated by calculating the RNA/DNA ratio in ridges and runnels. As seen in Figure 4.20 only a small number of ASVs were found to be highly active (RNA/DNA ratio>0) in ridges (6 ASVs) and in runnels (10 ASVs). All ASVs highly active in ridges were also active in runnels. Unlike for AOB, there was no evident link between activity of individual AOA ASVs and PNR measured in individual structure (Figure 4.20). When comparing ridges and runnels, it was observed that out of 11 ASVs with significantly different activities, 10 were more active in runnels and their activities correlated well with overall PNR (Figure 4.21). 5 out of these 10 ASVs were also identified as highly active in runnels (ASV_175: Unknown *Nitrosopumilus*, ASV_238: *Nitrosopumilus oxyclinae*, ASV_377: unknown, ASV_230: *Nitrosopumilus spD3C*, and ASV_7: *Nitrosopumilus spD3C*).

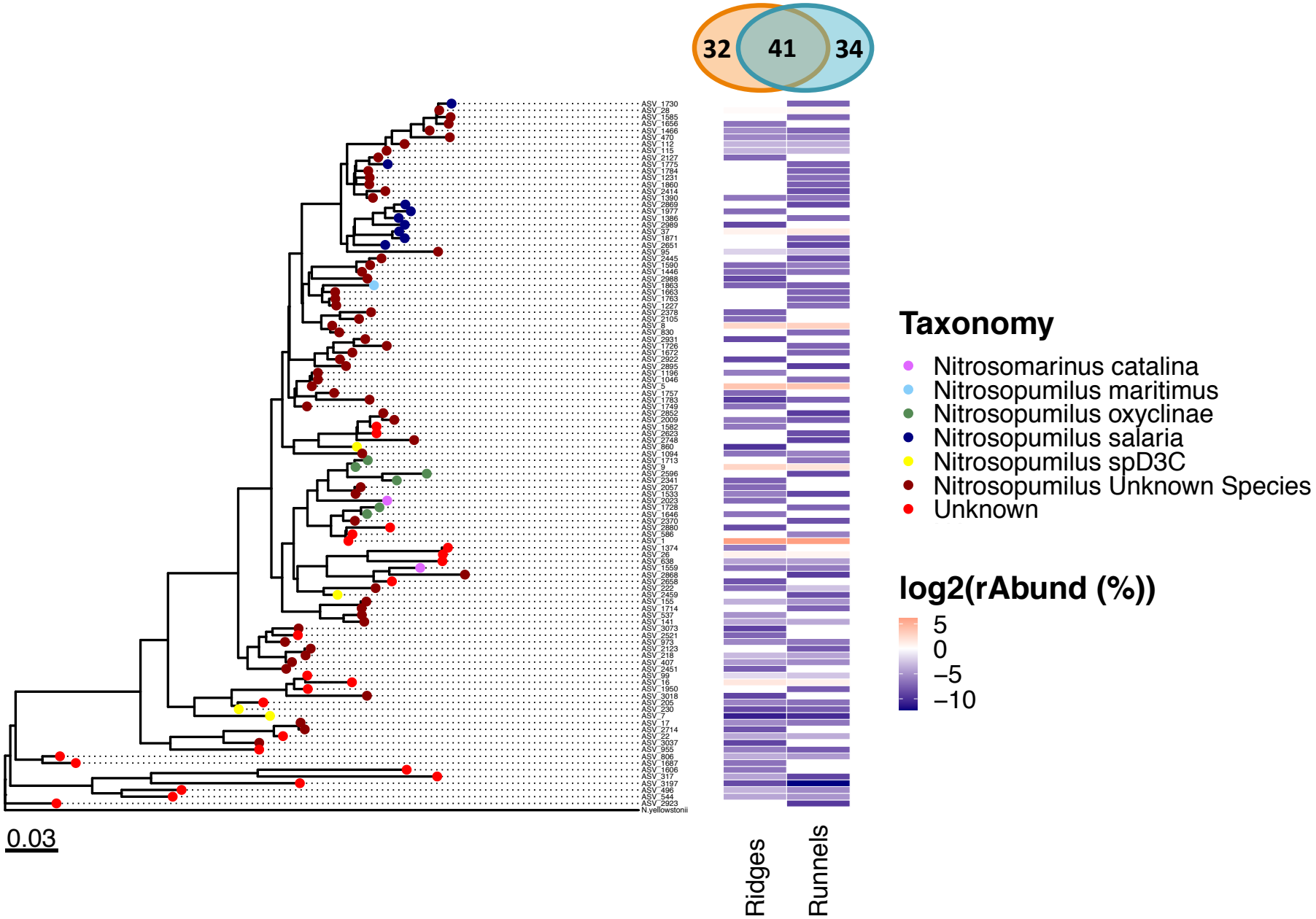


Figure 4.19 AOA community in ridges and runnels. Colour of tips represents the taxonomy attributed to ASVs using the BLCA method. The heatmap represents the relative abundance of each ASV in % (log2 transformed: red indicates relative abundance >1% and darker red indicates higher abundance; blue indicates

relative abundance <1% and darker blue indicates lower abundance; white indicated that the ASV is absent) in ridges and runnels. Venn diagram on top of the heatmap represents the number of ASVs present in ridges only (orange oval), present in runnels only (blue oval) and shared (intercept). Phylogenetic tree was drawn using FastTree, from Maftt-aligned ASVs sequences and rooted using the amoA gene from *N.yellowstonii*. Tree colouring and heatmap was done in R using ggtree.

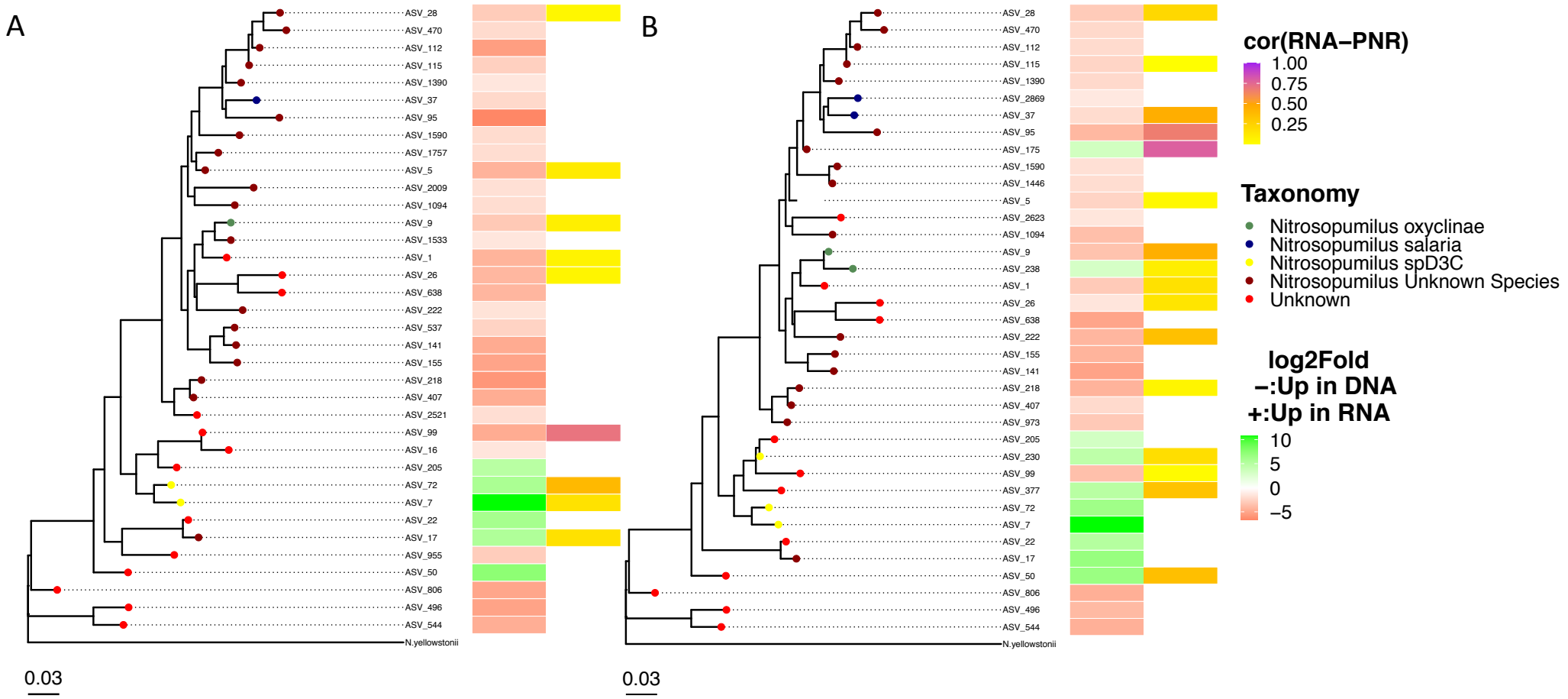


Figure 4.20 Identification of highly active AOA in ridges (A) and runnels (B). Colour of tips represents the taxonomy attributed to ASVs using the BLCA method. The first heatmap represents the log2 fold change (*i.e.* fold change, log2 transformed) in abundance of individual ASV between the DNA and RNA libraries (only ASVs with significant differences in abundance between DNA and RNA libraries are represented; red indicates higher abundance in DNA and green indicates higher abundance in RNA). The second heatmap represents the correlation between activity of individual ASV (abundance in the RNA library)

and PNR in each separate structure (PNR +0µM ammonia; only positive values are represented). Phylogenetic tree was drawn using FastTree, from Maftt-aligned ASVs sequences and rooted using the amoA gene from *N.yellowstonii*. Tree colouring and heatmap was done in R using ggtree.

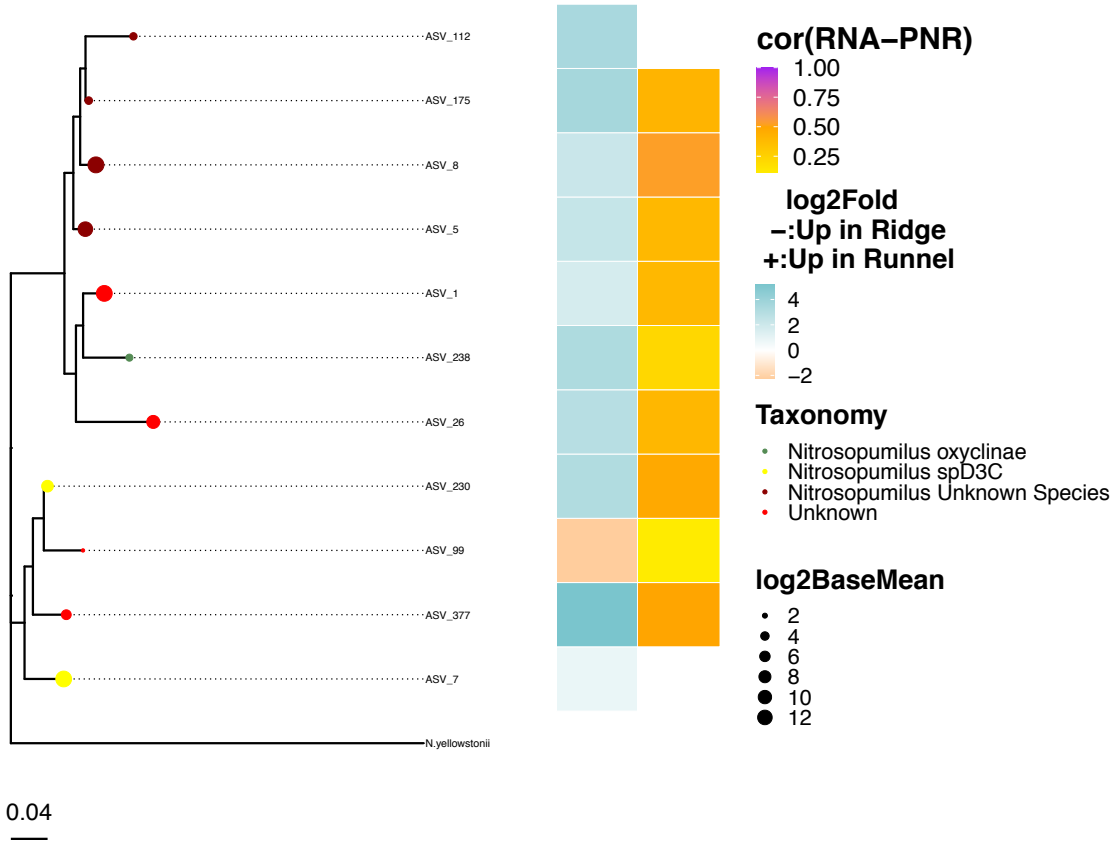


Figure 4.21 AOA ASVs differentially active between ridges and runnels. Colour of tips represents the taxonomy attributed to ASVs using the BLCA method. The first heatmap represents the log2 fold change (*i.e.* fold change, log2 transformed) in activity (abundance in the RNA library) of individual ASV between ridges and runnels (only ASVs with significant differences are represented; blue indicates higher activity in runnels and orange indicates higher activity in ridges). The second heatmap represents the correlation between activity of individual ASV and PNR (PNR +0µM ammonia; only positive values are represented). Size of tips represents the mean activity (log2 transformed) between ridges and runnels. Phylogenetic tree was drawn using FastTree, from Maftt-aligned ASVs sequences and rooted using the amoA gene from *N.yellowstonii*. Tree colouring and heatmap was done in R using ggtree.

4.4 Discussion

The ridges/runnels formation are sedimentary structures caused by the combined action of waves, sediment deposition and, possibly, bio-stabilisation. Because they are located on an intertidal mudflat, it can be expected that at high tide the conditions experienced by the communities in both structures are identical. Since the physio-chemical conditions are probably only different at low tide and the structures themselves only semi-permanent (Gouleau *et al.*, 2000), we first hypothesised that the microbial communities hosted in ridges and runnels would be identical. However, results obtained from sequencing of the *16S rRNA* amplicons showed that the communities are, in fact, different between the two structures. This is an important

finding as it shows that when interested in the distribution of microbial communities in an ecosystem, small-scale physical differences in the physical arrangement of the sediment can be significant. Significant differences were found when considering the Bray-Curtis distances between ridges and runnels. The Bray-Curtis dissimilarity index is based on the differences in ASVs proportions between conditions being compared. A significant difference in the Bray-Curtis index therefore indicates that the ridges/runnels structure affects the proportions of ASVs found in ridges and runnels but does not inform on their phylogenetic distances. On the other hand, the Unifrac distance is measured as the proportion of branch length that are not shared between conditions relative to the total branch length in the phylogenetic tree. For *16S rRNA*, a high number of ASVs was not shared between ridges and runnels explaining the highly significant effect of the ridges/runnels on the Unifrac distances. Finally, the Weighted Unifrac is similar to the Unifrac but weights branches by the differences in abundances between conditions being compared. In other words, the Unifrac distance between two communities with the same members but at different proportions is zero but the WUnifrac is not (Chang *et al.*, 2011). For, *16S rRNA* the ridges/runnels structures affect both the proportions of ASVs (Bray-Curtis) and the number of unique branches in the phylogenetic tree (Unifrac). Combined together, this results in strong differences in the WUnifrac distance. The majority of ASVs (80%) were not shared between ridges and runnels. However, these unique ASVs only represented 36% of the total abundance indicating that most were low-abundance ASVs (Table 4.6). This could be the reflection of the new approaches used here. Indeed, the construction of ASVs is based on single differences between sequences (*i.e.* 100% similarity threshold) as opposed to the OTU where sequences are clustered at a defined similarity threshold, usually 97%. By constructing OTUs, more sequences are therefore merged together, and fewer differences are likely to be found. It is debatable whether beta-diversity measures should be undertaken at the single ASV level, especially given the possibility of errors introduced during the PCR and sequencing steps. We then merged ASVs at higher taxonomic levels and evaluated the effect of the ridges/runnels structures using the Bray-Curtis dissimilarly index (Figure 4.13). At all phylogenetic level tested (Family, Order and Class), the effect was significant (*p*.value PERMANOVA<0.001) confirming that the ridges/runnels structures do affect the overall microbial community assemblages in the Montportail-Brouage mudflat. In contrast, the effect of the sampling day was not significant, indicating that the microbial communities were homogenous among sample type between different locations. Further investigation using CCA revealed that pH, TOC and the sediment grain size (% of silt+ clay) are the possible drivers of the differences observed in the microbial communities. To explain the lower pH in the ridges,

it can be hypothesised that the higher organic content found in the ridges favours heterotrophic metabolisms, which in turn results in higher production of CO2, inducing acidification of the sediment.

AOB were found to be dominant over AOA in both ridges and runnels and were found to also be more active. This is in accordance with previous studies showing the dominance of AOB in ammonia-rich/high salinity estuarine sediment (Mosier and Francis, 2008; Santoro *et al.*, 2008; Magãlhaes *et al.*, 2009; Zheng *et al.*, 2013; Smith *et al.*, 2014; Urakawa *et al.*, 2014; Li *et al.*, 2015; Damashek *et al.*, 2015; Duff *et al.*, 2017; Zhang *et al.*, 2018) but contradicts other reports of AOA dominating the AOM communities in marine sites (Caffrey *et al.*, 2007; Marton *et al.*, 2015; Bernhard *et al.*, 2019; He *et al.*, 2018), including high ammonia marine sediments sites (Moin *et al.*, 2009; Cao *et al.*, 2011). AOA were however active, as some AOA *amoA* transcripts could be amplified and sequenced. This could reflect the ability of AOA to remain active in AOB dominated environments, albeit to a much lower level. Interestingly, it was observed that the ratio of AOB/AOA gene abundance was significantly influenced by the ridges/runnels structures, with a higher ratio observed in the ridges (≈50 in ridges vs. ≈30 in runnels), indicating an influence of the ridges/runnels structures on AOB/AOA niche differentiation. Additionally, this study shows that the ridges/runnels structures strongly influence the nitrifiers' communities. Interestingly, a lower proportion of the variance was explained by the ridges/runnels structures when considering Unifrac distances (Figure 4.13) indicating a lower number of unique branches in the phylogenetic tree. Nevertheless, highly significant differences were found when considering WUnifrac distances indicating an overall strong effect of the sediment type on AOA and AOB communities. Plus, for AOA, a high number of ASVs, representing a significant proportion of the total abundance, were not shared between ridges and runnels (Table 4.6). For AOA and AOB, ASVs sequences were checked for length and sequence and combined together if they translated to the same protein. It is therefore likely that these differences have a real ecological significance, indicating a real impact of the ridge/runnel structures on nitrifier communities. CCA analysis indicates a possible role of SGS, ammonia and pH for AOA and ammonia and pheophytin for AOB. Ammonia is a well-known parameter that shapes the distribution of nitrifiers in natural environments (Erguder *et al.*, 2009; Schleper and Nicol, 2010; Shen *et al.*, 2012; Sims *et al.*, 2012) and in particular in estuaries (Urakawa *et al.*, 2014; Damashek *et al.*, 2015; Damashek and Francis, 2018; Zhang *et al.*, 2018). Indeed, different concentrations of the substrate will advantage different AOA/AOB species based on the affinity of their enzymes for ammonia.

pH is also a known factor influencing nitrifiers (Nicol *et al.*, 2008; Wessén *et al.*, 2010) in part because it controls the NH4+/NH3 balance. The effect of SGS could link to the oxygenation of the sediment. Indeed, finer and more compact sediment will likely reduce O2 penetration in the top layers, impeding aerobic metabolisms such as nitrification. The role of pheophytin on AOB communities could possibly indicate an interaction between phototrophs and AOB in the Montportail-Brouage mudflat. Dense mats of microbial phototrophs form on the surface of the sediment in ridges at low tide (Lavergne *et al.*, 2017) but not in the runnels. In can therefore be hypothesised that in ridges a negative interaction takes place between phototrophs and AOB through competition for ammonia. It has been shown before that the presence of phototrophs can strongly inhibit bacterial nitrification in sediments (Laima *et al.*, 2002; Risgaard-Petersen, 2003; Risgaard-Petersen *et al.*, 2004) and bioreactors (Choi *et al.*, 2010) due to the reduction of available ammonia as these phototrophs exhibit higher uptake ability. Plus, the formation of dense microalgae mats can lessen O2 penetration in the sediment, further reducing the available reactant pools for AOB metabolisms (Risgaard-Petersen, 2003). On the other hand, in runnels, it can be expected that the competition between AOB and phototrophs is reduced due to the much higher ammonia concentration. In fact, in this type of non-nitrogen-limiting conditions, positive interaction can be expected via the production of O2 as a by-product of photosynthesis that can be consumed by AOB as shown before to occur in estuarine sediments (Petersen *et al.*, 1994; Joy, 2001).

Despite the intensive effort deployed here to monitor relevant physio-chemical parameters to explain the difference in nitrification rates, it should be noted that other factors might play a role but have not been measured. For example, sulphide concentration has been shown to significantly inhibit nitrification in pure cultures of AOB (Juliette *et al.*, 1993), sludge (Bejarano-Ortiz *et al.*, 2015; Delgado Vela *et al.*, 2018) and sediment (Joye and Hollibaugh, 1995; Erguder *et al.*, 2009). During the campaign to collect the samples used in this study, a black colouration was observed in the sediment of the Montportail-Brouage mudflat at a few centimetres below the surface (data not shown), which is an indicator of the presence of iron sulphide due to the likely production of sulphide in the lower anoxic layers of the sediment. Laima *et al.* (2002) showed that the desiccation of ridges during emersion (low tide) induces an increase in salinity and a flow of pore water from the bottom layer of the sediment to the surface layer. If sulphide is present in those bottom sediment layers, it is possible that it will be transported to the top layers via this mechanism and inhibits AOB activity in ridges. Whether sulphide significantly affects nitrification in this environment should be further investigated in

future studies. As discussed previously, another important parameter influencing nitrifier activity, and that wasn't measured here, is the O2 penetration in the sediment. During the 2016 campaign, a higher O2 penetration was measured in the runnels compared to the ridges (Figure S.2). Laima *et al.* (2002) also showed a higher O2 penetration in runnels compared to ridges in the Montportail-Brouage mudflat sediment. This is in accordance with the higher measured PNR rates and higher *amoA* transcription measured in the runnels here. It would therefore be interesting to test how sulphide and oxygen concentration change in these sediments following sediment exposure at low tide. For example, to determine if a flow of sulphide-rich water from the bottom layers upon sediment drying at low tide is occurring, sulphide concentration changes at different depth and different time points could be recorded using micro-sensors. A similar depth/time approach could be used to investigate how oxygen penetration changes in sediments (especially ridges) at low tide. Finally, differences in sediment erodability and macro-fauna composition between the two sediment structures could have an impact on sediment mixing, organic matter remineralisation and ammonia release, both of which could affect microbial assemblages and activities (Gouleau *et al.*, 2000; Laima *et al.*, 2002). It would be interesting to construct ridges and runnels microcosms with and without native macrofauna and measure PNR/AOB activity to see how a phenomenon such as bioturbation influences nitrification. A summary of the main findings and hypotheses generated by this study is presented in Figure 4.22.

Previously, the ridge/runnel structures present in the Montportail-Brouage mudflat have been shown to significantly impact the rates of microbial nitrification (Laima *et al.*, 1999, 2002) and this observation was further confirmed in this study. Other studies have shown the importance of soil physical arrangement to boost microbial nitrogen cycle metabolisms in constructed wetlands (Su *et al.*, 2018; Wang *et al.*, 2018). In these studies, authors reported the presence of nitrogen cycle hotspots in riparian zones and were able to increase microbial nitrogen metabolisms, especially nitrification, by increasing the interface between land and water (Su *et al.*, 2018). Together with our present study, these results indicate that landscape morphology is important to create hotspots of microbial activity.

Interestingly, Su *et al.* (2018) found that higher activity in their engineered systems did not correlate with total AOB abundance. Similarly, when comparing nitrification rates between two different estuaries, Duff *et al.* (2017) showed significant differences that were not correlated with nitrifiers abundances. Here, we found that the abundance of AOB was higher

in ridges compared to runnels, in contradiction with the measured PNR. On the other hand, a comparison of bacterial *amoA* transcript abundances revealed an opposite trend with higher abundance in runnels compared to ridges. A similar trend was observed for *nxrB* (higher abundance in ridges compared to runnels at the DNA level and inversely at cDNA level). This shows the importance of quantifying functional genes at the transcript level to obtain a better understanding of microbial processes in the natural environment (see also Graham *et al.*, 2011). The success of transcript-based studies depends on the integrity of the starting RNA material after extraction from the environment (Chapter II; Cholet *et al.*, 2019) and on an RT protocol that faithfully produces cDNA (Chapter III; Cholet *et al.*, 2020). In this study, one sample was excluded because it showed lower integrity compared to other samples, illustrating how to use RNA integrity indexes in concrete situations.

Results from the sequencing of the Bacterial *amoA* genes revealed the existence of three separate clusters (C1, C2 and C3) based on the phylogenetic tree (Figure 4.15): Cluster C1, which was dominant in terms of abundance in both ridges and runnels, and in particular ASVs 1,2,8,12 and 18 which represented a cumulative 95.1% and 88.1% of the total abundance in ridges and runnels, respectively. Cluster C2: containing a majority of unknown ASVs, generally of low abundances and some representative of *Nitrosomonas aestuarii, Nitrosomonas eutropha* and *Nitrosomonas oligotropha*. C3: containing some unknown AOB and representatives of *Nitrosospira*, also low in abundance. Since clusters C1and C2 were more closely related to each other compared to C3 and that the known ASVs in C1 and C2 were only identified as *Nitrosomonas*, while those in C3 were only identified as *Nitrosospira* it can be hypothesised that the unknown ASVs in C1 and C2 are of the genus *Nitrosomonas* while the unknown in C3 is of the genus *Nitrosospira*.

Strikingly, when looking at the RNA/DNA ratio of individual ASVs, it was found that the ASVs in C2 (the low abundant cluster; Figure 4.15 and 4.18) were generally highly active (RNA/DNA ratio>1) and their abundance at RNA generally correlated with PNR measured in each structure. On the other hand, the ASVs in C1 (the high abundance cluster; Figure 4.15) displayed low activity (RNA/DNA ratio<1) (Figure 4.16) indicating that, although present at high proportion, many did not transcribe their *amoA* genes. No transcripts form cluster C3 was recovered indicating that *Nitrosospira* likely does not play a significant role in nitrification in this ecosystem, or at least at the time of sampling. When comparing ridges and runnels, it was observed that the highly active AOB ASVs from C2 were generally more active in runnels and

that their abundance at RNA level correlated with overall PNR (Figure 4.17 and 4.18). From these observations, we therefore hypothesise that these highly active but low abundance ASVs are the drivers of the nitrification process in the Montportail-Brouage mudflat and that the higher AOB abundance in ridges reflects the higher presence of inactive *Nitrosomonas* in this structure, which is also reflected by the lower recorded PNR per AOB cells in ridges (Figure 4.12). This, again, shows the importance to study microbial communities at both DNA and RNA level, including when doing amplicon sequencing, to get a better insight into their function.

For AOA, the community (at DNA level) was also dominated by a few ASVs with ASV_1 representing, on its own, 64.7% and 66.7% of the total community in ridges and runnels, respectively. When adding the next 6 most abundant ASVs (ASV 5,9,8,16,26 and 37), this accounted for a total of 96.9% and 97.4% in ridges and runnels, respectively. Interestingly, there were more unique AOA ASVs (not shared between ridges and runnels) compared to AOB and they represented a higher proportion of the community (Figure 4.19 and Table 4.6) indicating that some AOA might preferentially grow in ridges and others in runnels. However, overall, both ridges and runnels were dominated by *Nitrosopumilus* (which probably included ASV_1). When looking at the RNA/DNA ratio for individual AOA ASVs, it was observed that, like for AOB, some ASVs were highly active while others, including the dominant ASV_1, 5 and 8 and 26 were of low activity (Figure 4.20). Yet, despite their low RNA/DNA ratio, these ASVs were more active in runnels than in ridges and their abundance at RNA level correlated with PNR (Figure 4.21) indicating that they could still participate in the nitrification process, albeit to a much lower extent compared to AOB.

Understanding which are the active microbial groups involved in the global nutrient cycle is of paramount importance to try to anticipate their responses to future environmental changes (Chapin *et al.*, 2009; Dutta and Dutta, 2016; York, 2018; Cavicchioli *et al.*, 2019). In particular, it is expected that nitrification could be substantially impacted, in both terrestrial and aquatic ecosystems by global climate change (Szukics *et al.*, 2010; Beman *et al.*, 2011; Liu *et al.*, 2015; Hu *et al.*, 2016; Breider *et al.*, 2019). In this study, we found that the most active AOB, which likely drive the difference in nitrifications between ridges and runnels, were unknown (likely unknown *Nitrosomonas*) therefore indicating that could represent novel *Nitrosomonas* isolates. More efforts are therefore needed to directly identify active nitrifiers from the environments.

In conclusion, this study shows that the ridges/runnels sedimentary structures significantly impact the distributions of microorganisms in the Montportail-Brouage mudflat and highlights the importance of taking into consideration the sediment spatial heterogeneity in environmental microbiology studies. We also showed that the ridges/runnels structures have a strong impact on nitrification, with PNR significantly higher in runnels. Nitrification in this ecosystem is likely driven by a low abundance but highly active AOB cluster while the most abundant AOB cluster displayed low activity. AOA was also shown to be active in both ridges and runnels albeit to a lower level. Differences in nitrification rates are better explained by comparisons at RNA level illustrating the need for reliable and reproducible transcriptomic workflow when studying biogeochemical cycles. This study also highlights the need for the development of new techniques for the identification of active nitrifiers in natural environments.

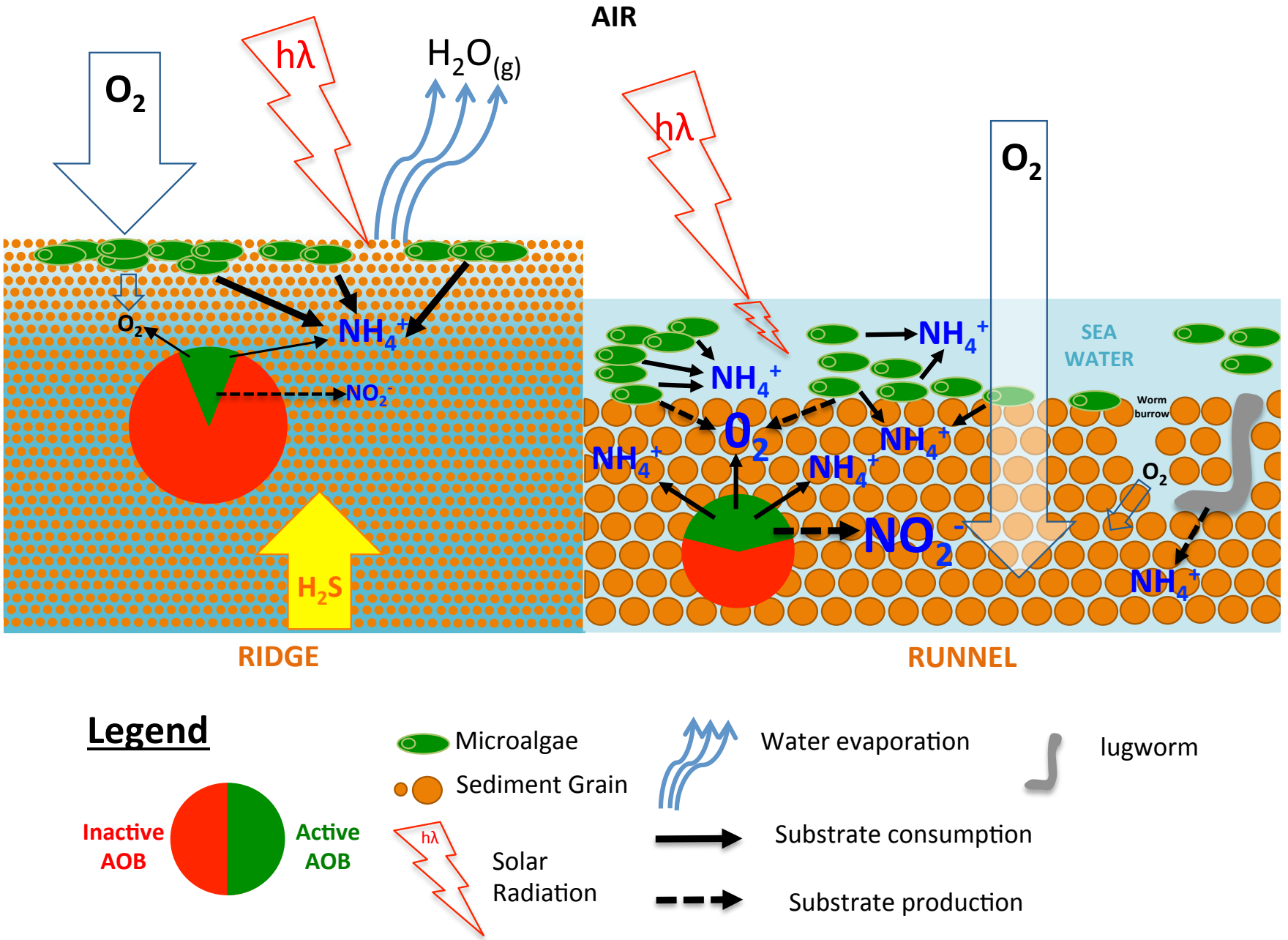

AIR
O_2
hλ
$H_2O_{(g)}$
NH_4^+
NO_2^-
H_2S
RIDGE
SEA WATER
Worm burrow
RUNNEL
Legend
Microalgae
Sediment Grain
Inactive AOB
Active AOB
Solar Radiation
Water evaporation
Substrate consumption
Substrate production
lugworm

Figure 4.22 Schematic representation of the hypothesised AOB dynamics in the Montportail-Brouage mudflat. The hypothesised higher ammonia uptake by microalgae in ridges (left) compared to AOB is represented by different arrow size. The dense microalgae mats on the top of the sediment also possibly reduce O_2 penetration, as represented by a smaller arrow. The desiccation of the most upper sediment layer in ridges is represented by a pore water gradient (blue font), which results in a possible suction of H_2S rich water form the bottom sediment layers (yellow arrow). The finer and denser sediment in ridges is represented by smaller grain size. In runnels (right) the higher ammonia availability reduces the competition between microalgae and AOB and a possible positive interaction via the O_2 produced by the phototrophic metabolism. Combined with a higher O_2 penetration, this creates a well-oxygenated sediment (in the top layers). AOB are represented by circles in ridges and runnels and the proportion of active and inactive cells by green and red colours in the circles, respectively. The smaller circle in runnels represents the lower AOB abundance compared to ridges. Substrate consumption and production are represented by solid and dashed arrows respectively. The possible role of the macrofauna is represented in runnels with the creation of a worm burrow by a polychaete, inducing the diffusion of O_2 in deeper sediment layer and the production of ammonia in dejections.

General Conclusion and Future Perspectives

5.1 Importance of RNA Integrity and Possible Improvement of the R_{amp}

In the first part of this thesis, the importance of measuring RNA integrity has been shown in a series of laboratory-controlled RNA degradation experiments. Indeed, RNA degradation significantly alters RT-Q-PCR results, inducing strong increases in the Ct of the target transcripts and significantly affected the community composition structure recovered by RT-PCR-sequencing of functional transcripts. While both the RIN and the R_{amp}, reflected the degradation status of the RNA preparation, it was shown that the R_{amp} generally reflected the degradation status of the mRNA better.

A drawback of the R_{amp} approach is the fact that it relies on the presence of multiple conserved regions along the target transcript (*glnA*) which poses several problems: i) primer design is restricted to these conserved regions which limit the size of the amplicons that can be produced and therefore, the possibilities for the length of both the long and small amplicons and ii) the degree of conservation of the different regions might vary among bacterial, which might result in the different primers having different binding efficiencies depending on the microbiome present in the sample. Furthermore, the use of Q-PCR limits the size for the long amplicon, with efficiencies generally reduced for amplicons >400 bp. Finally, the SYBR fluorescence detection method is affected by primer dimer when considering small amplicons.

Figure 5.1 presents a possible method that would resolve some of the issues with the current R_{amp} method by adapting the 5'RACE method (Schaefer, 1995): This method starts with 1) the attachment of the reverse gene specific primer on the target mRNA and 2) the synthesis of cDNA 3) The RNA template is then degraded and the cDNA is purified. 4) A poly C tail is attached at the 3' end of the cDNA which is used for 5) the attachment of an anchor sequence. The anchor contains a known sequence and a ploy G/I (deoxyino-sine) site 6) The anchor site and the gene specific sequences are used as forward and reverse sites respectively for PCR amplification. Finally, 7) the PCR is run on an electropherogram to check the size distribution of the PCR products. In the case of non-degraded RNA, only one size should be observed, equal to the distance between the binding site of the reverse primer and the transcription initiation site of the target gene. For degraded RNA, multiple peaks are observed with smaller lengths than the one expected.

The ratio of peak area of expected size to the total peak area can be used as a measure of RNA integrity, which should be bound between 1 (perfect RNA) and 0 (totally degraded RNA). In theory, this method has the advantage of relying on only one conserved site and the use of PCR instead of Q-PCR allows for the amplification of longer fragments making this approach, in theory, a more sensitive measure of RNA integrity. Developing new methods for the evaluation of RNA integrity, and in particular, mRNA is crucial and should be the focus of more studies future.

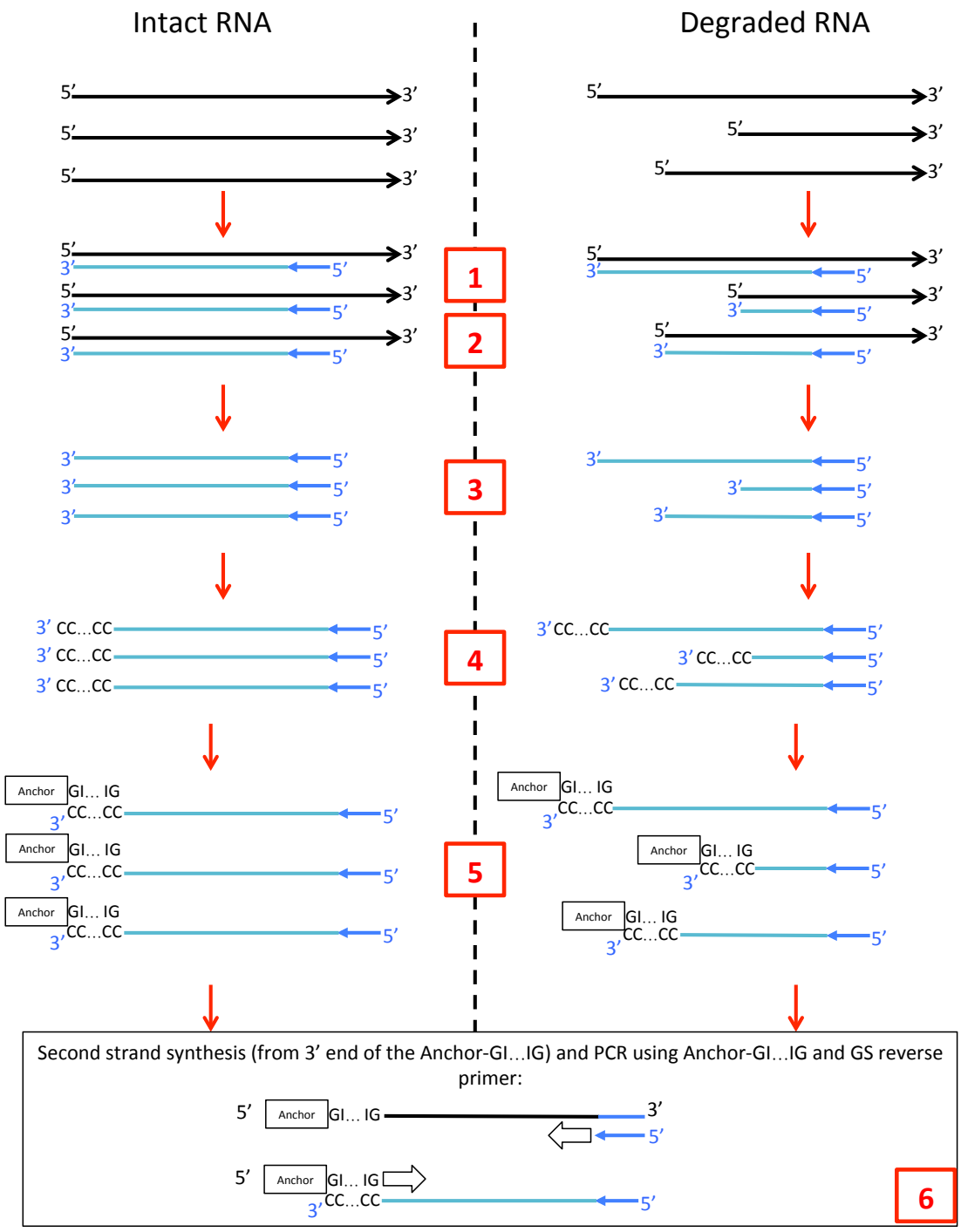


Figure 5.1 Schematic representation of the adaptation of the RACE technique for the evaluation of RNA integrity. Two different situation are presented: left: Intact RNA and right: degraded RNA, which are truncated at the 5'end. The target mRNA is represented as black arrows. Gene specific reverse primer is represented by the blue arrows and the cDNA by blue lines. Red numbers represent the different steps described in the text above. x (2; X; Y; Z)^n: Expected copy number after PCR (X: Y; Z the efficiencies for the different lengths). In 7, the green dashed arrow represents the expected size.

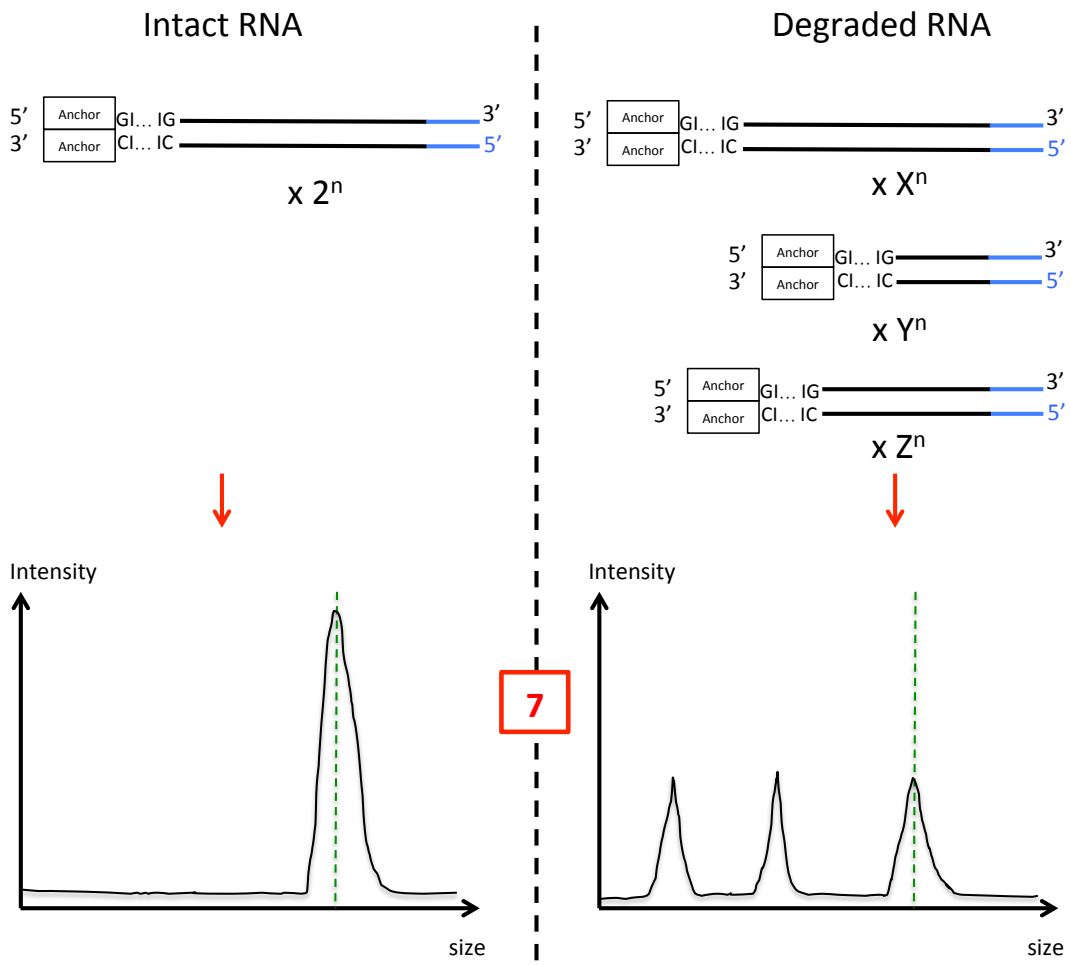


Figure 5.1. (continued)

5.2 Importance of the Reverse-Transcriptase Step

As shown in chapter III, having an RT protocol that faithfully reverse-transcribes RNA into cDNA is essential to achieve accurate and sensitive detection of environmental transcripts. In this thesis, only four different RT enzymes and two different priming strategies have been tested. The general conclusion was that the SSIV enzyme and GS priming was the most efficient combination. It should be noted that several other kits exist and might exhibit better performances than the SSIV enzyme. For example, it could have been interesting to test the QuantiNova and QuantiTech kits from Qiagen, or the ProtoScript kit from New England that advertises the sensitive detection of mRNA. Other aspects of the RT reaction like the removal of the RNA template using RNaseH enzyme or the removal of rRNA before RT of the mRNA have been shown to affect subsequent Q-PCR and sequencing and would need to be further investigated in environmental microbiology. Even though these were not investigated here, we still show that the SSIV/GS combination accurately and precisely reverse-transcribes RNA into cDNA. Improvements are always possible and should be attempted in future. However, it is

unlikely that a flawless method will ever be developed. The important is to understand the limits of the methods used and to use a consistent approach for all samples tested. In our case, this consistent approach means 1) to systematically measure RNA integrity and 2) use the SSIV enzyme and GS priming for functional gene studies.

In order to compare the RT strategies, mock communities containing RNA with known sequences were constructed, seeded into environmental RNA background and retrieved via RT-PCR-sequencing (chapter III). Such validations should be more routinely carried out to ensure the validity of the results obtained. Here, it informed us that transcriptomics diversity comparison between samples (β diversity) is valid while, on the other hand, small differences in α diversity are not always meaningful. Such conclusion significantly helped interpret results obtained in Chapter IV, when an actual ecological hypothesis was being tested.

This approach, although being unusually thorough in an environmental microbiology context, has limits that need to be discussed. The RNA used for the mock was obtained via *in vitro* transcription, which might be different from *in vivo* transcripts. Indeed, it is now recognised that bacteria use RNA modifications as post-transcriptional regulation mechanisms to modulate translation (Hoernes *et al.*, 2016; Hör *et al.*, 2018). These base-modifications have been shown to interfere with the RT reaction, introducing, base substitutions, insertions/deletions and truncation in the cDNA (Kietrys *et al.*, 2018; Potapov *et al.*, 2018). If we were to move from OTUs to ASVs, with single resolutions as the cut-off for distinguishing two sequences, the impact of these RNA modifications on the RT fidelity would also need to be determined. Furthermore, their significance as post-transcriptional regulations mechanisms in the environment as been yet to be explored and could constitute the next breakthrough in our understanding of microbial dynamics. Nevertheless, we aimed to introduce greater rigour into mRNA quantification from environmental samples, and introduce best practices that can be more widely adopted by the community to would improve the reliability of the approach and therefore translate to meaningful data from which ecological understanding can be derived.

5.3 Effect of the Ridge/Runnel Structures on Microbial Composition and Activity

5.3.1 Limits of the primers for the detection of AOB

In Chapters IV, members of the family Nitrosococcaceae (containing the genus of AOB *Nitrosococcus* in the SILVA taxonomy classification) were detected by sequencing of the *16S rRNA* genes but were not detected by sequencing of the *amoA* gene. As discussed in 4.5, this is due to the current AOB primers that were designed to specifically target beta-proteobacterial AOB. This is a major drawback of this primer pair as *Nitroscoccus oceani* is distributed worldwide in the ocean (Ward and O'Mullan, 2002) and has been found in sediment (Zaccone *et al.*, 1996; Dohra *et al.*, 2019). It is therefore necessary to design new primers that allow the simultaneous detection of beta and gamma AOB to further our understanding of AOB dynamics in the environment.

To determine if such primers could be designed, A list of full length or quasi full length (containing the start codon and length superior to 700bp) AOB *amoA* sequences were downloaded from NCBI including a representative of β-proteobacterial AOB (*Nitrosomonas* and *Nitrosospia*) and γ-proteobacterial AOB (*Nitrosococcus*). The *amoA* sequence of the newly discovered acid-tolerant γ-proteobacterial AOB was retrieved by blasting the *amoA* sequence of *Nitrosococcus Halophilus* (NC_013960) against *Candidatus Nitrosoglobus Terrae* (AP014836.1) full genome. A significant match was found (71.85% identity) between positions 1665209 and 1664496 on the minus strand of the genome and added to the database. Sequences were then aligned using Mafft and primers were searched using PrimerProspector with the default settings. DeNovo Primers were then aligned against the initial full-length AOB *amoA* sequences and compared based on the following criteria: i) the total number of perfectly matching bases ii) total number of perfectly matching bases at the 3'end of the primer iii) the total number of degeneracies and iv) the total number of unambiguous bases at the 3'end of the primer. When multiple primers were found at the same position, only the best candidate was chosen based on the above criteria.

The PrimerProspector search returned four and two possible regions for the design of the forward and reverse primers respectively. Comparison of the different primers showed that primers 163f and 492r were the best candidates as universal AOB primers as they displayed

the higher number of perfectly matching bases (163f and 492r will be referred to as unAOBf and unAOBr, respectively, thereafter).

To compare the newly designed primer pair to the existing ones, the WS of the unAOB primer pair was displayed as a heatmap on the phylogenetic tree of the AOB *amoA* sequences used for primer design, along with the BamoA primer pair and the 310.627 primer pair (specific to *Nitrosococcus*) (see Table 5.1 for the full list of Primers). The specificity of the primers was evaluated by calculating the WS of the primers against a list of full-length *pmoA* bacterial sequences downloaded from NCBI, including marine sediment methanotrophs such as *Methylocaldum marinum*, *Methylomicrobium japanense* and *Methyloprofundus sediment*.

Table 5.1 Primers being compared for the amplification of AOB *amoA*.

Primer	Primer pair	Sequence	Orientation	Reference	Target
BacamoA1-F	BamoA	GGGGHTTYTACTGGTGGT	Forward	(Hornek *et al.*, 2006)	β-AOB
BacamoA2-R		CCCCTCBGSAAAVCCTTCTTC	Reverse		
310-F	310.627	GGTGAGTGGGCTAACCG	Forward	(Norton *et al.*, 2002) (Junier *et al.*, 2008)	γ-AOB
627-R		GGTACCTTTCTCAAYCATTC	Reverse		
uniAOB-F	unAOB	GWGACTGGGAYTTCTGG	Forward	This study	β and γ - AOB
uniAOB-R		SCAAASAKCGGCCAGTT	Reverse		

Sequences are reported in 5' to 3' orientation.

As expected, the BamoA primer pair had a low WS against the β-proteobacterial *amoA* sequences and a high WS against the γ-proteobacterial AOB (and inversely for the 310.627 primer pair). The unAOB primer pair generally had a low WS against both β and γ AOB *amoA* sequences, indicating that they might be a good candidate as universal AOB primers. However, it is worth noticing that the unAOB primer pair generally had slightly higher WS against the β-proteobacterial *amoA* sequences compared to the BamoA primer pair and a high WS against the *Nitrosoglobus amoA* sequence (Figure 5.2). Furthermore, the unAOB primers generally had lower WS against the full-length pmoA compared to the BamoA primer pair, indicating that they could possibly generate unspecific amplification of the *pmoA* gene (Figure 5.3).

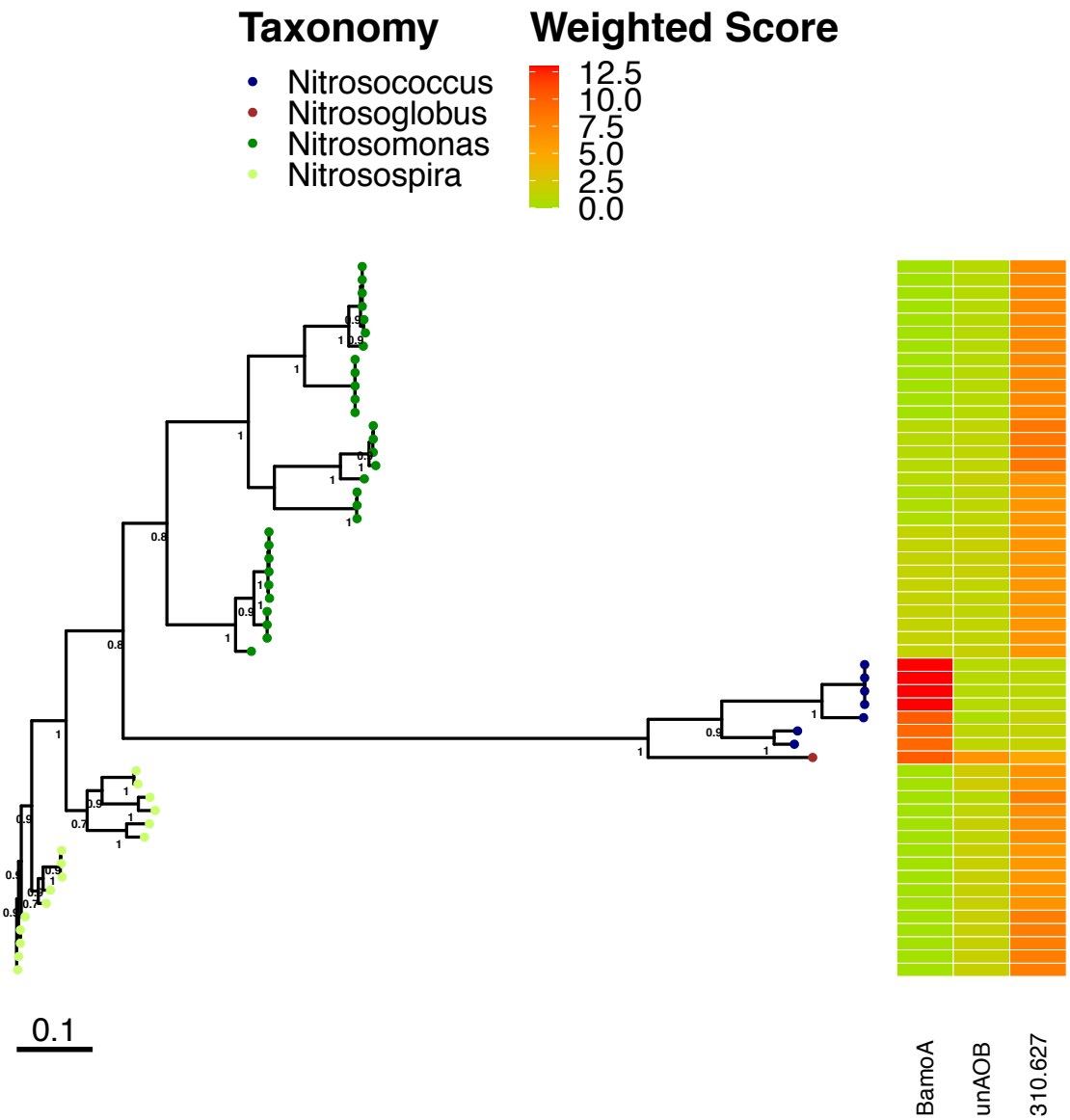


Figure 5.2 WS of the selected primers along the diversity of the AOB *amoA* sequences used for deNovo primer design. The Phylogenetic tree of the AOB *amoA* sequences is presented along with the WS of each primer as represented by the heatmap with green colours indicating lo WS and inversely for red colours. The Genus of the AOB *amoA* sequence is indicated by coloured points on the tips of the tree branches as indicated in the legend. Numbers on the tree represent bootstrap values (1000 repetitions).

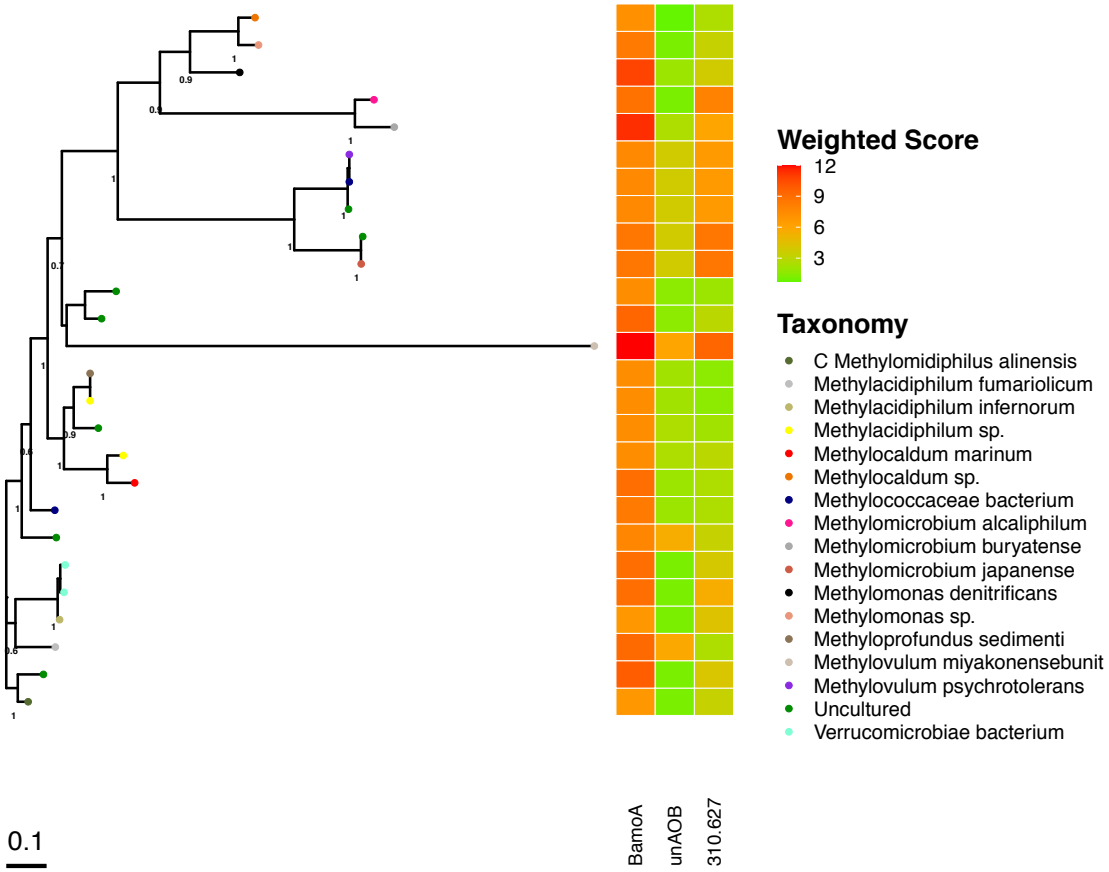


Figure 5.3 WS of the AOB primers against full-length pmoA sequences. The Phylogenetic tree of the *pmoA* sequences is presented along with the WS of each primer as represented by the heatmap with green colours indicating lo WS and inversely for red colours. The Genus and species of the *pmoA* sequence is indicated by coloured points on the tips of the tree branches as indicated in the legend. Numbers on the tree represent bootstrap values (1000 repetitions).

5.3.2 ASVs vs. OTUs

The Amplicon Sequence Variant (ASV) (or exact sequence variant) is a new method for clustering NGS reads that doesn't require the clustering of sequences based on a user-defined threshold, as is the case for OTU selection. Indeed, new methods can now control errors sufficiently such that sequences can be resolved at the single nucleotide difference level (Callahan *et al.*, 2017). This is particularly advantageous for the study of functional genes where the relationship between percentage identity of the gene and taxonomic assignment is unclear, unlike for the *16S rRNA* gene for which a 97% similarity threshold usually defines the difference between two species. By using the ASV method in chapter IV, a clustering between highly abundant but inactive and low abundance and highly active ASVs could be observed for AOB (Figure 4.16). The abundance of some ASVs in the highly active cluster was higher in runnels compared to ridges and correlated with the nitrification rates, indicating that they were the drivers of the differences in PNR. To determine if the same would be observed when

using the OTU method, the analysis was repeated but using OTU clustering (97% identity) rather than ASVs.

As shown in Figure 5.4, although two separate clusters could be distinguished in ridges ad runnels (one cluster containing mainly *Nitrosomonas* Group A and Group B and a second cluster containing mostly Unknown AOB), there was no evidence for one of these clusters being the more active one. Interestingly, 10 out of the 11 OTUs found to be more abundant in the runnels at RNA level (Figure 5.5) were also highly active in runnels (Figure 5.4 B). This would significantly change the conclusions drawn in chapter IV, as it would indicate that both *Nitrosomonas* group A and B and the unknown AOB are responsible for the nitrification process in the Montportail-Brouage mudflat. This could indicate that the OTU method has less resolution than the ASV method. In other words, the clustering of sequences based on a similarity threshold most likely combines together sequences that are in fact different, masking relevant features of the microbial community. Indeed, it had previously been shown hat the ASV method is more accurate and sensitive compared to the OTU approach for marker-gene analysis (Callahan *et al.*, 2017; Caruso *et al.*, 2019). Here we propose a new method to analyse functional gene sequencing data based on ASV construction, sequences filtering for size and clustering at the protein level. We show that this approach reveals more differences between libraries (e.g. DNA vs. RNA and ridges vs. runnels) and we therefore hypothesise that this method is more resolutive than the OTU approach. However, more research should be done to compare the two approaches in future, especially for functional gene analysis.

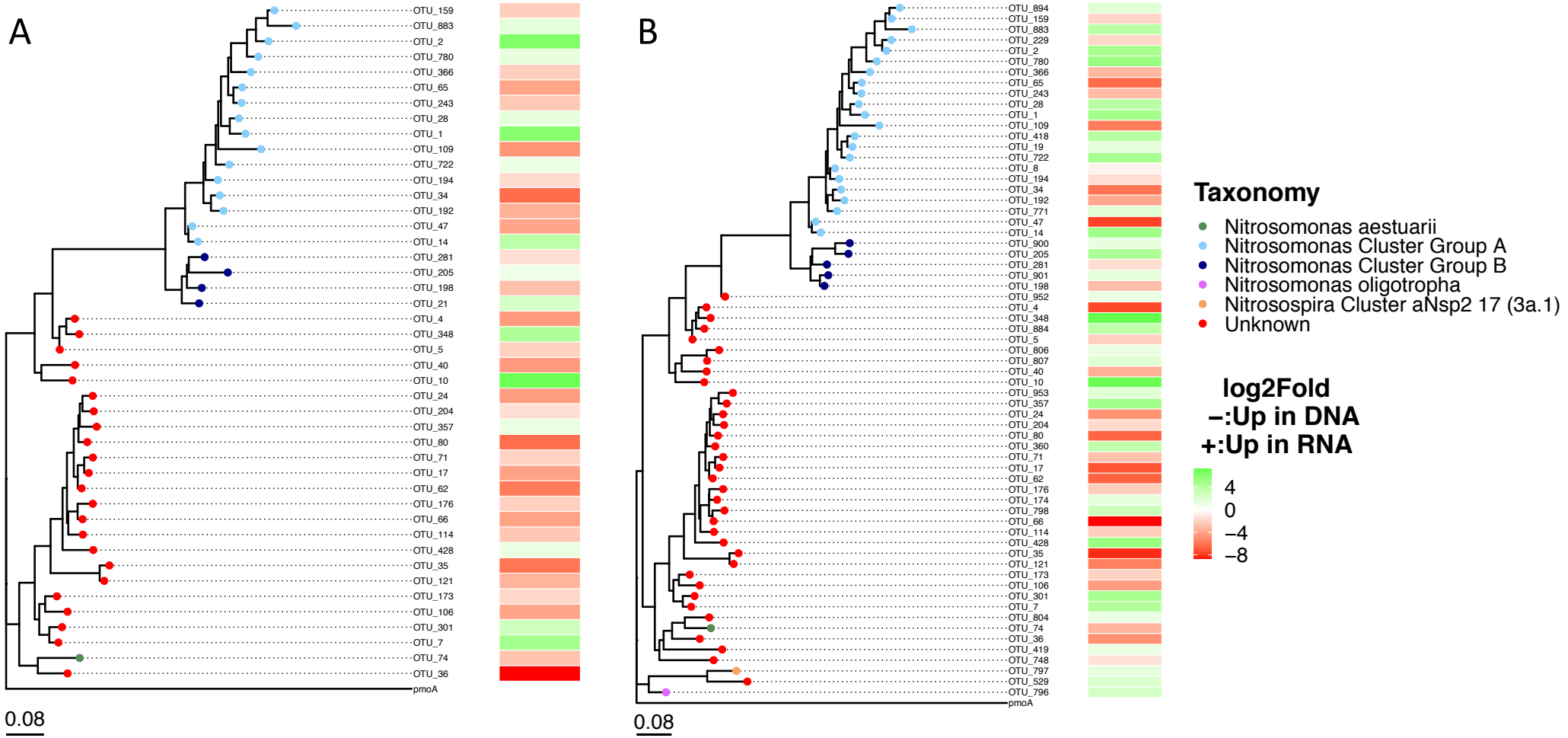


Figure 5.4 Identification of highly active AOB in ridges (A) and runnels (B) based on OTU construction. Colour of tips represents the taxonomy attributed to OTUs using the BLCA method. The heatmap represents the log2 fold change in abundance of individual OTU between the DNA and RNA libraries (only OTUs with significant differences in abundance between DNA and RNA libraries are represented; red indicates higher abundance in DNA and green indicates higher abundance in RNA). Phylogenetic tree was drawn using FastTree, from Maftt-aligned OTUs sequences and rooted using the particulate methane monooxygenase (*pmoA*) gene. Tree colouring and heatmap was done in R using ggtree.

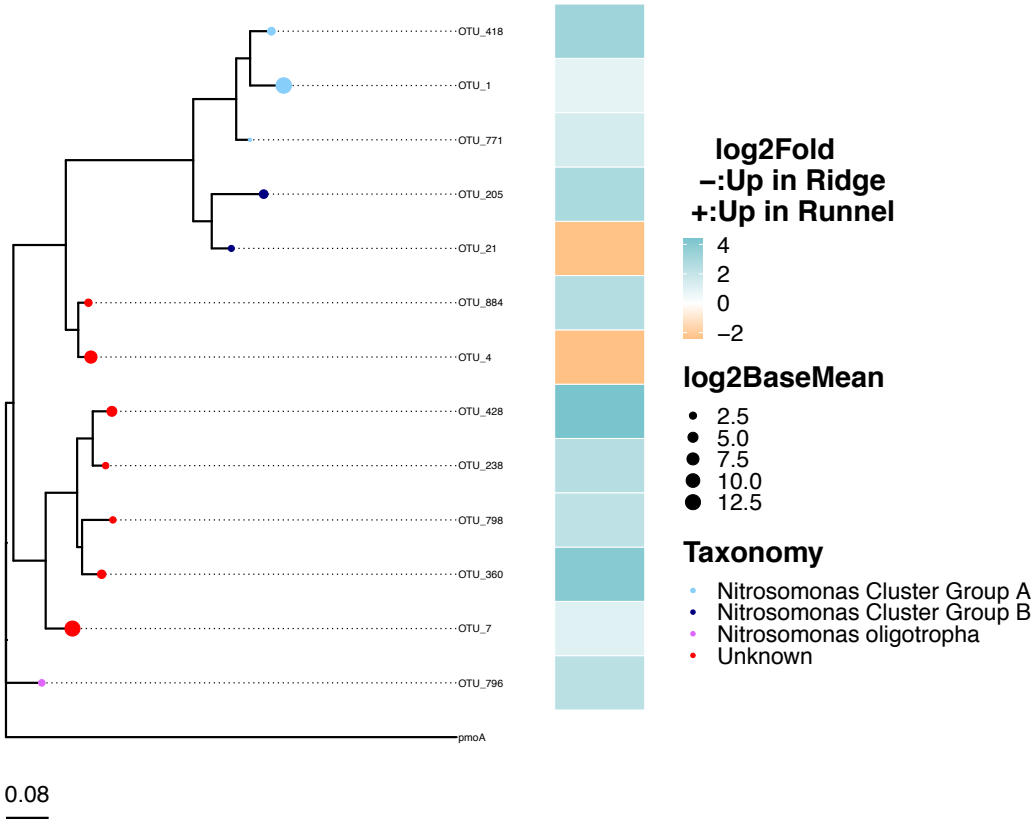


Figure 5.5 AOB OTUs differentially active between ridges and runnels. Colour of tips represents the taxonomy attributed to OTUs using the BLCA method. The heatmap represents the log2 fold change (*i.e.* fold change, log2 transformed) in activity (abundance in the RNA library) of individual OTU between ridges and runnels (only OTUs with significant differences are represented; blue indicates higher activity in runnels and orange indicates higher activity in ridges). Size of tips represents the mean activity (log2 transformed) between ridges and runnels. Phylogenetic tree was drawn using FastTree, from Maftt-aligned OTUs sequences and rooted using the particulate methane monooxygenase (*pmoA*) gene. Tree colouring and heatmap was done in R using ggtree.

5.3.3 Need for a combination of techniques to identify active nitrifiers

By measuring differences in AOB *amoA* gene and transcripts abundances between ridges and runnels, it was concluded that nitrification in the Montportail-Brouage mudflat is likely driven by a group of AOB that are low in abundance but highly active. The use of the BCLA method could not resolve the phylogeny of this group, which limits hypotheses that can be made from this observation. We also showed that the PCR-based methods limit our ability to target the full diversity of sequences in the environment (*Nitrosococcus* not currently targeted). This highlights the need for new techniques that would allow the direct isolation of active organisms from the environment.

The RAMAN-activated cell sorting method allows for the isolation of carbon fixing microorganisms directly from the environment. In this method, cells that incorporate heavy isotopes such as ^{13}C during autotrophic growth are separated from complex microbial communities. First, cells are separated from the matrix, like for flow cytometry, and passed on a microfluidic device and their single-cell RAMAN spectra are measured. Cells that have incorporated ^{13}C will display a shift in their spectra, which is then used to separate them from the rest of the microbial community (Wang *et al.*, 2013; McIlvenna *et al.*, 2016). This method has been previously used to isolate phototrophs from the Yellow Sea, China. In combination with metagenomics, it can constitute a formidable tool for reconstructing full genomes of active carbon fixers in the environment (Jing *et al.*, 2018). Since AOA and AOB grow autotrophically, it could in theory be possible to apply this method to isolate active nitrifiers, including AOM and NOB directly from the environment. This would constitute a major breakthrough for the discovery of the metabolic potential, microbe-microbe interactions and factors regulating the behaviour of nitrifiers in the environment.

The presence of active and inactive AOB in the Montportail-Brouage mudflat probably reflects differences in nitrifiers' physiologies that make them respond and adapt differently to environmental parameters. However, when using a targeted approach, as the sequencing of the *amoA* gene and transcript alone, it is impossible to identify what these differences might be, especially since the taxonomy of the active group could not be resolved. Untargeted approaches such as metagenomics, either from enrichment cultures or directly from environmental DNA, could help understand the genetic basis for these differences by reconstructing full genomes from representatives of the active and inactive clusters. Previously, the use of metagenomics has greatly helped expand knowledge on AOM and NOB by discovering the genetic potential to use alternative substrates as sources of energy such as the urease genes in *Nitrosospira multiformis* (Norton *et al.*, 2008), *Nitrosococcus oceani* (Klotz *et al.*, 2006) and the cyanase genes in the AOA *Nitrososphaera gargensis* and in NOB (Palatinszky, Herbold, Jehmlich, Pogoda, Han, Von Bergen, *et al.*, 2015).

5.3.4 Ecological and environmental implications

In chapter IV, strong differences in potential nitrification rates were detected between ridges and runnels. These observations would need to be validated by measuring actual *in situ* rates. Whether small-scale variations in sediment morphology should be taken into account when

calculating nutrient-processing rates is important as it would affect global budgets. Beyond nitrification and the amount of ammonia that is transferred from land to sea, the effect of sediment structures on nitrogen processing rates is important to understand to accurately predict the amount of nitrogen oxides produced by estuaries. As discussed previously, nitrogen oxides (NO_x) represent a global challenge in the context of climate change and, in natural environments, their production and consumption are mainly mediated by microorganisms. Are structures like runnels- where nitrification and potentially coupled nitrification-denitrification enhanced- hotspots for NO_x production? Or, alternatively, are ridges producing more NO_x than runnels because of oxygen limitation? What is the difference in production and consumption of NO_x in ridges and runnels compared to mudflats without such structures? These questions are some examples of the ecological implications of the findings of this thesis and should be further investigated in future.

Results from this thesis also have important ecological implications. Indeed, we also showed that AOB were more abundant in ridges while PNR was higher in runnels and hypothesise that the majority of AOB quantified at DNA level were actually of low activity. Alternatively, could this observation indicate that AOB are active and growing in ridges but simply "doing something else"? Whether AOB are able of heterotrophic growth in the environment is still under debate but results from pure cultures (Abeliovich *et al.*, 1992; Hommes *et al.*, 2003; Schmidt, 2009) and identification of genes involved in heterotrophic growth in AOM genomes suggest they might (Arp *et al.*, 2007; Prosser and Nicol, 2012). Could it therefore be that the low-activity AOB cluster (*Nitrosomonas* Cluster group A and B) are only facultative autotrophs and are also able to gain energy and grow on organic carbon, while the high-activity-unknown AOB cluster are obligate autotrophs? The ability of the AOB to use pyruvate as a carbon and energy source would theoretically increase their maximum growth yield from 0.13-0.16 gBio/gN-NH_3 (grams of biomass per gram of ammonia nitrogen) to 0.35-0.61 gBio/gN-NH_3 (González-Cabaleiro *et al.*, 2019). Could this higher growth yield explain the higher abundance of AOB in ridges? If this is indeed the case, it would seriously put into question the utility of measuring *amoA* gene abundance alone and it's correlation with PNR as this assumes that nitrification is the only energy source in AOM. This further illustrates the utility of transcriptomics when studying microbial function in the environment. The use of SIP would help identify the AOM clusters that actively fix CO_2 in this environment and therefore grow autotrophically; by incubating ridges and runnels samples under nitrifying conditions (presence of ammonia; oxygenation), would it be possible to show the active cluster more labelled than

the inactive one, therefore showing that they are the main autotrophic nitrifiers? With the differential abundance and activity of AOB clusters in close proximity, the Montportail-Brouage mudflat would be an ideal environment to test these hypotheses in future.

References

Abeliovich, A., Vonshak, A., Unit, E.M., and Boker, S. (1992) Anaerobic metabolism of *Nitrosomonas europaea*. *Microbiology* **158**: 267–270.

Abell, G.C.J., Banks, J., Ross, D.J., Keane, J.P., Robert, S.S., Revill, A.T., and Volkman, J.K. (2011) Effects of estuarine sediment hypoxia on nitrogen fluxes and ammonia oxidizer gene transcription. *FEMS Microbiol. Ecol.* **75**: 111–122.

Agardy, T., Alder, J., Dayton, P., Curran, S., Kitchingman, A., Wilson, M., et al. (2005) Coastal Systems. *Ecosyst. Hum. Well-being Curr. Status Trends* 513–550.

Agogué, H., Brink, M., Dinasquet, J., and Herndl, G.J. (2008) Major gradients in putatively nitrifying and non-nitrifying Archaea in the deep North Atlantic. *Nature* **456**: 788–91.

Aigle, A., Gubry-Rangin, C., Thion, C., Estera-Molina, K.Y., Richmond, H., Pett-Ridge, J., et al. (2020) Experimental testing of hypotheses for temperature- and pH-based niche specialization of ammonia oxidizing archaea and bacteria. *Environ. Microbiol.* **22**: 4032–4045.

Aigle, A., Prosser, J.I., and Gubry-Rangin, C. (2019) The application of high-throughput sequencing technology to analysis of amoA phylogeny and environmental niche specialisation of terrestrial bacterial ammonia-oxidisers. *Environ. Microbiome* **14**: 3.

Altschul, S F, Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–10.

Altschul, Stephen F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic Local Alignment Search Tool. 403–410.

Alves, R.J.E., Minh, B.Q., Urich, T., von Haeseler, A., and Schleper, C. (2018) Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on amoA genes. *Nat. Commun.* **9**: 1517.

Alzerreca, J.J., Norton, J.M., and Klotz, M.G. (1999) The amo operon in marine, ammonia-oxidizing G-proteobacteria. *FEMS Microbiol. Lett.* **180**: 21–29.

An, S. and Joye, S.B. (2001) Enhancement of coupled nitrification-denitrification by benthic photosynthesis in shallow estuarine sediments. *Limnol. Oceanogr.* **46**: 62–74.

Arp, D.J. and Bottomley, P.J. (2006) Nitrifiers: More than 100 years from isolation to genome sequences . *Microbe* **1**: 229–234.

Arp, D.J., Chain, P.S.G., and Klotz, M.G. (2007) The impact of genome analyses on our understanding of ammonia-oxidizing bacteria. *Annu. Rev. Microbiol.* **61**: 503–528.

Arp, D.J., Sayavedra-Soto, L.A., and Hommes, N.G. (2002) Molecular biology and biochemistry of ammonia oxidation by Nitrosomonas europaea. *Arch. Microbiol.* **178**: 250–255.

Arrigo, K.R. (2005) Marine microorganisms and global nutrient cycles. *Nature* **437**: 343–348.

Baggs, E.M. (2008) A review of stable isotope techniques for N2O source partitioning in soils: Recent progress, remaining challenges and future considerations. In, *Rapid Communications in Mass Spectrometry*., pp. 1664–1672.

Baker, B.J., Lazar, C.S., Teske, A.P., and Dick, G.J. (2015) Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome* **3**: 1–12.

Bartlett, R., Mortimer, R.J.G., and Morris, K. (2008) Anoxic nitrification: Evidence from Humber Estuary sediments (UK). *Chem. Geol.* **250**: 29–39.

Bartossek, R., Nicol, G.W., Lanzen, A., Klenk, H.P., and Schleper, C. (2010) Homologues of nitrite reductases in ammonia-oxidizing archaea: Diversity and genomic context. *Environ. Microbiol.* **12**: 1075–1088.

Bayer, B., Vojvoda, J., Offre, P., Alves, R.J.E., Elisabeth, N.H., Garcia, J.A.L., et al. (2016) Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. *ISME J.* **10**: 1051–1063.

Beaumont, H.J.E., Lens, S.I., Reijnders, W.N.M., Westerhoff, H. V., and Van Spanning, R.J.M. (2004) Expression of nitrite reductase in Nitrosomonas europaea involves NsrR, a novel nitrite-sensitive transcription repressor. *Mol. Microbiol.* **54**: 148–158.

Bejarano-Ortiz, D.I., Huerta-Ochoa, S., Thalasso, F., Cuervo-López, F. de M., and Texier, A.C. (2015) Kinetic Constants for Biological Ammonium and Nitrite Oxidation Processes Under Sulfide Inhibition. *Appl. Biochem. Biotechnol.* **177**: 1665–1675.

Beman, J.M., Chow, C.E., King, A.L., Feng, Y., Fuhrman, J.A., Andersson, A., et al. (2011) Global declines in oceanic nitrification rates as a consequence of ocean acidification. *Proc. Natl. Acad. Sci. U. S. A.* **108**: 208–213.

Berg, C., Listmann, L., Vandieken, V., Vogts, A., and Jürgens, K. (2015) Chemoautotrophic growth of ammonia-oxidizing Thaumarchaeota enriched from a pelagic redox gradient in the Baltic Sea. *Front. Microbiol.* **6**: 1–10.

Bernhard, A.E., Chelsky, A., Giblin, A.E., and Roberts, B.J. (2019) Influence of local and regional drivers on spatial and temporal variation of ammonia-oxidizing communities in Gulf of Mexico salt marshes. *Environ. Microbiol. Rep.* **11**: 825–834.

Bernhard, A.E., Landry, Z.C., Blevins, A., De La Torre, J.R., Giblin, A.E., and Stahl, D.A. (2010) Abundance of ammonia-oxidizing archaea and bacteria along an estuarine salinity gradient in relation to potential nitrification rates. *Appl. Environ. Microbiol.* **76**: 1285–1289.

Berounsky, V.M. and Nixon, S.W. (1990) Temperature and annual cycle of nitrification in waters of Narragansett Bay. *Limnol. Oceanogr.* **35**: 1610–1617.

Bianchi, T.S. (2007) Biogeochemistry of Estuaries.

Biswas, B. and Gresshoff, P.M. (2014) The role of symbiotic nitrogen fixation in sustainable production of biofuels. *Int. J. Mol. Sci.* **15**: 7380–7397.

Björkman, J., Švec, D., Lott, E., Kubista, M., and Sjöback, R. (2016) Differential amplicons (δAmp)-a new molecular method to assess RNA integrity. *Biomol. Detect. Quantif.* **6**: 4–12.

Blainey, P.C., Mosier, A.C., Potanina, A., Francis, C.A., and Quake, S.R. (2011) Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS One* **6**:.

Blazewicz, S.J., Barnard, R.L., Daly, R.A., and Firestone, M.K. (2013) Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME J.* **7**: 2061–2068.

Boatman, C.D. and Murray, W. (1982) Modeling exchangeable NH4+ adsorption in marine sediments: Process and controls of adsorption. *Limnol. Oceanogr.* **27**: 99–110.

Botes, M., De Kwaadsteniet, M., and Cloete, T.E. (2013) Application of

quantitative PCR for the detection of microorganisms in water. *Anal. Bioanal. Chem.* **405**: 91–108.

Bower, C.E. and Bidwell, J.P. (1978) Ionization of ammonia in seawater: effects of temperature, pH and, salinity. *J. Fish. Res. Board Can* **35**: 1012–1016.

Breider, F., Yoshikawa, C., Makabe, A., Toyoda, S., Wakita, M., Matsui, Y., et al. (2019) Response of N2O production rate to ocean acidification in the western North Pacific. *Nat. Clim. Chang.* **9**: 954–958.

Brisco, M.J. and Morley, A.A. (2012) Quantification of RNA integrity and its use for measurement of transcript number. *Nucleic Acids Res.* **40**:.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008) Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.* **6**: 245–252.

Burgin, A.J., Yang, W.H., Hamilton, S.K., and Silver, W.L. (2011) Beyond carbon and nitrogen: How the microbial energy economy couples elemental cycles in diverse ecosystems. *Front. Ecol. Environ.* **9**: 44–52.

Burton, S.A.Q. and Prosser, J.I.M.I. (2001) Dynamical system-Guckenheimer. *Appl. Environ. Microbiol.* **67**: 2952–2957.

Bustin, S. and Nolan, T. (2017) Talking the talk, but not walking the walk: RT-qPCR as a paradigm for the lack of reproducibility in molecular research. *Eur. J. Clin. Invest.* **47**: 756–774.

Bustin, S.A. (2002) Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): Trends and problems. *J. Mol. Endocrinol.* **29**: 23–39.

Bustin, S.A. and Nolan, T. (2004) Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J. Biomol. Tech.* **15**: 155–66.

Caffrey, Jane M, Bano, N., Kalanetra, K., and Hollibaugh, J.T. (2007) Ammonia oxidation and ammonia-oxidizing bacteria and archaea from estuaries with differing histories of hypoxia. *ISME J.* **1**: 660–662.

Caffrey, Jane M., Bano, N., Kalanetra, K., and Hollibaugh, J.T. (2007) Ammonia oxidation and ammonia-oxidizing bacteria and archaea from estuaries with differing histories of hypoxia. *ISME J.* **1**: 660–662.

Callahan, B.J., McMurdie, P.J., and Holmes, S.P. (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **11**: 2639–2643.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016a) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**: 581–583.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016b) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**: 581–583.

Campbell, M.A., Chain, P.S.G., Dang, H., El Sheikh, A.F., Norton, J.M., Ward, N.L., et al. (2011) Nitrosococcus watsonii sp. nov., a new species of marine obligate ammonia-oxidizing bacteria that is not omnipresent in the world's oceans: Calls to validate the names "Nitrosococcus halophilus" and "Nitrosomonas mobilis." *FEMS Microbiol. Ecol.* **76**: 39–48.

Canfield, D.E., Glazer, A.N., and Falkowski, P.G. (2010) The Evolution and Future of Earth ' s Nitrogen Cycle. *Science (80-.).* **330**: 192–196.

Cao, H., Hong, Y., Li, M., and Gu, J.D. (2011) Diversity and abundance of ammonia-oxidizing prokaryotes in sediments from the coastal Pearl River estuary to the South China Sea. *Antonie van Leeuwenhoek, Int. J.*

Gen. Mol. Microbiol. **100**: 545–556.

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

Caranto, J.D. and Lancaster, K.M. (2017) Nitric oxide is an obligate bacterial nitrification intermediate produced by hydroxylamine oxidoreductase. *Proc. Natl. Acad. Sci.* 201704504.

Caruso, V., Song, X., Asquith, M., and Karstens, L. (2019) Performance of Microbiome Sequence Inference Methods in Environments with Varying Biomass. *mSystems* **4**: 1–19.

Carvalhais, L.C., Dennis, P.G., Tyson, G.W., and Schenk, P.M. (2012) Application of metatranscriptomics to soil environments. *J. Microbiol. Methods* **91**: 246–251.

Castignetti, D. and Hollocher, T.C. (1984) Heterotrophic nitrification among denitrifiers. *Appl. Environ. Microbiol.* **47**: 620–623.

Cavicchioli, R., Ripple, W.J., Timmis, K.N., Azam, F., Bakken, L.R., Baylis, M., et al. (2019) Scientists' warning to humanity: microorganisms and climate change. *Nat. Rev. Microbiol.* **17**: 569–586.

Chang, Q., Luan, Y., and Sun, F. (2011) Variance adjusted weighted UniFrac: a powerful beta diversity measure for.

Chapin, F.S., McFarland, J., David McGuire, A., Euskirchen, E.S., Ruess, R.W., and Kielland, K. (2009) The changing global carbon cycle: Linking plant-soil carbon dynamics to global consequences. *J. Ecol.* **97**: 840–850.

Charif, D. and Lobry, J.R. (2007) SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis Delphine. In, *Structural approaches to sequence evolution: Molecules, networks, populations*., pp. 207–232.

Chen, X.P., Zhu, Y.G., Xia, Y., Shen, J.P., and He, J.Z. (2008) Ammonia-oxidizing archaea: Important players in paddy rhizosphere soil? *Environ. Microbiol.* **10**: 1978–1987.

Choi, O., Das, A., Yu, C.P., and Hu, Z. (2010) Nitrifying bacterial growth inhibition in the presence of algae and cyanobacteria. *Biotechnol. Bioeng.* **107**: 1004–1011.

Cholet, F., Ijaz, U.Z., and Smith, C.J. (2019) Differential ratio amplicons (R amp) for the evaluation of RNA integrity extracted from complex environmental samples. *Environ. Microbiol.* **21**: 827–844.

Cholet, F., Ijaz, U.Z., and Smith, C.J. (2020) Reverse transcriptase enzyme and priming strategy affect quantification and diversity of environmental transcripts. *Environ. Microbiol.* **22**: 2383–2402.

Christiansen, N.A., Fryirs, K.A., Green, T.J., and Hose, G.C. (2019) The impact of urbanisation on community structure, gene abundance and transcription rates of microbes in upland swamps of Eastern Australia. *PLoS One* **14**: 1–20.

Costa, E., Pérez, J., and Kreft, J.U. (2006) Why is metabolic labour divided in nitrification? *Trends Microbiol.* **14**: 213–219.

Costanza, R., de Groot, R., Sutton, P., van der Ploeg, S., Anderson, S.J., Kubiszewski, I., et al. (2014) Changes in the global value of ecosystem services. *Glob. Environ. Chang.* **26**: 152–158.

Crutzen, P.J. (1970) The influence of nitrogen oxides on the atmospheric ozone content. *Q. J. R. Meteorol. Soc.* **96**: 320–325.

D'Amore, R., Ijaz, U.Z., Schirmer, M., Kenny, J.G., Gregory, R., Darby, A.C., et al. (2016) A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA

community profiling. *BMC Genomics* **17**:.

Daims, H., Lebedeva, E.V., Pjevac, P., Han, P., Herbold, C., Albertsen, M., et al. (2015) Complete nitrification by Nitrospira bacteria. *Nature* **528**: 504–509.

Daims, H., Lücker, S., and Wagner, M. (2016) A New Perspective on Microbes Formerly Known as Nitrite-Oxidizing Bacteria. *Trends Microbiol.* **24**: 699–712.

Damashek, J. and Francis, C.A. (2018) Microbial Nitrogen Cycling in Estuaries: From Genes to Ecosystem Processes. *Estuaries and Coasts* **41**: 626–660.

Damashek, J., Smith, J.M., Mosier, A.C., and Francis, C.A. (2015) Benthic ammonia oxidizers differ in community structure and biogeochemical potential across a riverine delta. *Front. Microbiol.* **6**: 1–18.

Delgado Vela, J., Dick, G.J., and Love, N.G. (2018) Sulfide inhibition of nitrite oxidation in activated sludge depends on microbial community composition. *Water Res.* **138**: 241–249.

Di, H.J., Cameron, K.C., Shen, J.P., Winefield, C.S., O'Callaghan, M., Bowatte, S., and He, J.Z. (2010) Ammonia-oxidizing bacteria and archaea grow under contrasting soil nitrogen conditions. *FEMS Microbiol. Ecol.* **72**: 386–394.

Die, J. V and Román, B. (2012) RNA quality assessment: a view from plant qPCR studies. *J. Exp. Bot.* **63**: 6069–6077.

Dimitri Kits, K., Sedlacek, C.J., Lebedeva, E. V., Han, P., Bulaev, A., Pjevac, P., et al. (2017) Kinetic analysis of a complete nitrifier reveals an oligotrophic lifestyle. *Nature* **549**: 269–272.

Dissanayake, N.G., Frid, C.L.J., Drylie, T.P., and Caswell, B.A. (2018) Ecological functioning of mudflats: Global analysis reveals both regional differences and widespread conservation of functioning. *Mar. Ecol. Prog. Ser.* **604**: 1–20.

Dohra, H., Arai, K., Urakawa, H., and Fujiwara, T. (2019) Draft Genome Sequence of Nitrosococcus oceani Strain NS58, a Marine Ammonia-Oxidizing Gammaproteobacterium Isolated from Tokyo Bay Sediment. *Microbiol Resour Announc* **8**: 29–31.

Dollhopf, S.L., Hyun, J., Smith, A.C., Adams, H.J., Brien, S.O., and Kostka, J.E. (2005) Quantification of Ammonia-Oxidizing Bacteria and Factors Controlling Nitrification in Salt Marsh Sediments. *Appl. Environ. Microbiol.* **71**: 240–246.

Douglas, A.E. (1998) Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria Buchnera. *Annu. Rev. Entomol.* **43**: 17–37.

Duff, A.M., Zhang, L.M., and Smith, C.J. (2017) Small-scale variation of ammonia oxidisers within intertidal sediments dominated by ammonia-oxidising bacteria Nitrosomonas sp. amoA genes and transcripts. *Sci. Rep.* **7**: 1–13.

Dutta, H. and Dutta, A. (2016) The microbial aspect of climate change. *Energy, Ecol. Environ.* **1**: 209–232.

Dyer, K.R., Christie, M.C., and Wright, E.W. (2000) The classification of intertidal mudflats. *Cont. Shelf Res.* **20**: 1039–1060.

Ehrich, S., Behrens, D., Lebedeva, E. V, Ludwig, W., and Bock, E. (1995) A New Obligately Chemolithoautotrophic, Nitrite-Oxidizing Bacterium, Nitrospira-Moscoviensis Sp-Nov and Its Phylogenetic Relationship. *Arch. Microbiol.* **164**: 16–23.

Eisenmann, E., Beuerle, J., Sulger, K., Kroneck, P.M.H., and Schumacher, W. (1995) Lithotrophic growth of Sulfurospirillum deleyianum with

sulfide as electron donor coupled to respiratory reduction of nitrate to ammonia. *Arch. Microbiol.* **164**: 180–185.

Elrod, M.J. (1999) Greenhouse Warming Potentials from the Infrared Spectroscopy of Atmospheric Gases. *J. Chem. Educ.* **76**: 1702.

Erguder, T.H., Boon, N., Wittebolle, L., Marzorati, M., and Verstraete, W. (2009) Environmental factors shaping the ecological niches of ammonia-oxidizing archaea. *FEMS Microbiol. Rev.* **33**: 855–869.

Eriksson, P.G., Svensson, J.M., and Carrer, G.M. (2003) Temporal changes and spatial variation of soil oxygen consumption, nitrification and denitrification rates in a tidal salt marsh of the Lagoon of Venice, Italy. *Estuar. Coast. Shelf Sci.* **58**: 861–871.

Ettwig, K.F., Butler, M.K., Le Paslier, D., Pelletier, E., Mangenot, S., Kuypers, M.M.M., et al. (2010) Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* **464**: 543–548.

Feng, G., Sun, W., Zhang, F., Karthik, L., and Li, Z. (2016) Inhabitancy of active Nitrosopumilus-like ammonia-oxidizing archaea and Nitrospira nitrite-oxidizing bacteria in the sponge Theonella swinhoei. *Sci. Rep.* **6**: 1–11.

Feng, G., Sun, W., Zhang, F., Orlić, S., and Li, Z. (2018) Functional Transcripts Indicate Phylogenetically Diverse Active Ammonia-Scavenging Microbiota in Sympatric Sponges. *Mar. Biotechnol.* **20**: 131–143.

Ferry, J.G. and Lessner, D.J. (2008) Methanogenesis in marine sediments. *Ann. N. Y. Acad. Sci.* **1125**: 147–157.

Firestone, M.. and Davidson, E.. (1989) Microbial Basis of NO and N2O production and Consumption in soil. In, *Exchange of Trace Gases Between Terestrial Ecosystems and the Athmosphere*., pp. 7–21.

Fleige, S. and Pfaffl, M.W. (2006) RNA integrity and the effect on the real-time qRT-PCR performance. *Mol. Aspects Med.* **27**: 126–139.

Forootan, A., Sjöback, R., Björkman, J., Sjögreen, B., Linz, L., and Kubista, M. (2017) Methods to determine limit of detection and limit of quantification in quantitative real-time PCR (qPCR). *Biomol. Detect. Quantif.* **12**: 1–6.

Foster, N.M., Hudson, M.D., Bray, S., and Nicholls, R.J. (2013) Intertidal mudflat and saltmarsh conservation and sustainable use in the UK: A review. *J. Environ. Manage.* **126**: 96–104.

Francis, C.A., Roberts, K.J., Beman, J.M., Santoro, A.E., and Oakley, B.B. (2005) Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 14683–14688.

Frankland, P. and Frankland, G. (1890) II. The nitrifying process and its specific ferment. In, *The nitrifying process and its specific ferment. Part I. Philos T Roy Soc B*., pp. 107–28.

Freitag, T.E., Chang, L., and Prosser, J.I. (2006) Changes in the community structure and activity of betaproteobacterial ammonia-oxidizing sediment bacteria along a freshwater-marine gradient. *Environ. Microbiol.* **8**: 684–696.

Fuhrman, J. a (2009) Microbial community structure and its functional implications. *Nature* **459**: 193–9.

Gadkar, V.J. and Filion, M. (2013) Quantitative real-time polymerase chain reaction for tracking microbial gene expression in complex environmental matrices. *Curr. Issues Mol. Biol.* **15**: 45–58.

Gao, H., Yang, Z.K., Gentry, T.J., Wu, L., Schadt, C.W., and Zhou, J. (2007) Microarray-based analysis of microbial community RNAs by whole-community RNA

amplification. *Appl. Environ. Microbiol.* **73**: 563–571.

Gao, J., Hou, L., Zheng, Y., Liu, M., Yin, G., Yu, C., and Gao, D. (2018) Shifts in the Community Dynamics and Activity of Ammonia-Oxidizing Prokaryotes Along the Yangtze Estuarine Salinity Gradient. *J. Geophys. Res. Biogeosciences* **123**: 3458–3469.

Gao, X., Lin, H., Revanna, K., and Dong, Q. (2017) A Bayesian taxonomic classification method for 16S rRNA gene sequences with improved species-level accuracy. *BMC Bioinformatics* **18**: 1–10.

Garbeva, P., Baggs, E.M., and Prosser, J.I. (2007) Phylogeny of nitrite reductase (nirK) and nitric oxide reductase (norB) genes from Nitrosospira species isolated from soil. *FEMS Microbiol. Lett.* **266**: 83–89.

Giblin, A., Tobias, C., Song, B., Weston, N., Banta, G., and Rivera-Monroy, V. (2013) The Importance of Dissimilatory Nitrate Reduction to Ammonium (DNRA) in the Nitrogen Cycle of Coastal Ecosystems. *Oceanography* **26**: 124–131.

Glass, J.B. and Orphan, V.J. (2012) Trace metal requirements for microbial enzymes involved in the production and consumption of methane and nitrous oxide. *Front. Microbiol.* **3**: 1–20.

Godbold, J.A. and Calosi, P. (2013) Ocean acidification and climate change: Advances in ecology and evolution. *Philos. Trans. R. Soc. B Biol. Sci.* **368**:.

Gonçalves, J., Gutiérrez-Aguirre, I., Balasubramanian, M.N., Zagorščak, M., Ravnikar, M., and Turk, V. (2018) Surveillance of human enteric viruses in coastal waters using concentration with methacrylate monolithic supports prior to detection by RT-qPCR. *Mar. Pollut. Bull.* **128**: 307–317.

González-Cabaleiro, R., Curtis, T.P., and Ofiţeru, I.D. (2019) Bioenergetics analysis of ammonia-oxidizing bacteria and the estimation of their maximum growth yield. *Water Res.* **154**: 238–245.

Gouleau, D., Jouanneau, J.M., Weber, O., and Sauriau, P.G. (2000) Short- and long-term sedimentation on Montportail-Brouage intertidal mudflat, Marennes-Oleron Bay (France). *Cont. Shelf Res.* **20**: 1513–1530.

Graham, J.E., Wantland, N.B., Campbell, M., and Klotz, M.G. (2011) Characterizing bacterial gene expression in nitrogen cycle metabolism with RT-qPCR 1st ed. Elsevier Inc.

Griffin, B.M., Schott, J., and Schink, B. (2007) Nitrite, an electron donor for anoxygenic photosynthesis. *Science* **316**: 1870.

Griffiths, R.I., Whiteley, A.S., O'Donnell, A.G., and Bailey, M.J. (2000) Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl. Environ. Microbiol.* **66**: 5488–5491.

Groeneweg, J., Sellner, B., and Tappe, W. (1994) Ammonia oxidation in nitrosomonas at NH3 concentrations near km: Effects of pH and temperature. *Water Res.* **28**: 2561–2566.

van Groenigen, J.W., Huygens, D., Boeckx, P., Kuyper, T.W., Lubbers, I.M., Rütting, T., and Groffman, P.M. (2015) The soil N cycle: new insights and key challenges. *Soil* **1**: 235–256.

Gruber, N. and Galloway, J.N. (2008) An Earth-system perspective of the global nitrogen cycle. *Nature* **451**: 293–6.

Gubry-Rangin, C., Nicol, G.W., and Prosser, J.I. (2010) Archaea rather than bacteria control nitrification in two agricultural acidic soils. *FEMS*

Microbiol. Ecol. **74**: 566–574.

Gutleben, J., Chaib De Mares, M., van Elsas, J.D., Smidt, H., Overmann, J., and Sipkema, D. (2018) The multi-omics promise in context: from sequence to microbial isolate. *Crit. Rev. Microbiol.* **44**: 212–229.

Gworek, B., Bemowska-Ka??abun, O., Kijeńska, M., and Wrzosek-Jakubowska, J. (2016) Mercury in Marine and Oceanic Waters - a Review. *Water. Air. Soil Pollut.* **227**:.

Hall-Spencer, J.M. and Harvey, B.P. (2019) Ocean acidification impacts on coastal ecosystem services due to habitat degradation. *Emerg. Top. Life Sci.* **3**: 197–206.

Hallin, S., Jones, C.M., Schloter, M., and Philippot, L. (2009) Relationship between n-cycling communities and ecosystem functioning in a 50-year-old fertilization experiment. *ISME J.* **3**: 597–605.

Happel, E., Bartl, I., Voss, M., and Riemann, L. (2018) Extensive nitrification and active ammonia oxidizers in two contrasting coastal systems of the Baltic Sea. *Environ. Microbiol.* **20**: 2913–2926.

Haroon, M.F., Hu, S., Shi, Y., Imelfort, M., Keller, J., Hugenholtz, P., et al. (2013) Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500**: 567–570.

Harse, G. a. (2011) Plastic, the Great Pacific Garbage Patch, and international misfires at a cure. *UCLA J. Environ. Law Policy* **29.2**: 331.

Hata, A., Katayama, H., and Furumai, H. (2015) Organic substances interfere with reverse transcription-quantitative PCR-based virus detection in water samples. *Appl. Environ. Microbiol.* **81**: 1585–1593.

Hättenschwiler, S. and Vitousek, P.M. (2000) The role of polyphenols in terrestrial ecoystems nutrient cycling. *Tree* **15**: 238–243.

Haubois, A.G., Sylvestre, F., Guarini, J.M., Richard, P., and Blanchard, G.F. (2005) Spatio-temporal structure of the epipelic diatom assemblage from an intertidal mudflat in Marennes-Oléron Bay, France. *Estuar. Coast. Shelf Sci.* **64**: 385–394.

Hayatsu, M., Tago, K., and Saito, M. (2008) Various players in the nitrogen cycle: Diversity and functions of the microorganisms involved in nitrification and denitrification. *Soil Sci. Plant Nutr.* **54**: 33–45.

Hayatsu, M., Tago, K., Uchiyama, I., Toyoda, A., Wang, Y., Shimomura, Y., et al. (2017) An acid-tolerant ammonia-oxidizing γ-proteobacterium from soil. *ISME J.* **11**: 1130–1141.

He, H., Zhen, Y., Mi, T., Fu, L., and Yu, Z. (2018) Ammonia-oxidizing archaea and bacteria differentially contribute to ammonia oxidation in sediments from adjacent waters of Rushan Bay, China. *Front. Microbiol.* **9**: 1–14.

Hemp, J., Lücker, S., Schott, J., Pace, L. a, Johnson, J.E., Schink, B., et al. (2016) Genomics of a phototrophic nitrite oxidizer: insights into the evolution of photosynthesis and nitrification. *ISME J.* 1–10.

Herbold, C.W., Lehtovirta-Morley, L.E., Jung, M.Y., Jehmlich, N., Hausmann, B., Han, P., et al. (2017) Ammonia-oxidising archaea living at low pH: Insights from comparative genomics. *Environ. Microbiol.* **19**: 4939–4952.

Hink, L., Gubry-Rangin, C., Nicol, G.W., and Prosser, J.I. (2018) The consequences of niche and physiological differentiation of archaeal and bacterial ammonia oxidisers for nitrous oxide emissions. *ISME J.* **12**: 1084–1093.

Hink, L., Lycus, P., Gubry-rangin, C., Frostegård, Å., Nicol, G.W., Prosser, J.I., et al. (2017) Kinetics of NH3-oxidation, NO-turnover, N2O-production and electron flow during oxygen depletion in model bacterial

and archaeal ammonia oxidisers. *Env. Microbiol* **19**: 1–42.

Hink, L., Nicol, G.W., and Prosser, J.I. (2017) Archaea produce lower yields of N2O than bacteria during aerobic ammonia oxidation in soil. *Environ. Microbiol.* **19**: 4829–4837.

Hoernes, T.P., Clementi, N., Faserl, K., Glasner, H., Breuker, K., Lindner, H., et al. (2016) Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. *Nucleic Acids Res.* **44**: 852–862.

Hollibaugh, J.T., Gifford, S., Sharma, S., Bano, N., and Moran, M.A. (2011) Metatranscriptomic analysis of ammonia-oxidizing organisms in an estuarine bacterioplankton assemblage. *ISME J.* **5**: 866–78.

Hommes, N.G., Sayavedra-Soto, L.A., and Arp, D.J. (2003) Chemolithoorganotrophic Growth of Nitrosomonas europaea on Fructose. *J. Bacteriol.* **185**: 6809–6814.

Hommes, N.G., Sayavedra-Soto, L.A., and Arp, D.J. (2001) Transcript analysis of multiple copies of amo (encoding ammonia monooxygenase) and hao (encoding hydroxylamine oxidoreductase) in Nitrosomonas europaea. *J. Bacteriol.* **183**: 1096–1100.

Hör, J., Gorski, S.A., and Vogel, J. (2018) Bacterial RNA Biology on a Genome Scale. *Mol. Cell* **70**: 785–799.

Horak, R.E.A., Qin, W., Schauer, A.J., Armbrust, E.V., Ingalls, A.E., Moffett, J.W., et al. (2013) Ammonia oxidation kinetics and temperature sensitivity of a natural marine community dominated by Archaea. *ISME J.* **7**: 2023–33.

Hornek, R., Pommerening-Röser, A., Koops, H.P., Farnleitner, A.H., Kreuzinger, N., Kirschner, A., and Mach, R.L. (2006) Primers containing universal bases reduce multiple amoA gene specific DGGE band patterns when analysing the diversity of beta-ammonia oxidizers in the environment. *J. Microbiol. Methods* **66**: 147–155.

Hoshino, T. and Inagaki, F. (2013) A comparative study of microbial diversity and community structure in marine sediments using poly(A) tailing and reverse transcription-PCR. *Front. Microbiol.* **4**: 1–8.

Hou, L., Liu, M., Carini, S.A., and Gardner, W.S. (2012) Transformation and fate of nitrate near the sediment-water interface of Copano Bay. *Cont. Shelf Res.* **35**: 86–94.

Hou, L., Xie, X., Wan, X., Kao, S.-J., Jiao, N., and Zhang, Y. (2018) Niche differentiation of ammonia and nitrite oxidizers along a salinity gradient from the Pearl River estuary to the South China Sea. *Biogeosciences Discuss.* 1–61.

Howarth, R.W. and Marino, R. (2006) Nitrogen as the limiting nutrient for eutrophication in coastal marine ecosystems: Evolving views over three decades. *Limnol. Ocean.* **51**: 364–376.

Hu, H.W., Chen, D., and He, J.Z. (2015) Microbial regulation of terrestrial nitrous oxide formation: Understanding the biological pathways for prediction of emission rates. *FEMS Microbiol. Rev.* **39**: 729–749.

Hu, H.W., Macdonald, C.A., Trivedi, P., Anderson, I.C., Zheng, Y., Holmes, B., et al. (2016) Effects of climate warming and elevated CO2 on autotrophic nitrification and nitrifiers in dryland ecosystems. *Soil Biol. Biochem.* **92**: 1–15.

Hulth, S., Aller, R.C., Canfield, D.E., Dalsgaard, T., Engström, P., Gilbert, F., et al. (2005) Nitrogen removal in marine environments: Recent findings and future research challenges. *Mar. Chem.* **94**: 125–145.

Jahn, C.E., Charkowski, A.O., and Willis, D.K. (2008) Evaluation of isolation methods and RNA integrity for

bacterial RNA quantitation. *J. Microbiol. Methods* **75**: 318–324.

Jameson, E., Stephenson, J., Jones, H., Millard, A., Kaster, A.K., Purdy, K.J., et al. (2019) Deltaproteobacteria (Pelobacter) and Methanococcoides are responsible for choline-dependent methanogenesis in a coastal saltmarsh sediment. *ISME J.* **13**: 277–289.

Javanaud, C., Michotey, V., Guasco, S., Garcia, N., Anschutz, P., Canton, M., and Bonin, P. (2011) Anaerobic ammonium oxidation mediated by Mn-oxides: From sediment to strain level. *Res. Microbiol.* **162**: 848–857.

Jia, Z. and Conrad, R. (2009) Bacteria rather than Archaea dominate microbial ammonia oxidation in an agricultural soil. *Environ. Microbiol.* **11**: 1658–1671.

JIA, Z., ZHOU, X., XIA, W., FORNARA, D., WANG, B., WASSON, E.A., et al. (2020) Evidence for niche differentiation of nitrifying communities in grassland soils after 44 years of different field fertilization scenarios. *Pedosphere* **30**: 87–97.

Jing, X., Gou, H., Gong, Y., Su, X., Xu, L., Ji, Y., et al. (2018) Raman-activated cell sorting and metagenomic sequencing revealing carbon-fixing bacteria in the ocean. *Environ. Microbiol.* **20**: 2241–2255.

Jørgensen, B.B., Findlay, A.J., and Pellerin, A. (2019) The biogeochemical sulfur cycle of marine sediments. *Front. Microbiol.* **10**: 1–27.

Joshi, N.A. and Sickle, J.N.F. (2011) A sliding-window, adaptive, quality-based trimming tool for FastQ files. 1–9.

Joye, S.B. and Hollibaugh, J.T. (1995) Influence of sulfide inhibition of nitrification on nitrogen regeneration in sediments. *Science (80-.).* **270**: 623–625.

Juliette, L.Y., Hyman, M.R., and Arp, D.J. (1993) Inhibition of ammonia oxidation in Nitrosomonas europaea by sulfur compounds: Thioethers are oxidized to sulfoxides by ammonia monooxygenase. *Appl. Environ. Microbiol.* **59**: 3718–3727.

Jung, M.-Y., Well, R., Min, D., Giesemann, A., Park, S.-J., Kim, J.-G., et al. (2014) Isotopic signatures of N2O produced by ammonia-oxidizing archaea from soils. *ISME J.* **8**: 1115–1125.

Jung, M.Y., Park, S.J., Min, D., Kim, J.S., Rijpstra, W.I.C., Damst??, J.S.S., et al. (2011) Enrichment and characterization of an autotrophic ammonia-oxidizing archaeon of mesophilic crenarchaeal group I.1a from an agricultural soil. *Appl. Environ. Microbiol.* **77**: 8635–8647.

Junier, P., Kim, O.S., Molina, V., Limburg, P., Junier, T., Imhoff, J.F., and Witzel, K.P. (2008) Comparative in silico analysis of PCR primers suited for diagnostics and cloning of ammonia monooxygenase genes from ammonia-oxidizing bacteria. *FEMS Microbiol. Ecol.* **64**: 141–152.

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

Kampschreur, M.J., Tan, N.C.G., Picioreanu, C., Jetten, M.S.M., Schmidt, I., and van Loosdrecht, M.C.M. (2006) Role of nitrogen oxides in the metabolism of ammonia-oxidizing bacteria. *Biochem. Soc. Trans.* **34**: 179–181.

Kapoor, V., Pitkänen, T., Ryu, H., Elk, M., Wendell, D., and Santo Domingo, J.W. (2015) Distribution of human-specific Bacteroidales and fecal indicator bacteria in an urban watershed impacted by sewage pollution, determined using RNA- and DNA-based quantitative PCR assays. *Appl. Environ. Microbiol.* **81**: 91–99.

Karlsson, R., Karlsson, A., Bäckman, O., Johansson, B.R., and Hulth, S. (2009) Identification of key proteins involved in the anammox reaction. *FEMS Microbiol. Lett.* **297**: 87–94.

Katoh, K., Asimenos, G., and Toh, H. (2009) Multiple Alignment of DNA Sequences with MAFFT. In, *Bioinformatics for DNA Sequence Analysis, Methods in Molecular Biology*., pp. 39–64.

Ke, X., Angel, R., Lu, Y., and Conrad, R. (2013) Niche differentiation of ammonia oxidizers and nitrite oxidizers in rice paddy soil. *Environ. Microbiol.* **15**: 2275–2292.

van Kessel, M.A.H.J., Speth, D.R., Albertsen, M., Nielsen, P.H., Op den Camp, H.J.M., Kartal, B., et al. (2015) Complete nitrification by a single microorganism. *Nature* **528**: 555–559.

Kietrys, A.M.K., Velema, W.A., and Kool, E.T. (2018) Fingerprints of Modified RNA Bases from Deep Sequencing Profiles. *J Am Chem Soc* **139**: 17074–17081.

Kim, Y.M., Park, D., Lee, D.S., and Park, J.M. (2008) Inhibitory effects of toxic compounds on nitrification process for cokes wastewater treatment. *J. Hazard. Mater.* **152**: 915–921.

Kits, K.D., Sedlacek, C.J., Lebedeva, E. V., Han, P., Bulaev, A., Pjevac, P., et al. (2017) Kinetic analysis of a complete nitrifier reveals an oligotrophic lifestyle. *Nature* **549**: 269–272.

Kitzinger, K., Marchant, H.K., Bristow, L.A., Herbold, C.W., Padilla, C.C., Kidane, A.T., et al. (2020) Single cell analyses reveal contrasting life strategies of the two main nitrifiers in the ocean. *Nat. Commun.* **11**:.

Klotz, M.G., Arp, D.J., Chain, P.S.G., El-Sheikh, A.F., Hauser, L.J., Hommes, N.G., et al. (2006) Complete genome sequence of the marine, chemolithoautotrophic, ammonia-oxidizing bacterium Nitrosococcus oceani ATCC 19707. *Appl. Environ. Microbiol.* **72**: 6299–6315.

Kneip, C., Lockhart, P., Voss, C., and Maier, U.G. (2007) Nitrogen fixation in eukaryotes--new models for symbiosis. *BMC Evol Biol* **7**: 55.

Koch, H., Galushko, A., Albertsen, M., Schintlmeister, A., Spieck, E., Richter, A., et al. (2011) R es e a rc h | r e po r ts 11. **1696**: 761–763.

Koch, H., van Kessel, M.A.H.J., and Lücker, S. (2019) Complete nitrification: insights into the ecophysiology of comammox Nitrospira. *Appl. Microbiol. Biotechnol.* **103**: 177–189.

Koch, H., Lücker, S., Albertsen, M., Kitzinger, K., Herbold, C., Spieck, E., et al. (2015) Expanded metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus Nitrospira. *Proc. Natl. Acad. Sci. U. S. A.* **112**: 11371–6.

Könneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl, D. a (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**: 543–6.

Könneke, M., Schubert, D.M., Brown, P.C., Hügler, M., Standfest, S., Schwander, T., et al. (2014) Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO2 fixation. *Proc. Natl. Acad. Sci. U. S. A.* **111**: 8239–44.

Kool, D.M., Dolfing, J., Wrage, N., and Van Groenigen, J.W. (2011) Nitrifier denitrification as a distinct and significant source of nitrous oxide from soil. *Soil Biol. Biochem.* **43**: 174–178.

Kozlowski, Jessica A., Dimitri Kits, K., and Stein, L.Y. (2016) Comparison of nitrogen oxide metabolism among diverse ammonia-oxidizing bacteria. *Front. Microbiol.* **7**: 1–9.

Kozlowski, J.A., Price, J., and Stein, L.Y. (2014) Revision of N2O-producing pathways in the ammonia-oxidizing bacterium Nitrosomonas europaea

ATCC 19718. *Appl. Environ. Microbiol.* **80**: 4930–4935.

Kozlowski, Jessica A, Stieglmeier, M., Schleper, C., Klotz, M.G., and Stein, L.Y. (2016) Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and Thaumarchaeota. *ISME J.* **10**: 1–10.

Kuenen, J.G. (2008) Anammox bacteria: from discovery to application. *Nat. Rev. Microbiol.* **6**: 320–326.

Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**: 1870–1874.

Kuypers, M.M.M., Marchant, H.K., and Kartal, B. (2018) The microbial nitrogen-cycling network. *Nat. Rev. Microbiol.*

Laima, M., Brossard, D., Girard, M., Richard, P., Gouleau, D., and Joassard, L. (2002) The influence of long emersion on biota, ammonium fluxes and nitrification in intertidal sediments of Marennes-Oléron Bay, France. *Mar. Environ. Res.* **53**: 381–402.

Laima, M.J.C., Girard, M.F., Vouvé, F., Richard, P., Blanchard, G., and Gouleau, D. (1999) Nitrification rates related to sedimentary structures in an Atlantic intertidal mudflat, Marennes-Oleron Bay, France. *Mar. Ecol. Prog. Ser.* **191**: 33–41.

Lancaster, K.M., Caranto, J.D., Majer, S.H., and Smith, M.A. (2018) Alternative Bioenergy: Updates to and Challenges in Nitrification Metalloenzymology. *Joule* **2**: 421–441.

Lavergne, C., Agogué, H., Leynaert, A., Raimonet, M., De Wit, R., Pineau, P., et al. (2017) Factors influencing prokaryotes in an intertidal mudflat and the resulting depth gradients. *Estuar. Coast. Shelf Sci.* **189**: 74–83.

Lavergne, C., Beaugeard, L., Dupuy, C., Courties, C., and Agogué, H. (2014) An efficient and rapid method for the enumeration of heterotrophic prokaryotes in coastal sediments by flow cytometry. *J. Microbiol. Methods* **105**: 31–38.

Lavergne, C., Hugoni, Mylène, Dupuy, C., and Agogué, H. (2018) First evidence of the presence and activity of archaeal C3 group members in an Atlantic intertidal mudflat. *Sci. Rep.* **8**: 1–5.

Lavergne, C., Hugoni, M., Hubas, C., Debroas, D., Dupuy, C., and Agogué, H. (2018) Diel Rhythm Does Not Shape the Vertical Distribution of Bacterial and Archaeal 16S rRNA Transcript Diversity in Intertidal Sediments: a Mesocosm Study. *Microb. Ecol.* **75**: 364–374.

Lee, K.H., Wang, Y.F., Wang, Y., Gu, J.D., and Jiao, J.J. (2018) Abundance and Diversity of Aerobic/Anaerobic Ammonia/Ammonium-Oxidizing Microorganisms in an Ammonium-Rich Aquitard in the Pearl River Delta of South China. *Microb. Ecol.* **76**: 81–91.

Legendre, P. and Gallagher, E.D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**: 271–280.

Lehtovirta-Morley, L.E. (2018) Ammonia oxidation: Ecology, physiology, biochemistry and why they must all come together. *FEMS Microbiol. Lett.* **365**: 1–9.

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

Lekanne Deprez, R.H., Fijnvandraat, A.C., Ruijter, J.M., and Moorman, A.F.M. (2002) Sensitivity and accuracy of quantitative real-time polymerase chain reaction using SYBR green I depends on cDNA synthesis conditions. *Anal. Biochem.* **307**: 63–

69.

Levesque-Sergerie, J.-P., Duquette, M., Thibault, C., Delbecchi, L., and Bissonnette, N. (2007) Detection limits of several commercial reverse transcriptase enzymes: impact on the low- and high-abundance transcript levels assessed by quantitative RT-PCR. *BMC Mol. Biol.* **8**: 93.

Levy-Booth, D.J., Prescott, C.E., and Grayston, S.J. (2014) Microbial functional genes involved in nitrogen fixation, nitrification and denitrification in forest ecosystems. *Soil Biol. Biochem.* **75**: 11–25.

Li, J., Nedwell, D.B., Beddow, J., Dumbrell, A.J., McKew, B.A., Thorpe, E.L., and Whitby, C. (2015a) amoA gene abundances and nitrification potential rates suggest that benthic ammonia-oxidizing bacteria and not archaea dominate N cycling in the Colne estuary, United Kingdom. *Appl. Environ. Microbiol.* **81**: 159–165.

Li, J., Nedwell, D.B., Beddow, J., Dumbrell, A.J., McKew, B.A., Thorpe, E.L., and Whitby, C. (2015b) amoA gene abundances and nitrification potential rates suggest that benthic ammonia-oxidizing bacteria and not archaea dominate N cycling in the Colne estuary, United Kingdom. *Appl. Environ. Microbiol.* **81**: 159–165.

Li, M., Wei, G., Shi, W., Sun, Z., Li, H., Wang, X., and Gao, Z. (2018) Distinct distribution patterns of ammonia-oxidizing archaea and bacteria in sediment and water column of the Yellow River estuary. *Sci. Rep.* **8**: 1–10.

Li, X., Hou, L., Liu, M., Zheng, Y., Yin, G., Lin, X., et al. (2015) Evidence of Nitrogen Loss from Anaerobic Ammonium Oxidation Coupled with Ferric Iron Reduction in an Intertidal Wetland. *Environ. Sci. Technol.* **49**: 11560–11568.

Li, Y., Jia, Z., Sun, Q., Cheng, J., Yang, Y., Zhan, J., and Wang, D. (2017) Plant-Mediated Changes in Soil N-Cycling Genes during Revegetation of Copper Mine Tailings. *Front. Environ. Sci.* **5**: 1–11.

Li, Y., Yu, S., Strong, J., and Wang, H. (2012) Are the biogeochemical cycles of carbon, nitrogen, sulfur, and phosphorus driven by the "FeIII-FeII redox wheel" in dynamic redox environments? *J. Soils Sediments* **12**: 683–693.

Lin, H. and Taillefert, M. (2014) Key geochemical factors regulating Mn(IV)-catalyzed anaerobic nitrification in coastal marine sediments. *Geochim. Cosmochim. Acta* **133**: 17–33.

Liu, L.Y., Xie, G.J., Xing, D.F., Liu, B.F., Ding, J., Cao, G.L., and Ren, N.Q. (2021) Sulfate dependent ammonium oxidation: A microbial process linked nitrogen with sulfur cycle and potential application. *Environ. Res.* **192**: 110282.

Liu, S., Meng, C., Xu, G., Jian, H., and Wang, F. (2018) Validation of reference genes for reverse transcription real-time quantitative PCR analysis in the deep-sea bacterium Shewanella psychrophila WP2. *FEMS Microbiol. Lett.* **365**: 1–5.

Liu, Y., Beyer, A., and Aebersold, R. (2016) On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**: 535–550.

Liu, Y., Liu, J., Yao, P., Ge, T., Qiao, Y., Zhao, M., and Zhang, X.H. (2018) Distribution patterns of ammonia-oxidizing archaea and bacteria in sediments of the eastern China marginal seas. *Syst. Appl. Microbiol.* **41**: 658–668.

Liu, Y., Zhou, H., Wang, J., Liu, X., Cheng, K., Li, L., et al. (2015) Short-term response of nitrifier communities and potential nitrification activity to elevated CO2 and temperature interaction in a Chinese paddy field.

Appl. Soil Ecol. **96**: 88–98.

Lloyd, K.G., MacGregor, B.J., and Teske, A. (2010) Quantitative PCR methods for RNA and DNA in marine sediments: Maximizing yield while overcoming inhibition. *FEMS Microbiol. Ecol.* **72**: 143–151.

Lodhi, M.A.. (2016) Torrey Botanical Society Additional Evidence of Inhibition of Nitrifiers and Possible Cycling of Inhibitors Produced by Selected Plants in a Climax Community Author (s): M . A . K . Lodhi Source : Bulletin of the Torrey Botanical Club , Vol . 109 , No . *Bull. Torrey Bot. Club* **109**: 199–204.

Löscher, C.R., Kock, A., Könneke, M., Laroche, J., Bange, H.W., and Schmitz, R.A. (2012) Production of oceanic nitrous oxide by ammonia-oxidizing archaea. *Biogeosciences* **9**: 2419–2429.

Lu, L., Han, W., Zhang, J., Wu, Y., Wang, B., and Lin, X. (2012) Nitrification of archaeal ammonia oxidizers in acid soils is supported by hydrolysis of urea. *ISME J.* **6**: 1978–1984.

Lücker, S., Nowka, B., Rattei, T., Spieck, E., and Daims, H. (2013) The genome of Nitrospina gracilis illuminates the metabolism and evolution of the major marine nitrite oxidizer. *Front. Microbiol.* **4**: 1–19.

Lücker, S., Wagner, M., Maixner, F., Pelletier, E., Koch, H., Vacherie, B., et al. (2010) A Nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 13479–13484.

Lund, M.B., Smith, J.M., and Francis, C. a (2012) Diversity, abundance and expression of nitrite reductase (nirK)-like genes in marine thaumarchaea. *ISME J.* **6**: 1966–1977.

Magalhaes, C.M., MacHado, A., and Bordalo, A.A. (2009) Temporal variability in the abundance of ammonia-oxidizing bacteria vs. archaea in sandy sediments of the Douro River estuary, Portugal. *Aquat. Microb. Ecol.* **56**: 13–23.

Mäki, A. and Tiirola, M. (2018) Directional high-throughput sequencing of RNAs without gene-specifi c primers. *Biotechniques* **65**: 219–223.

Martens-Habbena, W., Berube, P.M., Urakawa, H., De La Torre, J.R., and Stahl, D.A. (2009) Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature* **461**: 976–979.

Marton, J.M., Roberts, B.J., Bernhard, A.E., and Giblin, A.E. (2015) Spatial and Temporal Variability of Nitrification Potential and Ammonia-Oxidizer Abundances in Louisiana Salt Marshes. *Estuaries and Coasts* **38**: 1824–1837.

McCarthy, A., Chiang, E., Schmidt, M.L., and Denef, V.J. (2015) RNA preservation agents and nucleic acid extraction method bias perceived bacterial community composition. *PLoS One* **10**: 1–14.

McCarty, G.W. (1999) Modes of action of nitrification inhibitors. *Biol. Fertil. Soils* **29**: 1–9.

McClain, M.E., Boyer, E.W., Dent, C.L., Gergel, S.E., Grimm, N.B., Groffman, P.M., et al. (2003) Biogeochemical Hot Spots and Hot Moments at the Interface of Terrestrial and Aquatic Ecosystems. *Ecosystems* **6**: 301–312.

McIlvenna, D., Huang, W.E., Davison, P., Glidle, A., Cooper, J., and Yin, H. (2016) Continuous cell sorting in a flow based on single cell resonance Raman spectra. *Lab Chip* **16**: 1420–1429.

McMurdie, P.J. and Holmes, S. (2013) Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**:.

Mellbye, B.L., Bottomley, P.J., and Sayavedra-Soto, L.A. (2015) Nitrite-

oxidizing bacterium Nitrobacter winogradskyi produces N-acyl-homoserine lactone autoinducers. *Appl. Environ. Microbiol.* **81**: 5917–5926.

Merbt, S.N., Bernal, S., Proia, L., Martí, E., and Casamayor, E.O. (2017) Photoinhibition on natural ammonia oxidizers biofilm populations and implications for nitrogen uptake in stream biofilms. *Limnol. Oceanogr.* **62**: 364–375.

Ming, H., Zhang, H., Chen, Q., Wang, Y., Su, J., Zhao, X., and Fan, J. (2020) Abundance and community structure of ammonium monooxygenase (amoA) genes in the wet season of liaohe estuary sediments. *Cont. Shelf Res.* **209**: 104253.

Miranda, J.A. and Steward, G.F. (2017) Variables influencing the efficiency and interpretation of reverse transcription quantitative PCR (RT-qPCR): An empirical study using Bacteriophage MS2. *J. Virol. Methods* **241**: 1–10.

Mogollón, J.M., Mewes, K., and Kasten, S. (2016) Quantifying manganese and nitrogen cycle coupling in manganese-rich, organic carbon-starved marine sediments: Examples from the Clarion-Clipperton fracture zone. *Geophys. Res. Lett.* **43**: 7114–7123.

Moin, N.S., Nelson, K.A., Bush, A., and Bernhard, A.E. (2009) Distribution and diversity of archaeal and bacterial ammonia oxidizers in salt marsh sediments. *Appl. Environ. Microbiol.* **75**: 7461–7468.

Monteiro, M., Séneca, J., and Magalhães, C. (2014) The history of aerobic ammonia oxidizers: From the first discoveries to today. *J. Microbiol.* **52**: 537–547.

Moreno-vivián, C., Cabello, P., Blasco, R., Castillo, F., Cabello, N., Marti, M., and Moreno-vivia, C. (1999) Prokaryotic Nitrate Reduction : Molecular Properties and Functional Distinction among Bacterial Nitrate Reductases MINIREVIEW Prokaryotic Nitrate Reduction : Molecular Properties and Functional Distinction among Bacterial Nitrate Reductases. **181**: 6573–6584.

Mosier, A.C. and Francis, C.A. (2008) Relative abundance and diversity of ammonia-oxidizing archaea and bacteria in the San Francisco Bay estuary. *Environ. Microbiol.* **10**: 3002–3016.

Muller, C., Laughlin, R.J., Spott, O., and R??tting, T. (2014) Quantification of N2O emission pathways via a 15N tracing model. *Soil Biol. Biochem.* **72**: 44–54.

Murray, N.J., Phinn, S.R., DeWitt, M., Ferrari, R., Johnston, R., Lyons, M.B., et al. (2019) The global distribution and trajectory of tidal flats. *Nature* **565**: 222–225.

Murray, R.H., Erler, D. V., and Eyre, B.D. (2015) Nitrous oxide fluxes in estuarine environments: Response to global change. *Glob. Chang. Biol.* **21**: 3219–3245.

Naqvi, S.W.A., Bange, H.W., FarÃ-As, L., Monteiro, P.M.S., Scranton, M.I., and Zhang, J. (2010) Marine hypoxia/anoxia as a source of CH 4 and N 2O. *Biogeosciences* **7**: 2159–2190.

Neufeld, J.D., Vohra, J., Dumont, M.G., Lueders, T., Manefield, M., Friedrich, M.W., and Murrell, C.J. (2007) DNA stable-isotope probing. *Nat. Protoc.* **2**: 860–866.

Nicol, G.W., Leininger, S., Schleper, C., and Prosser, J.I. (2008) The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria. *Environ. Microbiol.* **10**: 2966–2978.

Norton, J.M., Alzerreca, J.J., Suwa, Y., and Klotz, M.G. (2002) Diversity of ammonia monooxygenase operon in autotrophic ammonia-oxidizing bacteria. *Arch. Microbiol.* **177**: 139–

149.

Norton, J.M., Klotz, M.G., Stein, L.Y., Arp, D.J., Bottomley, P.J., Chain, P.S.G., et al. (2008) Complete genome sequence of Nitrosospira multiformis, an ammonia-oxidizing bacterium from the soil environment. *Appl. Environ. Microbiol.* **74**: 3559–3572.

Nsenga Kumwimba, M. and Meng, F. (2019) Roles of ammonia-oxidizing bacteria in improving metabolism and cometabolism of trace organic chemicals in biological wastewater treatment processes: A review. *Sci. Total Environ.* **659**: 419–441.

Okello, J.B.A., Rodriguez, L., Poinar, D., Bos, K., Okwi, A.L., Bimenya, G.S., et al. (2010) Quantitative assessment of the sensitivity of various commercial reverse transcriptases based on armored HIV RNA. *PLoS One* **5**:.

Oksanen, J., Kindt, R., and O'Hara, R.B. (2005) vegan: Community Ecology Package. Available from http://cc.oulu.fi/~jarioksa/. **3**:.

Otte, S., Kuenen, J.G., Nielsen, L.P., Paerl, H.W., Zopfi, J., Schulz, H.N., et al. (1999) Nitrogen, carbon, and sulfur metabolism in natural Thioploca samples. *Appl. Environ. Microbiol.* **65**: 3148–3157.

Ouyang, Y., Norton, J.M., and Stark, J.M. (2017) Ammonium availability and temperature control contributions of ammonia oxidizing bacteria and archaea to nitrification in an agricultural soil. *Soil Biol. Biochem.* **113**: 161–172.

Palatinszky, M., Herbold, C., Jehmlich, N., Pogoda, M., Han, P., von Bergen, M., et al. (2015) Cyanate as an energy source for nitrifiers. *Nature* **524**: 105–108.

Palatinszky, M., Herbold, C., Jehmlich, N., Pogoda, M., Han, P., Von Bergen, M., et al. (2015) Cyanate as an energy source for nitrifiers. *Nature* **524**: 105–108.

Palomo, A., Pedersen, A.G., Fowler, S.J., Dechesne, A., Sicheritz-Pontén, T., and Smets, B.F. (2018) Comparative genomics sheds light on niche differentiation and the evolutionary history of comammox Nitrospira. *ISME J.* **12**: 1779–1793.

Parada, A.E., Needham, D.M., and Fuhrman, J.A. (2016) Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**: 1403–1414.

Parro, V., Moreno-Paz, M., and González-Toril, E. (2007) Analysis of environmental transcriptomes by DNA microarrays. *Environ. Microbiol.* **9**: 453–464.

Pascal, P.-Y., Dupuy, C., Richard, P., Mallet, C., du Châtelet, E.A., and Niquil, N. (2009) Seasonal variation in consumption of benthic bacteria by meio- and macrofauna in an intertidal mudflat. *Limnol. Oceanogr.* **54**: 1048–1059.

Pelissari, C., Guivernau, M., Viñas, M., García, J., Velasco-Galilea, M., Souza, S.S., et al. (2018) Effects of partially saturated conditions on the metabolically active microbiome and on nitrogen removal in vertical subsurface flow constructed wetlands. *Water Res.* **141**: 185–195.

Pelissari, C., Guivernau, M., Viñas, M., de Souza, S.S., García, J., Sezerino, P.H., and Ávila, C. (2017) Unraveling the active microbial populations involved in nitrogen utilization in a vertical subsurface flow constructed wetland treating urban wastewater. *Sci. Total Environ.* **584**–**585**: 642–650.

Pérez-Novo, C.A., Claeys, C., Speleman, F., Van Cauwenberge, P., Bachert, C., and Vandesompele, J. (2005) Impact of RNA quality on reference gene expression stability. *Biotechniques* **39**: 52–56.

Pester, M., Maixner, F., Berry, D., Rattei, T., Koch, H., Lücker, S., et al. (2014) NxrB encoding the beta subunit of nitrite oxidoreductase as functional and phylogenetic marker for nitrite-oxidizing Nitrospira. *Environ. Microbiol.* **16**: 3055–3071.

Petersen, N.R., Rysgaard, S., Nielsen, L.P., and Revsbech, N.P. (1994) Diurnal variation of denitrification and nitrification in sediments colonized by benthic microphytes. *Limnol. Oceanogr.* **39**: 573–579.

Philips, S., Wyffels, S., Sprengers, R., and Verstraete, W. (2002) Oxygen-limited autotrophic nitrification/denitrification by ammonia oxidisers enables upward motion towards more favourable conditions. *Appl. Microbiol. Biotechnol.* **59**: 557–566.

Pierre, S., Hewson, I., Sparks, J.P., Litton, C.M., Giardina, C., Groffman, P.M., and Fahey, T.J. (2017) Ammonia oxidizer populations vary with nitrogen cycling across a tropical montane mean annual temperature gradient. *Ecology* **98**: 1896–1907.

Pilegaard, K. (2013) Processes regulating nitric oxide emissions from soils. *Philos Trans R Soc L. B Biol Sci* **368**: 20130126.

Pinto, A.J., Marcus, D.N., Ijaz, U.Z., Bautista-de lose Santos, Q.M., Dick, G.J., and Raskin, L. (2016) Metagenomic Evidence for the Presence of Comammox *Nitrospira* -Like Bacteria in a Drinking Water System. *mSphere* **1**: e00054-15.

Polumuri, S.K., Ruknudin, A., and Schulze, D.H. (2002) RNase H and its effects on PCR. *Biotechniques* **32**: 1224–1225.

Posman, K.M., DeRito, C.M., and Madsen, E.L. (2017) Benzene degradation by a Variovorax species within a coal tar-contaminated groundwater microbial community. *Appl. Environ. Microbiol.* **83**: 1–13.

Potapov, V., Fu, X., Dai, N., Corrêa, I.R., Tanner, N.A., and Ong, J.L. (2018) Base modifications affecting RNA polymerase and reverse transcriptase fidelity. *Nucleic Acids Res.* **46**: 5753–5763.

Poth, M. and Focht, D.D. (1985) 15N kinetic analysis of N2O production by Nitrosomonas europaea: An examination of nitrifier denitrification. *Appl. Environ. Microbiol.* **49**: 1134–1141.

Pratscher, J., Dumont, M.G., and Conrad, R. (2011) Ammonia oxidation coupled to CO(2) fixation by archaea and bacteria in an agricultural soil. *Proc. Natl. Acad. Sci. U. S. A.* **108**: 4170–4175.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* **5**:.

Prosser, J.. (2005) Nitrification. 1707–1717.

Prosser, J.I., Hink, L., Gubry-Rangin, C., and Nicol, G.W. (2019) Nitrous oxide production by ammonia oxidizers: Physiological diversity, niche differentiation and potential mitigation strategies. *Glob. Chang. Biol.* **26**: 103–118.

Prosser, J.I. and Nicol, G.W. (2012) Archaeal and bacterial ammonia-oxidisers in soil: The quest for niche specialisation and differentiation. *Trends Microbiol.* **20**: 523–531.

Puthiya Veettil, V., Abdulaziz, A., Chekidhenkuzhiyil, J., Kalanthingal Ramkollath, L., Karayadi Hamza, F., Kizhakkepat Kalam, B., et al. (2015) Bacterial Domination Over Archaea in Ammonia Oxidation in a Monsoon-Driven Tropical Estuary. *Microb. Ecol.* **69**: 544–553.

Qin, W., Zheng, Y., Zhao, F., Wang, Y., Urakawa, H., Martens-Habbena, W., et al. (2020) Alternative strategies of nutrient acquisition and energy conservation map to the biogeography of marine ammonia-oxidizing archaea. *ISME J.* **14**: 2595–

2609.

Reck, M., Tomasch, J., Deng, Z., Jarek, M., Husemann, P., and Wagner-Döbler, I. (2015) Stool metatranscriptomics: A technical guideline for mRNA stabilisation and isolation. *BMC Genomics* **16**: 494.

Rios-Del Toro, E.E. and Cervantes, F.J. (2016) Coupling between anammox and autotrophic denitrification for simultaneous removal of ammonium and sulfide by enriched marine sediments. *Biodegradation* **27**: 107–118.

Risgaard-Petersen, N. (2003) Coupled nitrification-denitrification in autotrophic and heterotrophic estuarine sediments: On the influence of benthic microalgae. *Limnol. Oceanogr.* **48**: 93–105.

Risgaard-Petersen, N., Nicolaisen, M.H., Revsbech, N.P., and Lomstein, B.A. (2004) Competition between ammonia-oxidizing bacteria and benthic microalgae. *Appl. Environ. Microbiol.* **70**: 5528–5537.

Ritchie, G.A.F. and Nicholas, D.J. (1972) Identification of the sources of nitrous oxide produced by oxidative and reductive processes in Nitrosomonas europaea. *Biochem. J.* **126**: 1181–1191.

Robertson, G.P. and Groffman, P.M. (2015) Chapter 14 - Nitrogen Transformations 4th ed. Elsevier Inc.

Robertson, L.A. and Kuenen, J.G. (1990) Combined heterotrophic nitrification and aerobic denitrification in Thiosphaera pantotropha and other bacteria. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **57**: 139–152.

Rocha, D.J.P., Santos, C.S., and Pacheco, L.G.C. (2015) Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis. *Antonie Van Leeuwenhoek* **108**: 685–93.

Rusmana, I. and Nedwell, D.B. (2004) Use of chlorate as a selective inhibitor to distinguish membrane-bound nitrate reductase (Nar) and periplasmic nitrate reductase (Nap) of dissimilative nitrate reducing bacteria in sediment. *FEMS Microbiol. Ecol.* **48**: 379–386.

Rysgaard, Søren, Thastum, P., Dalsgaard, T., Christensen, P.B., Sloth, N.P., and Rysgaard, Soren (1999) Effects of Salinity on NH 4 + Adsorption Capacity, Nitrification, and Denitrification in Danish Estuarine Sediments. *Estuaries* **22**: 21–30.

Saint-Béat, B., Dupuy, C., Agogué, H., Carpentier, A., Chalumeau, J., Como, S., et al. (2014) How does the resuspension of the biofilm alter the functioning of the benthos-pelagos coupled food web of a bare mudflat in Marennes-Oléron Bay (NE Atlantic)? *J. Sea Res.* **92**: 144–157.

Saint-Béat, B., Dupuy, C., Bocher, P., Chalumeau, J., De Crignis, M., Fontaine, C., et al. (2013) Key Features of Intertidal Food Webs That Support Migratory Shorebirds. *PLoS One* **8**:.

Saleh-Lakha, S., Miller, M., Campbell, R.G., Schneider, K., Elahimanesh, P., Hart, M.M., and Trevors, J.T. (2005) Microbial gene expression in soil: Methods, applications and challenges. *J. Microbiol. Methods* **63**: 1–19.

Saleh-Lakha, S., Shannon, K.E., Goyer, C., and Trevors, J.T. (2011) Challenges in quantifying microbial gene expression in soil using quantitative reverse transcription real-time PCR. *J. Microbiol. Methods* **85**: 239–243.

Sanders, R., Mason, D.J., Foy, C.A., and Huggett, J.F. (2014) Considerations for accurate gene expression measurement by reverse transcription quantitative PCR when analysing clinical samples. *Anal. Bioanal. Chem.* **406**: 6471–6483.

Sanders, T. and Laanbroek, H.J. (2018) The distribution of sediment and water column nitrification potential in the

hyper-turbid Ems estuary. *Aquat. Sci.* **80**: 1–13.

Santoro, A.E., Buchwald, C., McIlvin, M.R., and Casciotti, K.L. (2011) Isotopic Signature of N2O Produced by Marine Ammonia-Oxidizing Archaea. *Science (80-.).* **333**: 1282–1285.

Santoro, A.E., Casciotti, K.L., and Francis, C.A. (2010) Activity, abundance and diversity of nitrifying archaea and bacteria in the central California Current. *Environ. Microbiol.* **12**: 1989–2006.

Santoro, A.E., Francis, C.A., De Sieyes, N.R., and Boehm, A.B. (2008a) Shifts in the relative abundance of ammonia-oxidizing bacteria and archaea across physicochemical gradients in a subterranean estuary. *Environ. Microbiol.* **10**: 1068–1079.

Santoro, A.E., Francis, C.A., De Sieyes, N.R., and Boehm, A.B. (2008b) Shifts in the relative abundance of ammonia-oxidizing bacteria and archaea across physicochemical gradients in a subterranean estuary. *Environ. Microbiol.* **10**: 1068–1079.

Santos, J.P., Mendes, D., Monteiro, M., Ribeiro, H., Baptista, M.S., Borges, M.T., and Magalhães, C. (2018) Salinity impact on ammonia oxidizers activity and amoA expression in estuarine sediments. *Estuar. Coast. Shelf Sci.* **211**: 177–187.

Sayavedra-Soto, L.A., Gvakharia, B., Bottomley, P.J., Arp, D.J., and Dolan, M.E. (2010) Nitrification and degradation of halogenated hydrocarbons-a tenuous balance for ammonia-oxidizing bacteria. *Appl. Microbiol. Biotechnol.* **86**: 435–444.

Sayavedra-Soto, L.A., Hommes, N.G., Alzerreca, J.J., Arp, D.J., Norton, J.M., and Klotz, M.G. (1998) Transcription of the amoC, amoA and amoB genes in Nitrosomonas europaea and Nitrosospira sp. NpAV. *FEMS Microbiol. Lett.* **167**: 81–88.

Sayavedra-Soto, L.A., Hommes, N.G., Russell, S.A., and Arp, D.J. (1996) Induction of ammonia monooxygenase and hydroxylamine oxidoreductase mRNAs by ammonium in Nitrosomonas europaea. *Mol. Microbiol.* **20**: 541–548.

Schaefer, B.C. (1995) Revolutions in rapid amplification of cDNA ends: New strategies for polymerase chain reaction cloning of full-length cDNA ends. *Anal. Biochem.* **227**: 255–273.

Schindler, D.W. and Bayley, S.E. (1993) The biosphere as an increasing sink for atmospheric carbon: Estimates from increased nitrogen depostion. *Global Biogeochem. Cycles* **7**: 717–733.

Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T., and Quince, C. (2015) Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* **43**:.

Schleper, C. and Nicol, G.W. (2010) Ammonia-oxidising archaea - physiology, ecology and evolution Elsevier Ltd.

Schmidt, I. (2009) Chemoorganoheterotrophic growth of Nitrosomonas europaea and Nitrosomonas eutropha. *Curr. Microbiol.* **59**: 130–138.

Schmidt, I., Bock, E., and Jetten, M.S.M. (2001) Ammonia oxidation by Nitrosomonas eutropha with NO2 as oxidant is not inhibited by acetylene. *Microbiology* **147**: 2247–2253.

Schmidt, Ingo, Sliekers, A.O., Schmid, M.C., Cirpus, I., Strous, M., Bock, E., et al. (2001) Aerobic and anaerobic ammonia oxidizing bacteria competitors or natural partners?vier Science B.V. All rights reserved. *FEMS Microbiol. Ecol.* **39**: 175–181.

Schott, J., Griffin, B.M., and Schink, B. (2010) Anaerobic phototrophic nitrite oxidation by Thiocapsa sp. strain KS1 and Rhodopseudomonas sp. strain LQ17. *Microbiology* **156**: 2428–

2437.

Schramm, A., De Beer, D., Van Den Heuvel, J.C., Ottengraf, S., and Amann, R. (1999) Microscale distribution of populations and activities of Nitrosospira and Nitrospira spp. along a macroscale gradient in a nitrifying bioreactor: Quantification by in situ hybridization and the use of microsensors. *Appl. Environ. Microbiol.* **65**: 3690–3696.

Schulz, H.N. and Jørgensen, B.B. (2001) Big acteria. *Annu. Rev. Microbiol* **55**: 105–137.

Schutte, C.A., Ahmerkamp, S., Wu, C.S., Seidel, M., De Beer, D., Cook, P.L.M., and Joye, S.B. (2018) Biogeochemical dynamics of coastal tidal flats.

Segall-Shapiro, T.H., Meyer, A.J., Ellington, A.D., Sontag, E.D., and Voigt, C.A. (2014) A "resource allocator" for transcription based on a highly fragmented T7 RNA polymerase. *Mol. Syst. Biol.* **10**: 742–742.

Seitz, R.D., Wennhage, H., Bergstrom, U., Lipcius, R.N., and Ysebaert, T. (2014) Ecological value of coastal habitats for commercially and ecologically important species. *ICES J. Mar. Sci.* **71**: 648–665.

Seitzinger, S.P., Gardner, W.S., and Spratt, A.K. (1991) The Effect of Salinity on Ammonium Sorption in Aquatic Sediments: Implications for Benthic Nutrient Recycling. *Estuaries* **14**: 167–174.

Seitzinger, S.P., Harrison, J.A., Böhlke, J.K., Bouwman, A.F., Lowrance, R., Tobias, C., and Drecht, G. Van (2006) Denitrification across Landscapes and Waterscapes : A Synthesis Published by : Wiley Stable URL : http://www.jstor.org/stable/40061944 2064 INVITED FEATURE ^^ f * ^. *Ecol. Appl.* **16**: 2064–2090.

Seitzinger, S.P., Kroeze, C., Bouwman, A.F., Caraco, N., Dentener, F., and Styles, R. V. (2002) Global patterns of dissolved inorganic and particulate nitrogen inputs to coastal systems: Recent conditions and future projections. *Estuaries* **25**: 640–655.

Semeniuk, V. (2019) Tidal Flats. *Encycl. Coast. Sci.* **53**: 1709–1725.

Shakoor, A., Ashraf, F., Shakoor, S., Mustafa, A., Rehman, A., and Altaf, M.M. (2020) Biogeochemical transformation of greenhouse gas emissions from terrestrial to atmospheric environment and potential feedback to climate forcing. *Environ. Sci. Pollut. Res.* **27**: 38513–38536.

Shakya, M., Lo, C.C., and Chain, P.S.G. (2019) Advances and challenges in metatranscriptomic analysis. *Front. Genet.* **10**: 1–10.

Shakya, M., Quince, C., Campbell, J.H., Yang, Z.K., Schadt, C.W., and Podar, M. (2013) Comparative metagenomic and rRNA microbial diversity characterization using archaeal and bacterial synthetic communities. *Environ. Microbiol.* **15**: 1882–1899.

Shaw, L.J., Nicol, G.W., Smith, Z., Fear, J., Prosser, J.I., and Baggs, E.M. (2006) Nitrosospira spp. can produce nitrous oxide via a nitrifier denitrification pathway. *Environ. Microbiol.* **8**: 214–222.

Shen, J.P., Zhang, L.M., Di, H.J., and He, J.Z. (2012) A review of ammonia-oxidizing bacteria and archaea in Chinese soils. *Front. Microbiol.* **3**: 1–7.

Sidova, M., Tomankova, S., Abaffy, P., Kubista, M., and Sindelka, R. (2015) Effects of post-mortem and physical degradation on RNA integrity and quality. *Biomol. Detect. Quantif.* **5**: 3–9.

Sieber, M.W., Recknagel, P., Glaser, F., Witte, O.W., Bauer, M., Claus, R.A., and Frahm, C. (2010) Substantial performance discrepancies among commercially available kits for reverse transcription quantitative

polymerase chain reaction: A systematic comparative investigator-driven approach. *Anal. Biochem.* **401**: 303–311.

Simon, J. and Klotz, M.G. (2013) Diversity and evolution of bioenergetic systems involved in microbial nitrogen compound transformations. *Biochim. Biophys. Acta - Bioenerg.* **1827**: 114–135.

Sims, A., Horton, J., Gajaraj, S., McIntosh, S., Miles, R.J., Mueller, R., et al. (2012) Temporal and spatial distributions of ammonia-oxidizing archaea and bacteria and their ratio as an indicator of oligotrophic conditions in natural wetlands. *Water Res.* **46**: 4121–4129.

Smith, C.J., Nedwell, D.B., Dong, L.F., and Osborn, A.M. (2007) Diversity and abundance of nitrate reductase genes (narG and napA), nitrite reductase genes (nirS and nrfA), and their transcripts in estuarine sediments. *Appl. Environ. Microbiol.* **73**: 3612–3622.

Smith, C.J., Nedwell, D.B., Dong, L.F., and Osborn, A.M. (2006) Evaluation of quantitative polymerase chain reaction-based approaches for determining gene copy and gene transcript numbers in environmental samples. *Environ. Microbiol.* **8**: 804–815.

Smith, C.J. and Osborn, A.M. (2009) Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS Microbiol. Ecol.* **67**: 6–20.

Smith, J.M., Mosier, A.C., and Francis, C.A. (2014) Spatiotemporal Relationships Between the Abundance, Distribution, and Potential Activities of Ammonia-Oxidizing and Denitrifying Microorganisms in Intertidal Sediments. *Microb. Ecol.* **69**: 13–24.

Sorokin, D.Y., Lücker, S., Vejmelkova, D., Kostrikina, N.A., Kleerebezem, R., Rijpstra, W.I.C., et al. (2012) Nitrification expanded: discovery, physiology and genomics of a nitrite-oxidizing bacterium from the phylum Chloroflexi. *ISME J.* **6**: 2245–56.

Ståhlberg, A., Håkansson, J., Xian, X., Semb, H., and Kubista, M. (2004) Properties of the Reverse Transcription Reaction in mRNA Quantification. *Clin. Chem.* **50**: 509–515.

StÅhlberg, A., Kubista, M., and Pfaffl, M. (2004) Comparison of Reverse Transcriptases in Gene Expres- sion Analysis. *Clin. Chem.* **50**: 1678–1680.

Stangegaard, M., Dufva, I.H., and Dufva, M. (2006) Reverse transcription using random pentadecamer primers increases yield and quality of resulting cDNA. *Biotechniques* **40**: 649–657.

Starkenburg, S.R., Chain, P.S.G., Sayavedra-Soto, L. a, Hauser, L., Land, M.L., Larimer, F.W., et al. (2006) Genome sequence of the chemolithoautotrophic nitrite-oxidizing bacterium *Nitrobacter winogradskyi* Nb-255. *Appl. Environ. Microbiol.* **72**: 2050–2063.

Stein, L.Y. (2011) Heterotrophic nitrification and nitrifier denitrification. *Nitrification* 95–114.

Stein, L.Y. and Arp, D.J. (1998) Loss of ammonia monooxygenase activity in Nitrosomonas europaea upon exposure to nitrite. *Appl. Environ. Microbiol.* **64**: 4098–4102.

Stein, L.Y. and Klotz, M.G. (2011) Nitrifying and denitrifying pathways of methanotrophic bacteria. *Biochem. Soc. Trans.* **39**: 1826–1831.

Steiner, P.A., De Corte, D., Geijo, J., Mena, C., Yokokawa, T., Rattei, T., et al. (2019) Highly variable mRNA half-life time within marine bacterial taxa and functional genes. *Environ. Microbiol.* **21**: 3873–3884.

Stieglmeier, M., Mooshammer, M., Kitzler, B., Wanek, W., Zechmeister-

Boltenstern, S., Richter, A., and Schleper, C. (2014) Aerobic nitrous oxide production through N-nitrosating hybrid formation in ammonia-oxidizing archaea. *ISME J.* **8**: 1135–46.

Su, Y., Wang, W., Wu, D., Huang, W., Wang, M., and Zhu, G. (2018) Stimulating ammonia oxidizing bacteria (AOB) activity drives the ammonium oxidation rate in a constructed wetland (CW). *Sci. Total Environ.* **624**: 87–95.

Suter, E.A., Pachiadaki, M.G., Montes, E., Edgcomb, V.P., Scranton, M.I., Taylor, C.D., and Taylor, G.T. (2020) Diverse nitrogen cycling pathways across a marine oxygen gradient indicate nitrogen loss coupled to chemoautotrophic activity. *Environ. Microbiol.* **00**: 1–18.

Suzuki, M.T., Taylor, L.T., Delong, E.F., and Long, E.F.D.E. (2000) Quantitative Analysis of Small-Subunit rRNA Genes in Mixed Microbial Populations via 5 ′ -Nuclease Assays Quantitative Analysis of Small-Subunit rRNA Genes in Mixed Microbial Populations via 5 J -Nuclease Assays. **66**: 4605–4614.

Szukics, U., Abell, G.C.J., Hödl, V., Mitter, B., Sessitsch, A., Hackl, E., and Zechmeister-Boltenstern, S. (2010) Nitrifiers and denitrifiers respond rapidly to changed moisture and increasing temperature in a pristine forest soil. *FEMS Microbiol. Ecol.* **72**: 395–406.

Taylor, A.E., Giguere, A.T., Zoebelein, C.M., Myrold, D.D., and Bottomley, P.J. (2017) Modeling of soil nitrification responses to temperature reveals thermodynamic differences between ammonia-oxidizing activity of archaea and bacteria. *ISME J.* **11**: 896–908.

Tolar, B.B., Ross, M.J., Wallsgrove, N.J., Liu, Q., Aluwihare, L.I., Popp, B.N., and Hollibaugh, J.T. (2016) Contribution of ammonia oxidation to chemoautotrophy in Antarctic coastal waters. *ISME J.* **10**: 2605–2619.

Torsvik, V. and Øvreås, L. (2002) Microbial diversity and function in soil: From genes to ecosystems. *Curr. Opin. Microbiol.* **5**: 240–245.

Tournier, E., Amenc, L., Pablo, A.L., Legname, E., Blanchart, E., Plassard, C., et al. (2015) Modification of a commercial DNA extraction kit for safe and rapid recovery of DNA and RNA simultaneously from soil, without the use of harmful solvents. *MethodsX* **2**: 182–191.

Treusch, A.H., Leininger, S., Kietzin, A., Schuster, S.C., Klenk, H.P., and Schleper, C. (2005) Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environ. Microbiol.* **7**: 1985–1995.

U.S. Environmental Protection Agency (2002) Greenhouse Gases and Global Warming Potential Values. *Environ. Prot.* 1–16.

Urakawa, H., Martens-Habbena, W., Huguet, C., de la Torre, J.R., Ingalls, A.E., Devol, A.H., and Stahl, D. a. (2014) Ammonia availability shapes the seasonal distribution and activity of archaeal and bacterial ammonia oxidizers in the Puget Sound Estuary. *Limnol. Oceanogr.* **59**: 1321–1335.

Urakawa, H., Tajima, Y., Numata, Y., and Tsuneda, S. (2008) Low temperature decreases the phylogenetic diversity of ammonia-oxidizing archaea and bacteria in aquarium biofiltration systems. *Appl. Environ. Microbiol.* **74**: 894–900.

Utåker, J.B., Andersen, K., Aakra, Å., Moen, B., and Nes, I.F. (2002) Phylogeny and Functional Expression of Carboxylase / Oxygenase from the Autotrophic Ammonia-Oxidizing Bacterium Phylogeny and Functional Expression of Ribulose 1 , 5-Bisphosphate Carboxylase / Oxygenase from the Autotrophic

Ammonia-Oxidizing Bacterium Nit. **184**: 468–478.

Vajrala, N., Martens-Habbena, W., Sayavedra-Soto, L. a, Schauer, A., Bottomley, P.J., Stahl, D. a, and Arp, D.J. (2013) Hydroxylamine as an intermediate in ammonia oxidation by globally abundant marine archaea. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 1006–11.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J. a, et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.

Venterea, R.T. (2007) Nitrite-driven nitrous oxide production under aerobic soil conditions: Kinetics and biochemical controls. *Glob. Chang. Biol.* **13**: 1798–1809.

Wakeham, S.G. (2020) Organic biogeochemistry in the oxygen-deficient ocean: A review. *Org. Geochem.* **149**: 1040–96.

Walker, C.B., de la Torre, J.R., Klotz, M.G., Urakawa, H., Pinel, N., Arp, D.J., et al. (2010) Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 8818–23.

Walters, W., R. Hyde, E., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., et al. (2015) Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *Msystems* **1**: 1–10.

Wang, B., Zhao, J., Guo, Z., Ma, J., Xu, H., and Jia, Z. (2015) Differential contributions of ammonia oxidizers and nitrite oxidizers to nitrification in four paddy soils. *ISME J.* **9**: 1062–75.

Wang, C., Tang, S., He, X., and Ji, G. (2020) The abundance and community structure of active ammonia-oxidizing archaea and ammonia-oxidizing bacteria shape their activities and contributions in coastal wetlands. *Water Res.* **171**: 115464.

Wang, S., Wang, W., Liu, L., Zhuang, L., Zhao, S., Su, Y., et al. (2018) Microbial Nitrogen Cycle Hotspots in the Plant-Bed/Ditch System of a Constructed Wetland with N2O Mitigation. *Environ. Sci. Technol.* **52**: 6226–6236.

Wang, Y., Hayatsu, M., and Fujii, T. (2012) Extraction of Bacterial RNA from Soil: Challenges and Solutions. *Microbes Environ.* **27**: 111–121.

Wang, Y., Ji, Y., Wharfe, E.S., Meadows, R.S., March, P., Goodacre, R., et al. (2013) Raman activated cell ejection for isolation of single cells. *Anal. Chem.* **85**: 10697–10701.

Wang, Y., Nagaoka, K., Hayatsu, M., Sakai, Y., Tago, K., Asakawa, S., and Fujii, T. (2012) A novel method for RNA extraction from Andosols using casein and its application to amoA gene expression study in soil. *Appl. Microbiol. Biotechnol.* **96**: 793–802.

Wang, Z. bin, Ni, S.Q., Zhang, J., Zhu, T., Ma, Y. guang, Liu, X. lin, et al. (2016) Gene expression and biomarker discovery of anammox bacteria in different reactors. *Biochem. Eng. J.* **115**: 108–114.

Wankel, S.D., Mosier, A.C., Hansel, C.M., Paytan, A., and Francis, C.A. (2011) Spatial variability in nitrification rates and ammonia-oxidizing microbial communities in the agriculturally impacted Elkhorn Slough estuary, California. *Appl. Environ. Microbiol.* **77**: 269–280.

Ward, B.B. (2013) Nitrification. *Ref. Modul. Earth Syst. Environ. Sci.* 1–8.

Ward, B.B. and O'Mullan, G.D. (2002) Worldwide Distribution of Nitrosococcus oceani, a Marine Ammonia- Oxidizing ⁄-Proteobacterium, Detected by PCR and Sequencing of 16S rRNA and amoA Genes. *Appl. Environ.*

Microbiol. **68**: 4153–4157.

Webster, F. and Hopkins, D. (1996) Contributions from different microbial processes to N 2 O emission from soil under different moisture regimes. *Biol. Fertil. Soils* **22**: 331–335.

Wei, B., Yu, X., Zhang, S., and Gu, L. (2011) Comparison of the community structures of ammonia-oxidizing bacteria and archaea in rhizoplanes of floating aquatic macrophytes. *Microbiol. Res.* **166**: 468–474.

Wellington, E.M.H., Berry, A., and Krsek, M. (2003) Resolving functional diversity in relation to microbial community structure in soil: Exploiting genomics and stable isotope probing. *Curr. Opin. Microbiol.* **6**: 295–301.

Werbrouck, H., Botteldoorn, N., Uyttendaele, M., Herman, L., and Van Coillie, E. (2007) Quantification of gene expression of Listeria monocytogenes by real-time reverse transcription PCR: Optimization, evaluation and pitfalls. *J. Microbiol. Methods* **69**: 306–314.

Wessén, E., Nyberg, K., Jansson, J.K., and Hallin, S. (2010) Responses of bacterial and archaeal ammonia oxidizers to soil organic and fertilizer amendments under long-term management. *Appl. Soil Ecol.* **45**: 193–200.

Winter, D.J. (2017) rentrez: An R package for the NCBI eUtils API. *R J.* **9**: 520–526.

Wrage, N., Van Groenigen, J.W., Oenema, O., and Baggs, E.M. (2005) A novel dual-isotope labelling method for distinguishing between soil sources of N2O. *Rapid Commun. Mass Spectrom.* **19**: 3298–3306.

Wrage, N., Velthof, G.L., Van Beusichem, M.L., and Oenema, O. (2001) Role of nitrifier denitrification in the production of nitrous oxide. *Soil Biol. Biochem.* **33**: 1723–1732.

Wrage, N., Velthof, G.L., Laanbroek, H.J., and Oenema, O. (2004) Nitrous oxide production in grassland soils: Assessing the contribution of nitrifier denitrification. *Soil Biol. Biochem.* **36**: 229–236.

Wuchter, C., Abbas, B., Coolen, M.J.L., Herfort, L., van Bleijswijk, J., Timmers, P., et al. (2006) Archaeal nitrification in the ocean. *Proc. Natl. Acad. Sci.* **103**: 12317–12322.

Xia, W., Zhang, C., Zeng, X., Feng, Y., Weng, J., Lin, X., et al. (2011) Autotrophic growth of nitrifying community in an agricultural soil. *ISME J.* **5**: 1226–1236.

Yan, Y.W., Zou, B., Zhu, T., Hozzein, W.N., and Quan, Z.X. (2017) Modified RNA-seq method for microbial community and diversity analysis using rRNA in different types of environmental samples. *PLoS One* **12**: 1–20.

Yang, X. (2014) Coastal ecosystems.

Yao, Z., Wang, F., Wang, C., Xu, H., and Jiang, H. (2019) Anaerobic ammonium oxidation coupled to ferric iron reduction in the sediment of a eutrophic lake. *Environ. Sci. Pollut. Res.* **26**: 15084–15094.

Yergeau, E., Arbour, M., Brousseau, R., Juck, D., Lawrence, J.R., Masson, L., et al. (2009) Microarray and real-time PCR analyses of the responses of high-arctic soil bacteria to hydrocarbon pollution and bioremediation treatments. *Appl. Environ. Microbiol.* **75**: 6258–6267.

York, A. (2018) Environmental microbiology: Marine biogeochemical cycles in a changing world. *Nat. Rev. Microbiol.* **16**: 259.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.Y. (2017) Ggtree: an R Package for Visualization and Annotation of Phylogenetic Trees With Their Covariates and Other Associated Data. *Methods Ecol. Evol.* **8**: 28–36.

Yung, Y.L., Wang, W.C. and Lacis, A.. (1976) Greenhouse effect due to atmospheric nitrous-oxide. *Geophys. Res. Lett.* **3**: 619–621.

Zaccone, R., Caruso, G., and Azzaro, M. (1996) Detection of Nitrosococcus oceanus in a Mediterranean lagoon by immunofluorescence. *J. Appl. Bacteriol.* **80**: 611–616.

Zhalnina, K. V., Dias, R., Leonard, M.T., De Quadros, P.D., Camargo, F.A.O., Drew, J.C., et al. (2014) Genome sequence of Candidatus nitrososphaera evergladensis from group I.1b enriched from everglades soil reveals novel genomic features of the ammonia-oxidizing archaea. *PLoS One* **9**: 21–23.

Zhang, L., Duff, A., and Smith, C. (2018) Community and functional shifts in ammonia oxidizers across terrestrial and marine (soil/sediment) boundaries in two coastal Bay ecosystems. *Environ. Microbiol.* **20**: 2834–2853.

Zhang, L.M., Offre, P.R., He, J.Z., Verhamme, D.T., Nicol, G.W., and Prosser, J.I. (2010) Autotrophic ammonia oxidation by soil thaumarchaea. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 17240–17245.

Zhang, Y., Chen, L., Dai, T., Sun, R., and Wen, D. (2015) Ammonia manipulates the ammonia-oxidizing archaea and bacteria in the coastal sediment-water microcosms. *Appl. Microbiol. Biotechnol.* **99**: 6481–6491.

Zhang, Y., Chen, L., Dai, T., Tian, J., and Wen, D. (2015) The influence of salinity on the abundance, transcriptional activity, and diversity of AOA and AOB in an estuarine sediment: a microcosm study. *Appl. Microbiol. Biotechnol.* **99**: 9825–9833.

Zhao, J., Bello, M.O., Meng, Y., Prosser, J.I., and Gubry-Rangin, C. (2020) Selective inhibition of ammonia oxidising archaea by simvastatin stimulates growth of ammonia oxidising bacteria. *Soil Biol. Biochem.* **141**:.

Zhao, J., Meng, Y., Drewer, J., Skiba, U.M., Prosser, J.I., and Gubry-Rangin, C. (2020) Differential Ecosystem Function Stability of Ammonia- Oxidizing Archaea and Bacteria following Short-Term Environmental Perturbation. *mSystems* **5**: 00309–20.

Zheng, L., Liu, T., Yuan, D., Wang, H., Zhang, S., Ding, A., and Xie, E. (2020) Abundance, diversity, and distribution patterns along with the salinity of four nitrogen transformation-related microbes in the Yangtze Estuary. *Ann. Microbiol.* **70**:.

Zheng, Y., Hou, L., Liu, M., Lu, M., Zhao, H., Yin, G., and Zhou, J. (2013) Diversity, abundance, and activity of ammonia-oxidizing bacteria and archaea in Chongming eastern intertidal sediments. *Appl. Microbiol. Biotechnol.* **97**: 8351–8363.

Zheng, Y., Hou, L., Newell, S., Liu, M., Zhou, J., Zhao, H., et al. (2013) Community dynamics and activity of ammonia-oxidizing prokaryotes in intertidal sediments of the yangtze estuary. *Appl. Environ. Microbiol.* **80**: 408–419.

Zhou, X., Li, B., Guo, Z., Wang, Z., Luo, J., and Lu, C. (2018) Niche Separation of Ammonia Oxidizers in Mudflat and Agricultural Soils Along the Yangtze River, China. *Front. Microbiol.* **9**: 1–12.

Zhu, X., Burger, M., Doane, T.A., and Horwath, W.R. (2013) Ammonia oxidation pathways and nitrifier denitrification are significant sources of N2O and NO under low oxygen availability. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 6328–33.

Zopfi, J., Kjær, T., Nielsen, L.P., and Jørgensen, B.B. (2001) Ecology of Thioploca spp.: Nitrate and Sulfur Storage in Relation to Chemical

Microgradients and Influence of Thioploca spp. on the Sedimentary Nitrogen Cycle. *Appl. Environ. Microbiol.* **67**: 5530–5537.

Zumft, W.G. (1997) Cell biology and molecular basis of denitrification. *Microbiol Mol Biol Rev* **61**: 533–616.

Annexes

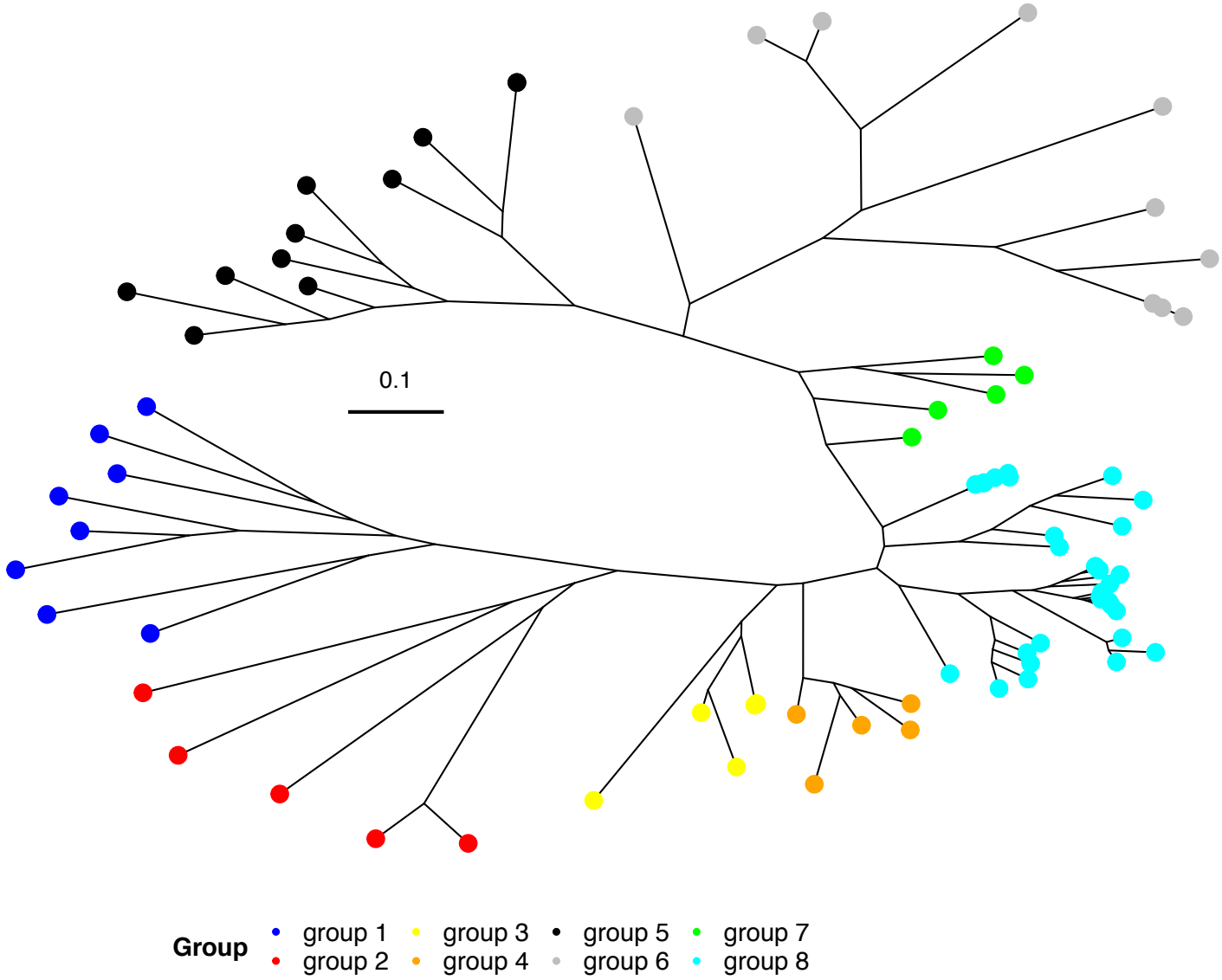


Figure S.1 Evolutionary relationships of the 84 bacterial *glnA* genes used to design new primers. The evolutionary history was inferred using the Neighbor-Joining method (Saitou & Nei, 1987). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site. The analysis involved 84 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 690 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2015). The different glanA groups are represented by coloured tips as indicated in the legend.

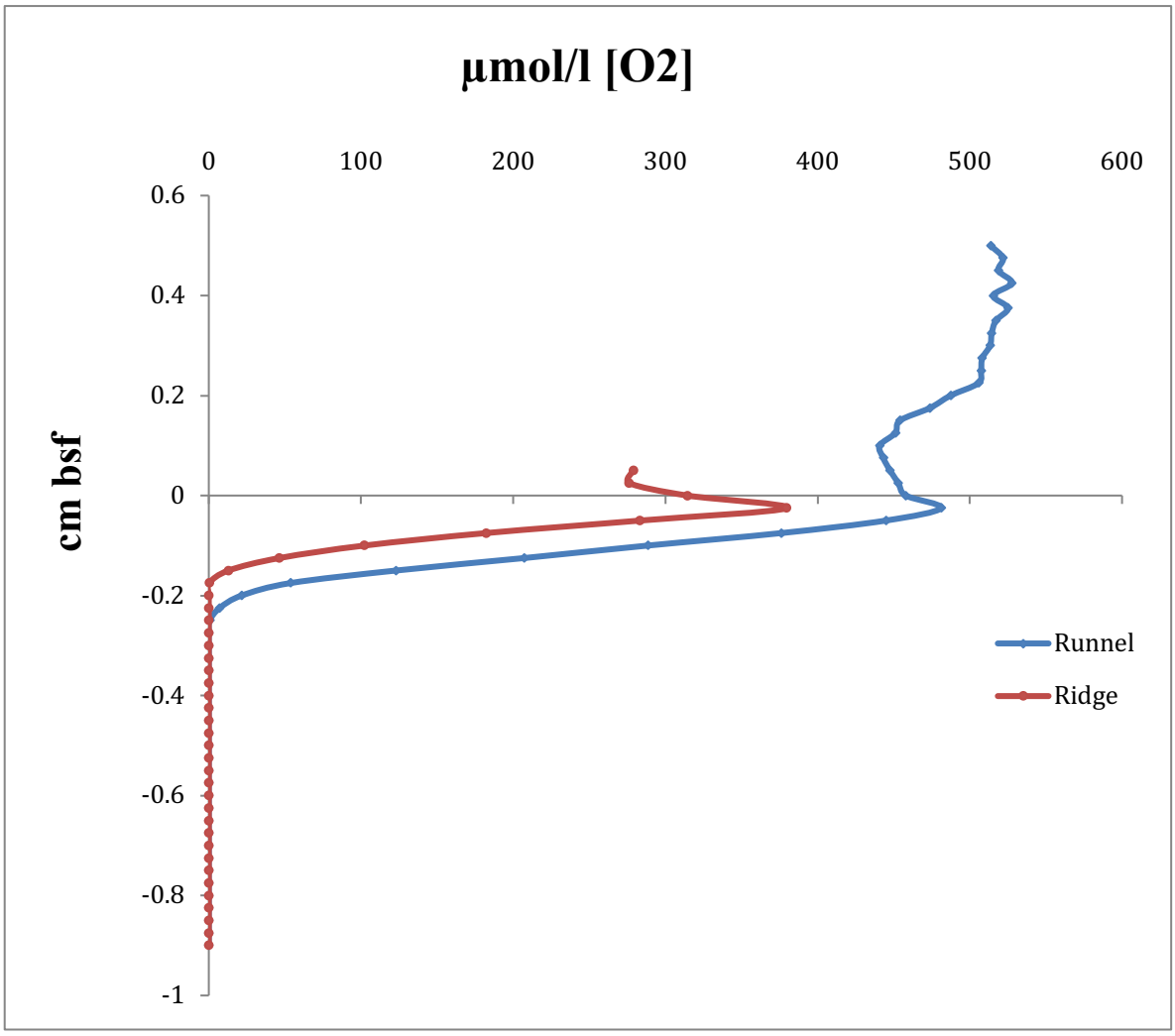


Figure S.2 Oxygen profiles measures in ridges and runnels on the Montportail-Brouage mudflat in 2016.

Supplementary Information 1: list of sequences used to design new *glnA* primers

Sequence ID	Group	Genus/Species
NZ HG322950.1	1	Pseudomonas knackmussii
NZ LUCV01000019.1	1	Pseudomonas putida
NZ CP007441.1	1	Pseudomonas stutzeri
NZ CP009323.1	1	Burkholderia gladioli
NZ CP012504.1	1	Aeromonas veronii
NZ LCWP01000011.1	1	Chromobacterium subtsugae
NZ CP015880.1	1	Ensifer adhaerens
NC 007493.2	1	Rhodobacter sphaeroides
NZ CP009048.1	2	Pseudomonas alkylphenolia
NC 009439.1	2	Pseudomonas mendocina
NC 014034.1	2	Rhodobacter capsulatus
NZ CP007045.1	2	Rhizobium leguminosarum
KX488606.1	2	Rhizobium leguminosarum
CP009124.1	3	Streptomyces lividans
NC 003888.3	3	Streptomyces coelicolor
CP012949.1	3	Streptomyces ambofaciens
LIAN01000232.1	3	Actinobacteria bacterium
NC 003450.3	3	Corynebacterium glutamicum
CP007790.1	3	Corynebacterium marinum
LN885086.1	4	Nitrospira sp
LT828648.1	4	Nitrospira japonica
FP929003.1	4	Nitrospira defluvii
CP011801.1	4	Nitrospira moscoviensis
KR873367.1	4	Nitrospira lenta
L05609.1	5	Calothrix
X00147.1	5	Anabaena
CP003642.1	5	Cylindrospermum stagnale
KX035102.1	5	Mastigocladus
LO018304.1	5	Planktothrix agardhii
HM130917.1	5	Arthrospira platensis
KY010035.1	5	Microcystis sp
CP000815.1	5	Paulinella chromatophora
NC 005042.1	5	Prochlorococcus marinus
AF026393.1	5	Synechococcus
AF026393.1	5	Synechococcus
CP000859.1	6	Desulfococcus oleovorans
NZ CP006951.1	6	Dehalococcoides mccartyi
CP000027.1	6	Dehalococcoides ethenogenes
X60160.1	6	Thiocapsa maritima
NC 012039.1	6	Campylobacter lari
NC 000964.3	6	Bacillus subtilis
NZ CP009828.1	6	Staphylococcus aureus
JX017366.1	6	Bacillus
CP006863.1	6	Bacillus toyonensis

CP003687.1	6	Bacillus thuringiensis
NC 000913.3	7	Escherichia coli
CP017671.1	7	Providencia rettgeri
NC 018632.1	7	Alteromonas macleodii
LPB0090	7	Thalassotalea sp
CP015411.1	7	Pseudoalteromonas luteoviolacea
NC 007643.1	8	Rhodospirillum rubrum
CP000744.1	8	Pseudomonas aeruginosa
CP004143.1	8	Pseudomonas denitrificans
JN676036.1	8	Pseudomonas nitroreducens
HG322950.1	8	Pseudomonas knackmussii
CP014158.1	8	Pseudomonas citronellolis
CP015878.1	8	Pseudomonas citronellolis
CP013923.1	8	Achromobacter denitrificans
AM490508.1	8	Herbaspirillum seropedicae
CP000267.1	8	Rhodoferax ferrireducens
CP019239.1	8	Rhodoferax saidenbachensis
CP016603.1	8	Comamonas aquatica
JX017371.1	8	Comamonas
HG326930.1	8	Rhizobium pusense
HG326939.1	8	Agrobacterium genomosp
HG326926.1	8	Agrobacterium fabrum
HG326927.1	8	Agrobacterium genomosp
JX017367.1	8	Rhizobium
KX490616.1	8	Rhizobium leguminosarum
AP014685.1	8	Bradyrhizobium diazoefficiens
CP002826.1	8	Oligotropha carboxidovorans
CP000319.1	8	Nitrobacter hamburgensis
CP000115.1	8	Nitrobacter winogradskyi
AP014854.2	8	Blastochloris viridis
AJ459585.1	8	Mesorhizobium
AJ459586.2	8	Mesorhizobium
JX017367.1	8	Rhizobium
KX490616.1	8	Rhizobium leguminosarum
HG326934.1	8	Agrobacterium tumefaciens
HG326941.1	8	Agrobacterium genomosp
NC 004463.1	8	Bradyrhizobium japonicum
AJ459590.1	8	Mesorhizobium
JN636822.1	8	Rhizobium
NC 003047.1	8	Sinorhizobium meliloti

Supplementary Information 2: Steps to generate *glnA* database and taxonomy file to work with Qiime

To make a database from Microbial Genome Database (MGBD) http://mbgd.genome.ad.jp, we need to follow the following steps:

Step 1:
From MBDG downloaded sequences for a glnA bacterial sequences and save it in a FASTA file called glnA.fa. Then generate a new FASTA file that only has organism names

```
awk '/>/{gsub("^>","",$0);gsub(":.*","",$0);$0=">F"++i"_"$0}1' glnA.fa >
glnA_symbol.fa
```

Step 2: Get the organism IDs and store them in IDs.csv

```
awk '/>/{a=$0;sub("^>F[0-9]+_","",$0); print a","$0}' glnA_symbol.fa >
IDs.csv
```

Step 3: Get mapping to NCBI gids. For this purpose, download http://mbgd.genome.ad.jp/dist/mbgd_2016-01/mbgd_2016-01_gene.gz from http://mbgd.genome.ad.jp/htbin/view_arch.cgi

```
awk -F"\t" 'BEGIN{while((getline k <
"IDs.csv")>0){split(k,a,",");m[a[1]]=a[2];o[a[2]]=1}}{if(o[$1]){if(length($
12)>1){o[$1]=$12}}}END{for(i in m){print i","o[m[i]]}}' mbgd_2016-01_gene >
mapping.csv
```

Step 4: Use the mapping file to change the FASTA headers

```
awk 'BEGIN{while((getline k <
"mapping.csv")>0){split(k,a,",");m[a[1]]=a[2]}}/^>/{b=$1;gsub("_.*","",b);$
0=b"_"m[$1]}1' glnA_symbol.fa > glnA_reference_db.fa
```

Step 5: Extract a comma-delimited IDs along with GIDs

```
awk -F"_" 'BEGIN{print "ID,GID"}/^>/{gsub(">","",$0);print $1","$2}'
glnA_reference_db.fa > glnA_reference_db_id.csv
```

Steps to generate amoA database and taxonomy file to work with Qiime

To make a database from Fungene (http://fungene.cme.msu.edu/), download amoA sequences to amoA.fa.

Step 1: Format so that we have >ID_GID

```
awk '/^>/{gsub(" .*","",$0); gsub(">","",$0); $0=">F"++i"_"$0}1' amoA.fa >
amoA_reference_db.fa
```

Step 2: Extract a comma-delimited IDs along with GIDs so that we can use an R package (taxize) to get the complete taxonomy out

```
awk -F"_" 'BEGIN{print "ID,GID"}/^>/{gsub(">","",$0);print $1","$2}'
amoA_reference_db.fa > amoA_reference_db_id.csv
```

R script to generate taxonomy files (common to both glnA and amoA):

```
library(rentrez)
#Load the mapping table up
mapping_table<-read.csv("glnA_reference_db_id.csv",row.names=1,header=T)
#extract gids
gids<-mapping_table$GID
taxa_levels<-NULL
for(i in seq(1:length(gids))){
  print(paste("Processing",i,"/",length(gids)))
  tmp<-tryCatch(paste(XML::xpathSApply(entrez_fetch(db="taxonomy",
id=entrez_summary(db="nucleotide", id=gids[i])$taxid,rettype="xml",
parsed=TRUE), "//LineageEx/Taxon/ScientificName",
```

```
XML::xmlValue),collapse=";"),error=function(e) "")
  if(is.null(taxa_levels)){taxa_levels<-tmp}else{taxa_levels<-
c(taxa_levels,tmp)}
}
taxa_levels[taxa_levels==""]<-"cellular organisms;unassigned"
data_to_write<-data.frame(ID=paste(rownames(mapping_table),
mapping_table[,1],sep="_"),Taxa=taxa_levels)
write.table(data_to_write,"glnA_reference_db.tax",sep="\t",
row.names=F,col.names=F,quote=F)
```

Supplementary Information 3: List of *nirS* sequences used to re-design qPCR primers

Phylum	Class	Order	Familly	Genus Species/ strain name	Accesion number (Nucleotide/Protein)
Aquificae	Aquificae	Aquificales	Aquificaceae	*Hydrogenobacter thermophilus TK-6*	CP002221/ADO44561
Aquificae	Aquificae	Aquificales	Hydrogenothermaceae	*Persephonella marina EX-H1*	CP001230/ACO04194
Aquificae	Aquificae	Aquificales	Hydrogenothermaceae	*Sulfurihydrogenibium sp. YO3AOP1*	CP001080/ACD66103
Chloroflexi	Anaerolineae	Anaerolineales	Anaerolineaceae	*Thermanaerothrix daxensis*	LGKO01000004/KPL83147
Chloroflexi	Anaerolineae	Anaerolineales	Anaerolineaceae	*Anaerolinea thermophila*	AP012029/BAJ64797
Chloroflexi	Anaerolineae	Anaerolineales	Anaerolineaceae	*Bellilinea caldifistulae*	LGHJ01000006/KPL78197
Deinococcus-Thermus	Deinococci	Thermales	Thermaceae	*Oceanithermus profundus DSM 14977*	CP002361/ADR37584
Deinococcus-Thermus	Deinococci	Thermales	Thermaceae	*Thermus oshimai JL-2*	CP003249/AFV76102
Deinococcus-Thermus	Deinococci	Thermales	Thermaceae	*Thermus scotoductus SA-01*	CP001962/ADW22372
Deinococcus-Thermus	Deinococci	Thermales	Thermaceae	*Thermus thermophilus*	FN666415/CBJ34318
Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus pneumoniae*	CLJU01000018/CJL48784
Proteobacteria	Acidithiobacillia	Acidithiobacillales	Acidithiobacillaceae	*Acidithiobacillales bacterium SG8 45*	LJTU01000018/KPK11659.1
Proteobacteria	Acidithiobacillia	Acidithiobacillales		*Acidithiobacillales bacterium SM23 46*	LJUK01000134/KPK69514
Proteobacteria	Alphaproteobacteria			*Polymorphum gilvum SL003B-26A1*	CP002568/ADZ69009
Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	*Bradyrhizobium oligotrophicum*	AP012603/BAM92788
Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	*Prosthecomicrobium hirschii*	LJYW01000001/KPL55974
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Litoreibacter	*Litoreibacter arenae DSM 19593*	AONI01000011/EPX78564
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Aliiroseovarius crassostreae*	LKBA01000024/KPN61790

Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Dinoroseobacter shibae DFL 12*	CP000830/ABV94913
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Labrenzia aggregata*	CXST01000004/CTQ46751
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Labrenzia alba*	CXWC01000007/CTQ69729
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Paracoccus denitrificans*	U75413/AAB17878
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Paracoccus denitrificans PD1222*	CP000489/ABL70574
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Paracoccus pantotrophus*	AJ401462/CAC03621
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Paracoccus sanguinis*	JRKT01000018/KGJ14663
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Paracoccus sp. MKU1*	LLWQ01000193/KRW95054
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Paucibacter aquatile strain CR182*	NZ/POSP01000003.1/WP/10276 9553.1
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Pseudovibrio axinellae*	LMCB01000159/KZL05178
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Pseudovibrio sp. Ad14*	LMCD01000020/KZL07807
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Pseudovibrio sp. Ad13*	LMCC01000003/KZK86619
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Pseudovibrio sp. Ad5*	LMCH01000004/KZL01520
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Rhodobacteraceae bacterium HLUCCO07*	LJSU01000028/KPP83431
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Roseobacter denitrificans OCh 114*	CP000362/ABG31193
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Roseobacter litoralis Och 149*	CP002623/AEI95094
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Ruegeria atlantica*	CYPU01000030/CUH47623
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Ruegeria pomeroyi DSS-3*	CP000032/AAV97354
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	*Thioclava sp. DT23-4*	AUNB01000023/KEO60092
Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	*Azospirillum brasilense*	CP012918/ALJ39431
Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	*Caenispirillum salinarum*	ANHY01000019/EKV27616
Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	*Magnetospirillum gryphiswaldense MSR-1*	CU459003/CAM74253
Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	*Magnetospirillum magneticum AMB-1*	AP007255/BAE52969
Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	*Magnetospirillum marisnigri*	LWQT01000048/OAN51092
Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	*Magnetospirillum sp. SO-1*	AONQ01000069/EME68346
Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	*Thalassospira lucentensis*	LPVY01000002/KZB68956

Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	*Thalassospira permensis NBRC 106175*	AUNC01000001/KEO59842
Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	*Thalassospira xiamenensis*	LPXM01000045/KZD06045
Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	*Thalassospira xiamenensis M-5*	AMRQ01000004/EKF13195
Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	*Achromobacter sp. DBTN3*	GU122964/ACY92300
Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	*Bordetella petrii*	AM902716/CAP44402
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Cupriavidus alkaliphilus*	FMAD01000008/SCB26603
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Cupriavidus metallidurans CH34*	CP000352/ABF10044
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Cupriavidus nantongensis*	CP014845/AMR81885
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Cupriavidus necator*	LMVF01000024/KUE88587
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Cupriavidus necator N-1*	CP002878/AEI81194
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Cupriavidus pauculus*	GQ504717/ADN28073
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Cupriavidus sp. HMR-1*	ANKP01000138/EKZ97934
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Cupriavidus sp. SK-4*	JFJW01000102/EYS87961
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Cupriavidus taiwanensis LMG 19424*	CU633750/CAQ72627
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Ralstonia eutropha H16*	X91394/CAA62740
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Ralstonia eutropha H16*	AM260480/CAJ97059
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Ralstonia pickettii DTP0602*	CP006668/AGW94840
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Acidovorax sp. 2FB7*	AY078273/AAL86942
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Acidovorax sp. JS42*	CP000539/ABM42092
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Alicycliphilus sp. B1*	BBSJ01000051/GAO23488
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Brachymonas denitrificans*	DQ865925/ABI96831
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Comamonas denitrificans*	DQ865926/ABI96832
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Curvibacter sp. PAE-UM*	LKCX01000031/KRH99151/
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Diaphorobacter polyhydroxybutyrativorans*	CP016278/ASI68082
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Diaphorobacter sp. J5-51*	JSYI01000121/KLR56966

Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Pelomonas sp. Root1444*	LMFP01000027/KQY82451
Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified Burkholderiales	*Leptothrix cholodnii SP-6*	CP001013/ACB33718
Proteobacteria	Betaproteobacteria	Burkholderiales		*Rubrivivax gelatinosus IL144*	AB536930/BAJ16230
Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	*Methylobacillus sp. MM2*	LXTQ01000001/OAJ71930
Proteobacteria	Betaproteobacteria	Neisseriales	Chromobacteriaceae	*Gulbenkiania indica*	CYHA01000002/CUA82351
Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae	*Chromobacterium sp. LK1*	LDUI01000036/KMN32840
Proteobacteria	Betaproteobacteria	Nitrosomonadales	Gallionellaceae	*Sideroxydans lithotrophicus ES-1*	CP001965/ADE11367
Proteobacteria	Betaproteobacteria	Nitrosomonadales	Gallionellaceae	*Sulfuricella denitrificans skB26*	BAFJ01000007/GAB72194
Proteobacteria	Betaproteobacteria	Nitrosomonadales	Sterolibacteriaceae	*Sulfuritalea hydrogenivorans sk43H*	AP012547/BAO30669
Proteobacteria	Betaproteobacteria	Nitrosomonadales	Thiobacillaceae	*Thiobacillus denitrificans*	LDUG01000028/KVW95088
Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	*Aromatoleum aromaticum*	CR555306/CAI06598
Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	*Azoarcus evansii*	AY078269/AAL86938
Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	*Azoarcus sp. CIB*	CP011072/AKU13501
Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	*Azoarcus sp. KH32C*	AP012304/BAL25833
Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	*Dechloromonas aromatica RCB*	CP000089/AAZ48052
Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	*Dechloromonas denitrificans*	LODL01000010/KXB31613
Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	*Propionivibrio sp. S1*	FLQY01000367/SBT10752
Proteobacteria	Betaproteobacteria	Rhodocyclales	Zoogloeaceae	*Thauera aminoaromatica S2*	AMXD01000019/ENO87324
Proteobacteria	Betaproteobacteria	Rhodocyclales	Zoogloeaceae	*Thauera aromatica*	AY078256/AAL86925
Proteobacteria	Betaproteobacteria	Rhodocyclales	Zoogloeaceae	*Thauera chlorobenzoica*	AY078261/AAL86930
Proteobacteria	Betaproteobacteria	Rhodocyclales	Zoogloeaceae	*Thauera humireducens*	CP014646/AMO36595
Proteobacteria	Betaproteobacteria	Rhodocyclales	Zoogloeaceae	*Thaurea sp. 28*	AMXA01000008/ENO93245
Proteobacteria	Betaproteobacteria	Rhodocyclales	Zoogloeaceae	*Thaurea sp. 27*	AMXB01000001/ENO83228
Proteobacteria	Betaproteobacteria	Rhodocyclales	Zoogloeaceae	*Thaurea linaloolentis 47Lol*	AMXB01000080/EN085266
Proteobacteria	Betaproteobacteria	Rhodocyclales	Zoogloeaceae	*Thaurea Terpenica*	ATJV01000048/EPZ16013

Proteobacteria	Deltaproteobacteria			*Deltaproteobacteria bacterium GWC2 55 46*	LVEI02000001/KYK97987
Proteobacteria	Delta/Epsilon subdivision	Epsilonproteobacteria		*Nitratiruptor sp. SB155-2*	AP009178/BAF70899
Proteobacteria	Gammaproteobacteria	Acidiferrobacterales	Acidiferrobacteraceae	*Sulfuricaulis limicola*	AP014879/BAV32586
Proteobacteria	Gammaproteobacteria	Acidiferrobacterales	Acidiferrobacteraceae	*Sulfurifustis variabilis*	AP014936/BAU46932
Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	*Marinobacter hydrocarbonoclasticus*	KT877013/AMO65337
Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	*Marinobacter sp. AK21*	ANIE01000004/KEF31766
Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	*Marinobacter sp. EhN04*	LXYN01000008/OAN94133
Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	*Marinobacter sp. EhC06*	LXYO01000033/OAN89896
Proteobacteria	Gammaproteobacteria	Alteromonadales	Colwelliaceae	*Colwellia psychrerythraea 34H*	CP000083/AAZ25602
Proteobacteria	Gammaproteobacteria	Chromatiales	Ectothiorhodospiraceae	*Thioalkalivibrio nitratireducens DSM 14787*	CP003989/AGA33316
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	*Enterobacter cloacae*	FKIY01000075/SAJ30203
Proteobacteria	Gammaproteobacteria	Nevskiales	Sinobacteraceae	*Steroidobacter denitrificans*	CP011971/AMN47484
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Hahellaceae	*Hahella chejuensis KCTC 2396*	CP000155/ABC31120
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	*Halomonas alimentaria*	FJ686149/ACN97377
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	*Halomonas campaniensis*	FJ686150/ACN97378
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	*Halomonas cerina*	GQ384052/ACV88083
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	*Halomonas chromatireducens*	CP014226/AMD00733
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	*Halomonas denitrificans*	GQ384047/ACV88078
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	*Halomonas fontilapidosi*	FJ686147/ACN97375
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	*Halomonas halodenitrificans*	FJ686155/ACN97381
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	*Halomonas nitroreducens*	FJ686148/ACN97376
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	*Halomonas shengliensis*	FJ686158/ACN97383
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	*Halomonas ventosae*	FJ686160/ACN97385
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	*Halomonas sp. PBN3*	AXCA01000197/ERS82918

Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oleiphilaceae	*Oleiphilus sp. HI0043*	LWEY01001511/KZY30905
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oleiphilaceae	*Oleiphilus sp. HI0050*	LWFB01001229/KZY45670
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oceanospirillaceae	*Oleispira antarctica RB-8*	FO203512/CCK76416
Proteobacteria	Gammaproteobacteria			*Sedimenticola thiotaurini*	CP011412/AKH19023
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas aeruginosa*	LLTT01000034/KSQ29041
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas aeruginosa VRFPA08*	AZHU01000087/ETD48699
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas aeruginosa VRFPA04*	CP008739/AID86920
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas aeruginosa SD9*	AMVN01000080/OPF35317
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas balearica*	LONE01000028/KXO68010
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas bauzanensis*	JFHS01000001/EZQ19309
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas brassicacearum*	CP012680/ALQ04046
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas brenneri*	LVWZ01000003/OAE17393
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas chloritidismutans*	AJ884572/CAI56317/
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas chlororaphis*	CP011020/AKJ99904
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas corrugata*	CP014262/AOE60879
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas flexibilis*	JTAK01000001/KHO66430
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas fluorescens*	AF197466/AAG34381
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas frederiksbergensis*	JQGJ02000001/KHK64801
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas kilonensis*	JZXC01000048/KA04128
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas libanensis*	LT629699/SDK81826
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas lini*	LFQO01000085/KNH44064
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas mandelii JR-1*	CP005960/AHZ67536
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas silesiensis*	CP014870/ANJ56620
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas marginalis*	LKEG01000046/OAJ47348
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas sp. YS-1p*	JPYP01000033/KGD88478
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas stutzeri*	CP002881/AEJ06785

Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas stutzeri*	LDWB01000026/KOR09159
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas synxantha*	LT629786/SDU39534
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas syringae*	LFQK01000013/KNH28426
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas thivervalensis*	LT629691/SDF46790
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas veronii*	JYLL01000006/KRP79617
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas xanthomarina*	MDEM01000020/OCX24666
Proteobacteria	Gammaproteobacteria	Thiotrichales	Thiotrichaceae	*Beggiatoa sp. PS*	ABBZ01001128/EDN67831
Proteobacteria	Gammaproteobacteria	Thiotrichales	Thiotrichaceae	*Thioploca ingrica*	AP014633/BAP57172
Proteobacteria	Hydrogenophilalia	Hydrogenophilales	Hydrogenophilaceae	*Tepidiphilus thermophilus*	CYHH01000010/CUB07681
Unclassified	*Candidatus Accumulibacter*	JDVG02000709/KFB70389			
Unclassified				*Candidatus Competibacter denitrificans*	CBTJ020000111/CDI04496
				Candidatus Kuenenia stuttgartiensis	CT573071/CAJ74898
				Candidatus Methylomirabilis oxyfera	FP565575/CBE69462
				Candidatus Thiomargarita nelsonii	LUTY01001461/OAD21685
				Dechlorosoma suillum PS	CP003153/AEV26447