THE UNIVERSITY OF WARWICK

Library Declaration and Deposit Agreement

1. STUDENT DETAILS

Please complete the following:

Full name: ……

University ID number: ……

2. THESIS DEPOSIT

2.1 I understand that under my registration at the University, I am required to deposit my thesis with the University in BOTH hard copy and in digital format. The digital version should normally be saved as a single pdf file.

2.2 The hard copy will be housed in the University Library. The digital version will be deposited in the University's Institutional Repository (WRAP). Unless otherwise indicated (see 2.3 below) this will be made openly accessible on the Internet and will be supplied to the British Library to be made available online via its Electronic Theses Online Service (EThOS) service.
[At present, theses submitted for a Master's degree by Research (MA, MSc, LLM, MS or MMedSci) are not being deposited in WRAP and not being made available via EthOS. This may change in future.]

2.3 In exceptional circumstances, the Chair of the Board of Graduate Studies may grant permission for an embargo to be placed on public access to the hard copy thesis for a limited period. It is also possible to apply separately for an embargo on the digital version. (Further information is available in the *Guide to Examinations for Higher Degrees by Research*.)

2.4 *If you are depositing a thesis for a Master's degree by Research, please complete section (a) below. For all other research degrees, please complete both sections (a) and (b) below:*

(a) Hard Copy

I hereby deposit a hard copy of my thesis in the University Library to be made publicly available to readers (please delete as appropriate) EITHER immediately OR after an embargo period of ……..... months/years as agreed by the Chair of the Board of Graduate Studies.

I agree that my thesis may be photocopied. YES / NO (*Please delete as appropriate*)

(b) Digital Copy

I hereby deposit a digital copy of my thesis to be held in WRAP and made available via EThOS.

Please choose one of the following options:

EITHER My thesis can be made publicly available online. YES / NO (*Please delete as appropriate*)

OR My thesis can be made publicly available only after…..[date] (Please give date)
YES / NO (*Please delete as appropriate*)

OR My full thesis cannot be made publicly available online but I am submitting a separately identified additional, abridged version that can be made available online.
YES / NO (*Please delete as appropriate*)

OR My thesis cannot be made publicly available online. YES / NO (*Please delete as appropriate*)

3. GRANTING OF NON-EXCLUSIVE RIGHTS

Whether I deposit my Work personally or through an assistant or other agent, I agree to the following:

Rights granted to the University of Warwick and the British Library and the user of the thesis through this agreement are non-exclusive. I retain all rights in the thesis in its present version or future versions. I agree that the institutional repository administrators and the British Library or their agents may, without changing content, digitise and migrate the thesis to any medium or format for the purpose of future preservation and accessibility.

4. DECLARATIONS

(a) I DECLARE THAT:

- I am the author and owner of the copyright in the thesis and/or I have the authority of the authors and owners of the copyright in the thesis to make this agreement. Reproduction of any part of this thesis for teaching or in academic or other forms of publication is subject to the normal limitations on the use of copyrighted materials and to the proper and full acknowledgement of its source.

- The digital version of the thesis I am supplying is the same version as the final, hard-bound copy submitted in completion of my degree, once any minor corrections have been completed.

- I have exercised reasonable care to ensure that the thesis is original, and does not to the best of my knowledge break any UK law or other Intellectual Property Right, or contain any confidential material.

- I understand that, through the medium of the Internet, files will be available to automated agents, and may be searched and copied by, for example, text mining and plagiarism detection software.

(b) IF I HAVE AGREED (in Section 2 above) TO MAKE MY THESIS PUBLICLY AVAILABLE DIGITALLY, I ALSO DECLARE THAT:

- I grant the University of Warwick and the British Library a licence to make available on the Internet the thesis in digitised format through the Institutional Repository and through the British Library via the EThOS service.

- If my thesis does include any substantial subsidiary material owned by third-party copyright holders, I have sought and obtained permission to include it in any version of my thesis available in digital format and that this permission encompasses the rights that I have granted to the University of Warwick and to the British Library.

5. LEGAL INFRINGEMENTS

I understand that neither the University of Warwick nor the British Library have any obligation to take legal action on behalf of myself, or other rights holders, in the event of infringement of intellectual property rights, breach of contract or of any other right, in the thesis.

Please sign this agreement and return it to the Graduate School Office when you submit your thesis.

Student's signature: .....…… Date: .....

Microbial Ecology of Anammox Bacteria in Estuarine and Oxygen Minimum Zone Environments

Simon Williams
BSc (Hons)
MSc (Hons)

Submitted for the Degree of Doctor of Philosophy

School of Life Sciences
University of Warwick

August 2013

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost I would like to thank my supervisor, Dr Kevin Purdy, for his help, guidance and patience during this project and for giving me the opportunity to study at the University of Warwick. His support has been invaluable throughout and is very much appreciated. I would also like to thank our collaborators at Queen Mary University, Professor Mark Trimmer, Dr Myrsini Chronopoulou and Jimmy Pritchard for their help and engaging discussions throughout this project (and for the mud).

I also wish to acknowledge the help and assistance of the crew and passengers of the RRS Discovery, Cruise D373 during sampling in the ETNP. I would especially like to thank Dr Laura Bristow without whose advice, experience and friendship I would have most likely turned (even more) mad during my time at sea.

I also wish to thank Josh Neufeld (University of Waterloo), Jakob Zopfi (University of Lausanne) and Amano Teruki (Kyoto University) and their respective groups for the provision of anammox DNA which has been invaluable to achieving the aims of this project.

I would also like to thank all the past and present members of the Purdy group and lab C126 for their friendship and aid during this project and for making the seemingly endless PCRs and DNA extractions that bit more bearable. A special thank you goes out to Viktoria for being my friend from the start and to EMM2 and Scoop for the rants and general joviality. Thanks also to everyone else in Life Sciences who have helped me during my research. It has truly been a pleasure to get to know and work with you all.

I would further like to thank the Animals of Farthing Wood for their kinship and TV watching abilities. Besides the odd Bolognese explosion and sore head it has been an honour to share a house with you.

Much love and thanks also to my family who, despite having no idea what I've been doing this last four years, have nevertheless been supportive and willing to listen.

Last but by no means least a big thank you to Kat, for your support, care and cuddles and for graciously being my landlady and PA in the weeks prior to submission.

This project was funded entirely by NERC.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. I declare that this thesis has been composed by myself and that all work within it is my own unless otherwise stated. This thesis has not been submitted in any previous application for any degree at this or any other institution.

Simon Williams

Abstract

Anammox (anaerobic ammonium oxidation) is an environmentally significant process with great importance for global biogeochemical cycles. This process is mediated by a unique suite of phylogenetically distinct chemolithoautotrophic bacteria which demonstrate novel physiological and metabolic characteristics. However, despite the importance of these organisms, there is still much which is poorly understood about them, specifically the diversity and distribution of these bacteria and their controlling environmental factors. Furthermore, genomic studies and observations from the field suggest that anammox bacteria may have a far greater metabolic diversity than previously thought, suggesting that the current understanding of these organisms is incomplete. This study aimed to elucidate these aspects of the ecology of anammox bacteria in estuarine and OMZ (oxygen minimum zone) environments. A clear community shift was observed in estuarine environments from *Ca.* Brocadia spp. dominated freshwater sites to *Ca.* Scalindua spp. dominated marine sites. The OMZ was dominated by *Ca.* Scalindua spp. though diversity within this clade was observed between organisms in the upper oxycline and those within the core of the OMZ. Microcosm experiments amended with organic substrates suggested that some anammox organisms (namely *Ca.* Brocadia spp., *Ca.* Jettenia spp. and *Ca.* Kuenenia spp.) may have the ability to assimilate carbon directly from organic substrates such as dimethylamine and urea. However, these data were inconclusive and further investigations are required to prove or disprove the hypothesis that anammox bacteria can utilise organic substrates. Nevertheless, this study improves the understanding of the ecology of anammox organisms in estuarine and OMZ environments, providing an unprecedented depth of data as to the diversity and distribution and unique insights into potentially novel metabolic capabilities of these organisms.

Abbreviations

AMZ	Anoxic Marine Zone
Anammox	Anaerobic Ammonium Oxidation
ATU	Allylthiourea
BOD	Biological Oxygen Demand
bp	(Nucleotide) Base pair
BSL	Below Sea-Level
Ca.	*Candidatus*
CCA	Canonical Correspondence Analysis
CTD	Instrument to measure conductivity, temperature in depth in oceanography
DGGE	Denaturing Gradient Gel Electrophoresis
DIN	Dissolved Inorganic Nitrogen
DMA	Di-methylamine
DNRA	Dissimilatory Nitrate Reduction to Ammonium
DOC	Dissolved Organic Carbon
DOM	Dissolved Organic Matter
DON	Dissolved Organic Nitrogen
ETNP	Eastern Tropical North Pacific
ETSP	Eastern Tropical South Pacific
FISH	Fluorescence *In Situ* Hybridisation
GC	Gas Chromatography/Chromatograph

HGT	Horizontal Gene Transfer
HPLC	High-performance Liquid Chromatography
IC	Ion Chromatography/Chromatograph
IPT	Isotope Pairing Technique
IRMS	Isotope-ratio Mass Spectrometry
MBM	Medway Bridge Marina
N_{def}	Nitrogen Deficit
ODZ	Oxygen Deficient Zone
OFN	Oxygen Free Nitrogen
OMZ	Oxygen Minimum Zone
OrgN	Organic Nitrogen
OTU	Operational Taxonomic Unit
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
qPCR	Quantitative PCR
Rep.	Replicate
RLIC	Read Length Incremental Clustering
SE	Standard Error
SIP	Stable Isotope Probing
sp.	Species (singular)
spp.	Species (plural)

TMA	Tri-methylamine

1. Introduction

1.1. Global Nitrogen Cycling

1.1.1. The Nitrogen Cycle

Nitrogen is an essential element for all life on this planet (Redfield, 1934; Francis, *et al.*, 2007) and hence is an important factor in biogeochemical cycling. Furthermore, it is also the primary nutrient-limiting factor for primary production in many environments (Seitzinger, *et al.*, 2002). It is therefore essential that a lucid understanding of the nitrogen cycle be obtained for the benefit of environmental and life sciences as a whole.

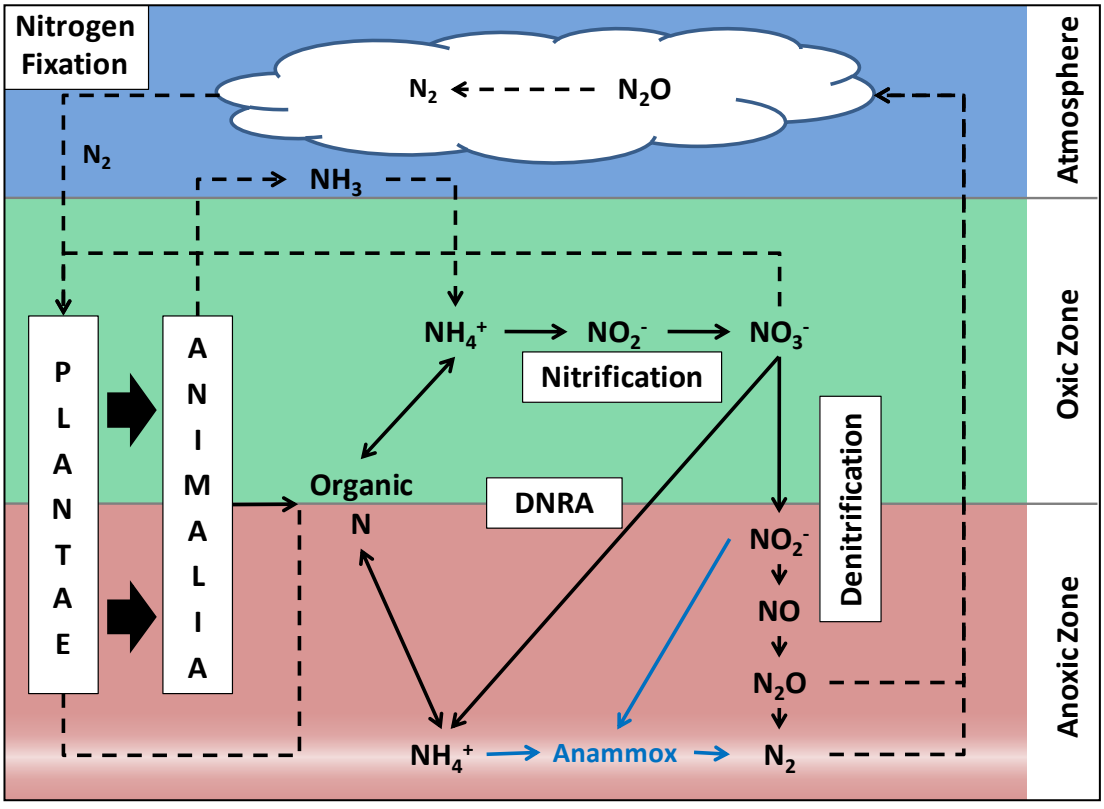


Figure 1.1: Diagram showing the currently accepted view of the nitrogen cycle. The anammox reaction (in blue) can be seen to play an important role in the anaerobic nitrogen cycle, providing an alternative route to N_2 production (other than denitrification) and a route to the removal of NH_4^+ (other than assimilation).

The nitrogen cycle (Manahan, 2005; Francis, *et al.*, 2007) can be divided into three broad stratifications: atmospheric, aerobic and anaerobic environments (Figure 1.1). Atmospheric di-nitrogen (N_2) gas is fixed into organic matter to fulfil various metabolic roles. Organic nitrogen (org-N) is released back to the environment either

through death or excretion and is, ultimately, converted to ammonium (NH_4^+) through putrification. In aerobic environments, NH_4^+ is oxidised to nitrate (NO_3^-) and nitrite (NO_2^-) via nitrification. NO_3^- and NO_2^- can then be reduced anaerobically to N_2 and nitrous oxide (N_2O) via denitrification, releasing these compounds back into the atmosphere. Until recently NH_4^+ was believed to be inert in anaerobic environments and it was not until the discovery of anaerobic ammonium oxidation (anammox) that this assumption was falsified (van de Graaf, *et al.*, 1995). Anammox provides an alternative pathway to N_2 production, using NO_2^- as an oxidising agent (Francis, *et al.*, 2007). Dissimilatory nitrate reduction to ammonium (DNRA) is a further element of the N-cycle, providing a novel source of NH_4^+ in anaerobic environments (An and Gardner, 2002; Dong, *et al.*, 2011).

1.1.2. Anammox, Denitrification and Nitrogen Cycling – N_2 loss

Gaseous N_2 represents 78% of the planet's atmosphere (Francis, *et al.*, 2007) which presents a significant source of nitrogen for fixation by primary producers. Thus it is important that a comprehensive understanding of the processes leading to the loss of fixed nitrogen from environments to the atmosphere is developed (Kuypers, *et al.*, 2006).

Prior to the discovery of the anammox process, denitrification was viewed as the only process producing N_2 in marine environments and one of the largest sinks for fixed nitrogen (Devol, 2003; Francis, *et al.*, 2007). Denitrification is the process by which NO_3^- is reduced sequentially to N_2 (Ward, *et al.*, 2007) as represented in Figure 1.1 and by the following equation:

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

(Devol, *et al.*, 2006)

Denitrification can couple either NO_x^- molecules to either N_2 or N_2O however a balanced reaction for the heterotrophic conversion of NO_3^- to N_2 can be represented as:

$$4NO_3^- + 5CH_2O + 4H^+ \rightarrow 2N_2 + 5CO_2 + 7H_2O$$

(Manahan, 2005)

Denitrification is a microbial process mediated by ~50 different genera of facultatively anaerobic, heterotrophic organisms (Francis, *et al.*, 2007) spanning the domains of bacteria, archaea and eukarya (Ward, *et al.*, 2007). These organisms are abundant in a wide range of anaerobic environments (Seitzinger, *et al.*, 2006).

However it had been evident for some time that our understanding of nitrogen loss from anaerobic systems was not complete. Richards (1965) noted that the amount of NH_4^+ found in anaerobic environments was much less than would be stoichiometrically expected because, if NH_4^+ was truly inert in these environments, it would steadily accumulate over time. Later Broda (1977) hypothesised, based on thermodynamic calculations, a mechanism for the conversion of NH_4^+ to N_2, mediated by a lithotrophic organism, suggesting that this could be the missing piece of the nitrogen-cycle. Despite such discrepancies being reported in the literature, it was not until much later that anammox was discovered in a wastewater reactor (Mulder, *et al.*, 1995) and the organisms responsible for this process identified (van de Graaf, *et al.*, 1995; Strous, *et al.*, 1999).

Anammox, in its most basic definition, is the oxidation of NH_4^+ to N_2 utilising NO_2^- as an electron (e^-) acceptor:

$$NH_4^+ + NO_2^- \rightarrow N_2 + 2H_2O$$

(Van de Graaf, *et al.*, 1996)

Anammox thus provided an alternative route to N_2 loss, potentially bypassing denitrification and dispelling the theory that this process was the only sink for fixed nitrogen in anaerobic environments (Devol, *et al.*, 2006; Francis, *et al.*, 2007). Following the discovery of anammox, much research was conducted investigating this process and evidence for it was soon found in a range of different environments (see section 1.2.1) but most notable in marine, coastal sediments and within the Oxygen Minimum Zones (OMZs). However, it was initially argued that anammox was environmentally insignificant, due to their slow growth rates (reviewed in Jetten, *et al.*, 2009), and that denitrification was the dominant nitrogen sink in these environments (Zehr and Ward, 2002; Ward, *et al.*, 2007). Indeed, theoretical calculations based on the Redfied ratio (Redfield, 1934) suggest that the maximum contribution of anammox to N_2 loss can only be 29% (Kuypers, *et al.*, 2006; Ward,

et al., 2007) though due to the diverse metabolisms of denitrifiers (Francis, *et al.*, 2007) anammox were likely to be outcompeted in most environments, posing interesting questions for the niche adaptation and environmental significance of these organisms.

This view was challenged when the contribution of anammox to global nitrogen cycling was investigated. This was made possible by the adaption of the Isotope Pairing Technique (IPT), devised by Nielsen (1992) for measuring denitrification rates, to include the anammox process (Thamdrup and Dalsgaard, 2002; Risgaard-Petersen, *et al.*, 2003). In IPT, samples are incubated with ^{15}N-labelled NO_x^- substrates and the amount of $^{29}N_2$ and $^{30}N_2$ is measured via mass spectrometry (Nielsen, 1992). The addition of experiments including $^{15}NH_4^+$ allows for the relative rate of anammox compared to denitrification to be measured as, with this substrate, $^{29}N_2$ can only be produced via the anammox reaction (Trimmer, *et al.*, 2003). Using this method, anammox was found to be a significant contributor to marine, continental shelf sediments, potentially contributing up to 60% of N_2 production (Thamdrup and Dalsgaard, 2002). Anammox was also shown to be significant in OMZs, producing 30-50% of N_2 in these environments (Devol, 2003). However, similar studies in estuarine environments (Trimmer, *et al.*, 2003; Dale, *et al.*, 2009) and stratified lakes (Hamersley, *et al.*, 2009) found the potential contribution of Anammox to be much lower (1 – 16.5% and 13% respectively).

However, it is likely that these measurements are not entirely accurate as this method does not (and cannot) account for the conversion of NO_3^- to NH_4^+ via DNRA (Trimmer, *et al.*, 2003) nor does it account for the potential for more diverse metabolisms of anammox bacteria (see section 1.2.2 and chapter 5). Thus, the potential contribution of anammox to the N-cycle may either be under- or over-estimated. DNRA is a problem for IPT methodologies as it could produce $^{15}NH_4^+$ from $^{15}NO_3^-$ which would lead to the acquisition of $^{30}N_2$ from the anammox process, thereby producing a "false denitrification" signal (Kartal, *et al.*, 2007a). Recently however, Dalsgaard, *et al.* (2012) succeeded in modelling the contribution of DNRA in OMZs to the "false denitrification" signal potentially obtained using this method and found that it was negligible. Regardless of these advances, our knowledge of the anaerobic nitrogen cycle based on stoichiometry and theoretical calculations do not

agree with observations from the field (Ward, *et al.*, 2007) and thus our understanding of these processes is still incomplete (Devol, 2003). Nevertheless, it is widely agreed that anammox is a significant process in these environments and is an essential component of our understanding of the global nitrogen and carbon cycles (Francis, *et al.*, 2007). As such it is imperative that the global significance of anammox is further quantified and the factors which regulate this process elucidated (Kuypers, *et al.*, 2006).

1.2. Anammox Microbiology and Ecology

1.2.1. Phylogeny, Diversity and Distribution

Shortly after the discovery of the anammox reaction, it was shown that anammox was a biologically mediated process (van de Graaf, *et al.*, 1995). These organisms were shown to be chemolithoautotrophic, solely using the redox of inorganic nitrogen compounds (*i.e.* NH_4^+ and NO_2^-) to create the energy in order to fix CO_2 into biomass (Van de Graaf, *et al.*, 1996). Strous, *et al.* (1999) later identified these organisms as being evolutionary distinct bacteria belonging to the order *Planctomycetales.* Recently however anammox have been assigned to their own order, *Brocadiales*, (Figure 1.2) within the phylum Planctomycetes (Gori, *et al.*, 2011).

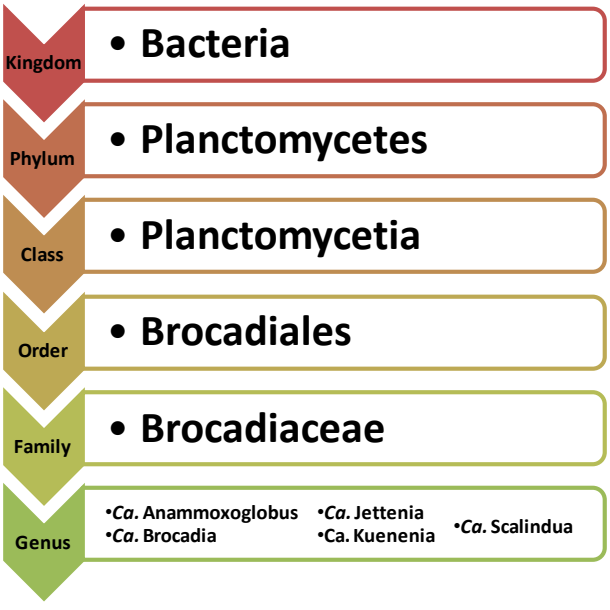


Figure 1.2: Classification of known anammox bacteria.

To date, five candidate distinct genera of anammox bacteria have been identified (Figure 1.3); *Ca.* Anammoxoglobus, *Ca.* Brocadia, *Ca.* Jettenia, *Ca.* Kuenenia and *Ca.* Scalindua (van Niftrik and Jetten, 2012). These organisms belong to a monophyletic clade within the phylum *Planctomycetes*, based on 16S rRNA gene diversity (van Niftrik and Jetten, 2012). All known anammox organisms have been shown to belong strictly to this clade. However despite this, the evolutionary history of anammox bacteria is relatively unknown and there are a number of theories as to their origins (van Niftrik and Jetten, 2012), although it does seem evident that anammox organisms are ancient (Fuerst, 2005) and thus may have been important in the biogeochemical cycles of the early earth and the formation of an N_2 rich atmosphere.

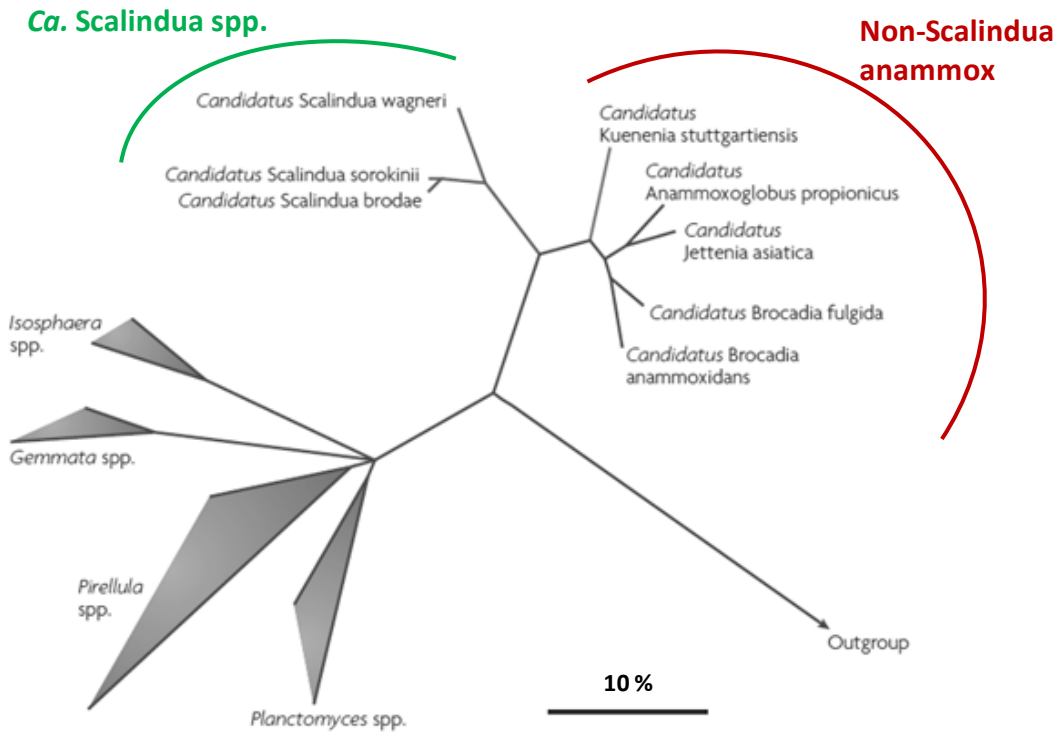


Figure 1.3: 16S rRNA phylogeny of anammox bacteria. Anammox 16S rRNA sequences cluster separately from other Planctomycetes. Ca. Scalindua spp. are distinct from other members of the anammox clade. Figure adapted from (Kuenen, 2008).

The first anammox organisms to be found in the natural environment were observed in the Black Sea and belonged to the *Ca.* Scalindua genus (Kuypers, *et al.*, 2003). Since then anammox have been found in a number of different marine environments, apparently dominated by *Ca.* Scalindua spp.; these include marine sediments (Dalsgaard and Thamdrup, 2002; Rysgaard, *et al.*, 2004; Engstrom, *et al.*, 2005;

Amano, *et al.*, 2007; Rich, *et al.*, 2008; Engstrom, *et al.*, 2009), OMZs (Hamersley, *et al.*, 2007; Woebken, *et al.*, 2008; Galan, *et al.*, 2009; Galan, *et al.*, 2012), estuarine sediments (Trimmer, *et al.*, 2003; Risgaard-Petersen, *et al.*, 2004; Dale, *et al.*, 2009; Nicholls and Trimmer, 2009), particulate matter in oceanic water columns (Hamersley, *et al.*, 2007; Jaeschke, *et al.*, 2007), arctic sea ice (Rysgaard and Glud, 2004), hydrothermal vents (Byrne, *et al.*, 2008), mangrove forests (Amano, *et al.*, 2011), marine sponges (Mohamed, *et al.*, 2010), and other anoxic marine environments and upwelling regions (Dalsgaard, *et al.*, 2003; Kuypers, *et al.*, 2003; Kuypers, *et al.*, 2005; Woebken, *et al.*, 2007). A greater diversity of anammox bacteria have also been found in numerous terrestrial environments including; the freshwater extents of river estuaries (Dale, *et al.*, 2009), stratified lakes (Schubert, *et al.*, 2006; Hamersley, *et al.*, 2009; Yoshinaga, *et al.*, 2011), wetland environments (Penton, *et al.*, 2006; Humbert, *et al.*, 2012), hot springs (Jaeschke, *et al.*, 2009a), petroleum reservoirs (Li, *et al.*, 2010a), anaerobic soils (Humbert, *et al.*, 2010), groundwater (Clark, *et al.*, 2008; Moore, *et al.*, 2011) and paddy fields (Sato, *et al.*, 2012). As anammox bacteria are found to be present and active in an increasing number of environmentally diverse ecosystems, it appears as if anammox may be ubiquitous in anaerobic environments (Francis, *et al.*, 2007; Kartal, *et al.*, 2008)[*]. However, to date, no anammox organisms has been successfully isolated and grown in pure culture and as such, all anammox organisms are classified as *Candidatus* organisms (Jetten, *et al.*, 2009) although highly enriched cultures are obtainable after considerable effort (Jetten, *et al.*, 2005).

1.2.2. Genetics

Despite the absence of cultured anammox organisms, successful attempts have been made to sequence anammox genomes which have significantly improved our understanding of these organisms. Strous, *et al.* (2006) were the first to sequence an anammox genome, *Ca.* Kuenenia stuttgartiensis, from a community metagenome, but were unable to obtain the complete genome for this organism. Strous, *et al.* (2006) found genes encoding the entire coenzyme-A (CoA) pathway of carbon assimilation as well as novel genes encoding for proteins for the synthesis and subsequent oxidation of N_2H_4, suggesting that previous hypotheses as to anammox

[*] For a more comprehensive review of anammox bacteria and diversity in the environment, especially in relation to estuaries and OMZs, please refer to chapter 5.

metabolism (*i.e.* that N_2H_4 was an important intermediate to the anammox process and that anammox bacteria are autotrophs, fixing CO_2 via the CoA pathway) may be correct (van de Graaf, *et al.*, 1997; Schalk, *et al.*, 2000). *Ca.* K. stuttgartiensis was also shown to contain genes encoding for the first steps of the denitrification pathway, narG (nitrate::nitrite oxidoreductase) and nirS (nitrite::nitric oxide oxidoreductase). Similar, subsequent studies have acquired metagenomes for *Ca.* Brocadia fulgida (Gori, *et al.*, 2011), *Ca.* Jettenia asiatica (Hu, *et al.*, 2012) and *Ca.* Scalindua profunda (van de Vossenberg, *et al.*, 2012). These studies reveal a greater potential metabolic diversity of these organisms than previously thought (see chapter 5 for a full review and discussion) however conclusions drawn from incomplete metagenomic data, without empirical observations, must be used with caution. Nevertheless, such studies have largely elucidated the principal functional genes involved in the anammox reaction, aiding future research.

1.2.3. Physiology and Structure

Anammox bacteria have been described as obligate anaerobes, and are active only in environments with O_2 concentrations less than 2 μM (Strous, *et al.*, 1997). However, anammox activity is only inhibited by higher O_2 concentrations, O_2 is not toxic to anammox bacteria (Strous, *et al.*, 1997) perhaps allowing anammox bacteria to survive in environments with transient exposure to oxygen (*e.g.* estuarine sediments). Anammox have also been shown to be active at a wide range of temperatures, from -1.8°C (Rysgaard and Glud, 2004) to 65°C (Jaeschke, *et al.*, 2009a) and 85°C (Byrne, *et al.*, 2008). Anammox bacteria have been reported to exhibit a high affinity for NO_2^- and NH_4^+, utilising these substrates even at concentrations < 5 μM (Jetten, *et al.*, 2009). However, the metabolic rate of anammox has been reported to be very low, producing 15 – 80 μmol N_2 per g cells (dry weight) min^{-1} (Jetten, *et al.*, 2009).

The anammox reaction occurs in a 1:1 ratio of NH_4^+:NO_2^- (van de Graaf, *et al.*, 1995). Anammox utilise the oxidation of NH_4^+ to synthesise adenosine triphosphate (ATP) as an energy source ($\Delta G^{\circ'}$ = -275 kJ mol^{-1} NH_4^+) which is then used to fix CO_2 into biomass and perform other metabolic functions (Güven, *et al.*, 2005; Jetten, *et al.*, 2009). The anammox reaction is in actuality a three-step reaction with hydrazine (N_2H_4) being an important intermediate (Kartal, *et al.*, 2011b). However, an overall equation for chemolithotrophic anammox can be expressed as:

$$NH_4^+ + 1.32NO_2^- + 0.066HCO_3^- + 0.13H^+ \rightarrow 0.26NO_3^- + 1.02N_2 + 0.066CH_2O_{0.5}N_{0.15} + 2.03H_2O$$

(Strous, *et al.*, 1998)

Anammox cells are coccoidal and are typically between 0.8 and 1.1 μm in diameter, with some *Ca.* Scalindua organisms exhibiting pili (van Niftrik and Jetten, 2012) and have been shown to form flocs or attach themselves to aggregate matter (Schmid, *et al.*, 2003; Hamersley, *et al.*, 2007; Jaeschke, *et al.*, 2007; Woebken, *et al.*, 2007; Quan, *et al.*, 2008). Anammox bacteria have also been shown to have unique characteristics in terms of their cell structure (van Niftrik and Jetten, 2012). Planctomycetes in general have interesting cellular properties and characteristics, such as large phylogenetic distances from other bacteria and a lack of peptidoglycan, adding to the suggestion that this phylum represents an evolutionarily historic group of organisms, perhaps representing one of the early evolutionary steps from prokaryotic to eukaryotic life (Fuerst, 2005). Most *Planctomycetes* have intracellular compartments which perform a variety of roles (Fuerst, 2005; Jetten, *et al.*, 2009; Fuerst and Sagulenko, 2011). Anammox bacteria are no different and contain a similar structure named the 'anammoxosome' (van Niftrik, *et al.*, 2010). It has been shown that the anammoxosome is the reaction centre for anammox bacteria (Figure 1.4), where N_2H_4 and NO are synthesised and oxidised (Neumann, *et al.*, 2011; van Niftrik and Jetten, 2012). The proteins associated with anammox metabolism, specifically hydrazine hydrolase/hydrazine synthase (HH/HZS) have been shown to be present only inside the anammoxosome (Karlsson, *et al.*, 2009) and a significant number of membrane-bound ATPases were found to be associated with the anammoxosome membrane (van Niftrik, *et al.*, 2010). The anammoxosome membrane is made from unique "ladderane" lipids, which have been hypothesised to protect the rest of the anammox cell and the nucleoid from potentially toxic compounds such as N_2H_4 and NO (Rattray, *et al.*, 2008; Boumann, *et al.*, 2009). As these ladderane lipids have only been reported in anammox bacteria, they have been postulated to be suitable as biomarkers for anammox in the environment (Jaeschke, *et al.*, 2009b).

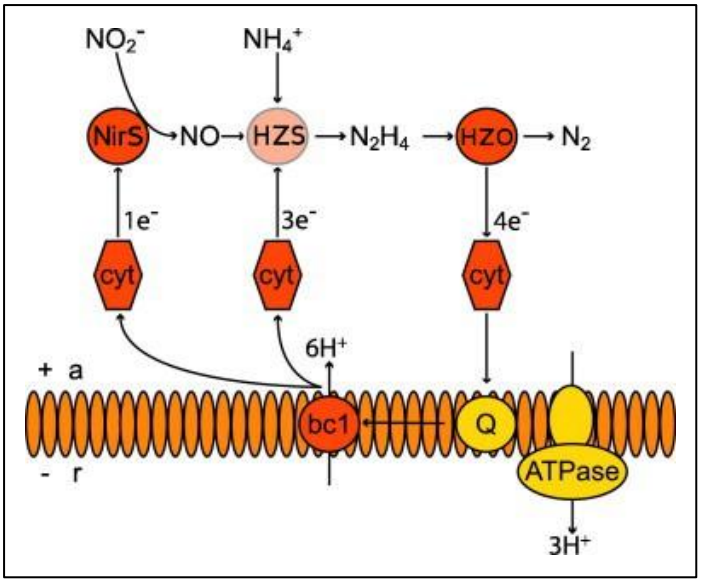


Figure 1.4: Diagram showing the current accepted view of the anammox reaction and the proteins involved coupled to the anammoxosome membrane. Figure adapted from **van Niftrik, *et al.* (2008)** and van Niftrik and Jetten (2012). Key: **bc1**, cytochrome bc_1 complex; **cyt**, cytochrome; **NirS**, nitrite reductase S; **HZS**, hydrazine synthase; **HZO**, hydrazine oxidireductase; **Q**, coenzyme Q; **a**, anammoxosome compartment; **r**, riboplasm compartment. This model clearly shows separation of the anammox reaction and toxic compounds from the rest of the anammox cell.

1.2.4. Microbial Ecology and Anammox

Ecology is the study of "the relationships of organisms to their environments" (Brock, 1966). Thus, the microbial ecologist wishes to study what an organism is, what it does, when it is active and how this organism affects both its environment and other organisms within its ecosystem (and *vice versa*). Classically, in order to study the ecology of a particular organism, it is necessary to isolate an organism from its environment in order to study its physiology and nature (Brock, 1966; McArthur, 2006). Laboratory experiments can be performed investigating factors which control the particular organism's activity using traditional microbiological, genomic, transcriptomic and proteomic approaches. In such experiments it is important to measure responses of the same organism under different conditions, to ensure that observations are comparable and that subsequent conclusions are valid. These findings can ultimately be taken out of the laboratory and tested *in situ* to further understand the organism in question as the effect of organisms on their environment can only be truly investigated *in-situ* (Brock, 1966; McArthur, 2006). Recently, the development of the "meta" age and associated technologies has

provided the ecologist with far more powerful tools (Huse, *et al.*, 2008; Kunin, *et al.*, 2010) to investigate his/her particular organism in its environments; though such technologies should be viewed as an addition to *in vitro* techniques and not a substitute.

Although enrichment cultures of anammox organisms do exist, to date, attempts to isolate these bacteria from the natural environment have been unsuccessful (Jetten, *et al.*, 2009). Thus, the studying the ecology of anammox organisms provides a unique challenge, as the classical methodology that one would hope to use (see above) is unavailable. This is further hindered by the fact that anammox bacteria are generally found to be of a low abundance, even in environments where they have been shown to be significantly active (Byrne, *et al.*, 2008; Hamersley, *et al.*, 2009). As such, anammox research has not progressed as rapidly as one might expect for such a unique and environmentally important group of organisms and much about these unique organisms is relatively unknown. Hence, the diversity and distribution of anammox organisms in natural environments is relatively unknown, as is the environmental factors affecting such diversity and the niche adaptation of these organisms (Hamersley, *et al.*, 2009; Humbert, *et al.*, 2010).

The acquisition of anammox genomes and metagenomes (section 1.2.2) has potentially revealed a far greater metabolic diversity of these organisms than has previously been suggested. However few investigations have investigated the potential for such metabolic diversity outside of the laboratory and what potential impact this may have for the ecology of anammox organisms. More detailed reviews of anammox metabolic and phylogenetic diversity and environmental distribution are included in the introduction sections of the subsequent chapters.

1.3. Aims, Hypotheses and Scientific Rationale

1.3.1. Overall Aims

This project aims to elucidate aspects of the ecology of anammox organisms:

a) To investigate anammox diversity and distribution and what drives this biogeography over environmental gradients.

b) To investigate whether evidence of anammox organisms utilising organic-nitrogen substrates can be found.

To achieve these aims a molecular approach was used implementing high throughput sequencing technologies as well as SIP.

1.3.2. Scientific Rationale and Hypotheses

1.3.2.1. Biogeography

1.3.2.1.1. Rationale

Anammox is an important and significant process in estuarine environments (Nicholls and Trimmer, 2009) and OMZs (Devol, 2003). To date, a detailed understanding of the biogeography of anammox organisms in such environments has yet to be achieved (Humbert, *et al.*, 2010) despite multiple studies investigating the diversity and distribution of these organisms (e.g. Rich, *et al.*, 2008; Dale, *et al.*, 2009; Galan, *et al.*, 2009). However, previous studies all have some issue relating to their methodologies (*e.g.* sampling strategy, sampling depth, specificity of PCR primers used *etc*) which prevented the findings of any particular study from being conclusive. Furthermore, no study exists within the literature which investigates both marine and estuarine anammox organisms in the same study. This creates an issue as no two investigations use exactly the same methodologies and therefore can never be directly comparable.

However, via comparison of all of these studies, it appears that "*Ca.* Scalindua spp." inhabit marine environments whereas the other four genera ("*Ca.* Brocadia spp.", "*Ca.* Jettenia spp.", "*Ca.* Kuenenia spp." and "*Ca.* Anammoxoglobus spp.") exist solely in freshwater or terrestrial environments (reviewed in Jetten, *et al.*, 2003; Jetten, *et al.*, 2009). From inference, one can assume that the biogeography of these organisms is dependent on salinity though whether this is as a direct result of changes in salinity or due to a secondary variable is unknown. Kartal, *et al.* (2006) however demonstrated that enrichments of terrestrial anammox bacteria were not greatly affected by increases in salinity. Furthermore, how the anammox community changes across a salinity gradient (*e.g.* along an estuary) is also unknown as no study

to date has comprehensively investigated both the freshwater and saline parts of an estuary in relation to the diversity of anammox bacteria.

Though OMZs appear to be dominated by *Ca.* Scalindua spp. and some hypotheses have been made concerning the micro-diversity of *Ca.* Scalindua in these ecosystems (Woebken, *et al.*, 2008; Galan, *et al.*, 2009), in truth few studies have investigated anammox diversity comprehensively in such environments due to small clone libraries and poor data resolutions. Though the core of the OMZ is relatively stable, in terms of its geochemistry, the periphery of the OMZ is typified by steep and potentially transient oxyclines and chemoclines (Paulmier and Ruiz-Pino, 2009). Such regions may present a key niche for anammox bacteria and a source of anammox diversity. Bae, *et al.* (2010) commented that anammox communities may be more diverse with increased O_2 concentrations (compared with complete anoxia) however, this observation was made from laboratory based anammox reactors and it is yet to be seen whether such an observation would hold true in the natural environment.

1.3.2.1.2. Specific Hypotheses

Anammox diversity was investigated across two environmental gradients; a salinity gradient (Medway Estuary, Kent, UK) and an oxygen gradient (ETNP OMZ). It was hypothesised that anammox diversity and community structure would change over these environmental gradients and that the observed anammox diversity would be dependent on varying environmental conditions.

1.3.2.1.2.1. Medway Estuary

Along the Medway Estuary, it was hypothesised that the anammox community would change from a *Ca.* Scalindua spp. dominated community in marine (high salinity) locations to a *Ca.* Brocadia spp., *Ca.* Jettenia spp. and *Ca.* Kuenenia spp. dominated community towards the freshwater end of the estuary (*i.e.* towards Allington Lock, see Chapter 2 for site descriptions). However as the specific niche adaptations of the different anammox genera is currently unknown, freshwater anammox communities may be dominated by a single, non-Scalindua genus or demonstrate a greater richness and diversity. It was hoped that the conclusions of this study would aid the understanding of the niche adaptations of these organisms.

It was further hypothesised that the transition between marine and freshwater anammox communities would be gradual, with decreasing salinity along the estuary, and would not demonstrate a sudden and distinct community shift. Although previous studies have suggested that *Ca.* Scalindua spp. are dominant in marine environments and non-Scalindua genera are adapted to less saline environments (reviewed in Song and Tobias, 2011), some studies have also reported the coexistence of several different anammox genera including suspected marine and freshwater species (Dale, *et al.*, 2009). As such it would seem that anammox bacteria (or at least some anammox bacteria) are reasonably tolerant to transient environmental conditions which would be advantageous to their survival in environments, such as estuaries which experience tidal surges and varying environmental parameters (*e.g.* salinity, temperature and concentrations of inorganic and organic compounds) throughout the year (Wharfe, 1977; Manahan, 2005; Dale, *et al.*, 2009). Therefore, it is unlikely that a distinct community shift between different anammox communities would be observed along the extent of the Medway Estuary but instead a gradual change in the proportion of different community members would be reported by the data.

Typically the range in salinity experienced towards the middle of the Medway Estuary is more variable throughout the year than at either extent (Figure 1.5) and so specific anammox community members, who may be more tolerant to such conditions, may outcompete their community rivals here. As such it was also hypothesised that, in addition to marine and freshwater dominated anammox communities, a separate anammox community may be observed which has adapted to meso-haline conditions and would be more dominant community members in such environments. However, as no such community has previously been reported in the literature, the nature of this anammox community, if it indeed exists, cannot be speculated on.

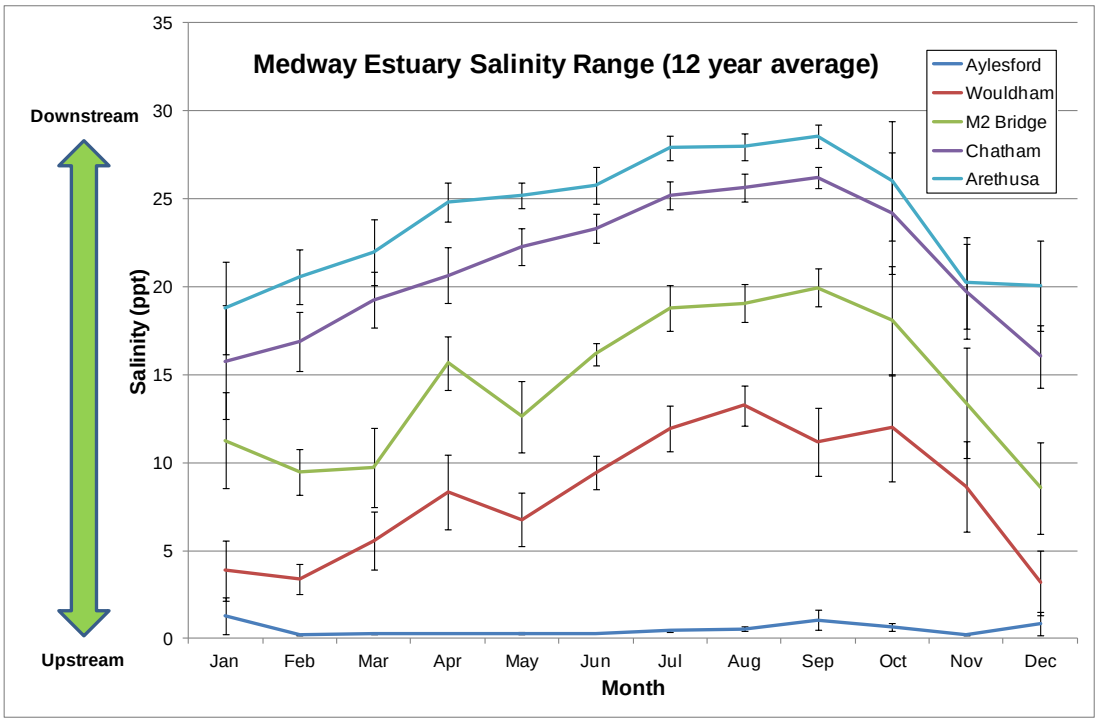


Figure 1.5: Mean annual salinity along the Medway Estuary. Mean annual salinity range across 5 sites along the Medway Estuary. Data obtained from The Environment Agency, UK. Error bars represent standard errors calculated from 12 year averages for each month. No data was available for the absolute marine extent of the estuary though salinities of ~ 30-35 ppt would be expected corresponding with global averages for seawater. The salinity at the most upstream site (Aylesford) was low throughout the year (mean value 0.54 ± 0.24 ppt) whereas salinities in downstream sites were more variable throughout the year (demonstrating an annual salinity range of approximately 10 ppt).

However, previous *in-vitro* research has suggested that anammox bacteria are highly-tolerant to varying salinities (Kartal, *et al.*, 2006). This appears to contradict *in-situ* anammox investigations which suggest that the diversity observed within the *Brocadiaceae* is determined by salinity (Song and Tobias, 2011). Therefore it was hypothesised that by comparing environmental data with measurements of anammox diversity along the extent of the Medway Estuary, other environmental parameters may be discovered which may have a controlling effect on anammox diversity (though this would have to be confirmed with *in-vitro* experimentation and can only be inferred from *in-situ* observations). As such, anammox diversity in estuarine environments may not be controlled by salinity, but rather by secondary variables which also increase or decrease along the estuary in conjunction with salinity. One such parameter may be the concentration of DOM which has been suggested to have a controlling influence on anammox bacteria (Trimmer, *et al.*, 2003; Rooks, *et al.*,

2012) and varies along estuaries with changing land use (*e.g.* from anthropogenic loadings).

1.3.2.1.2.2. ETNP OMZ

Previous studies concerning anammox diversity in OMZ settings have reported that these environments are dominated by *Ca.* Scalindua spp. but that the anammox community may also demonstrate a degree of micro-diversity within this genus (Woebken, *et al.*, 2008; Galan, *et al.*, 2009). As such it was hypothesised that this study's investigation into anammox community structure and diversity in the ETNP OMZ would also discover a *Ca.* Scalindua spp. dominated anammox population, despite the fact that this OMZ is the least well characterised, in terms of nitrogen cycling and anammox rates and diversity, of the planet's OMZ regions.

OMZ environments typically exhibit a stable anoxic (or suboxic) core with steep oxyclines and chemoclines above and below this core (Paulmier and Ruiz-Pino, 2009). The temporal variability of these periphery regions is unknown (largely due to the vast spatial extent of these OMZs and difficulties associated with regular and comprehensive sampling) though it is expected that they vary in depth with seasonal upwelling events or blooms in primary productivity (Paulmier and Ruiz-Pino, 2009). Furthermore, the steepness of the oxycline, certainly in the upper oxycline, where O_2 concentrations can rise as steeply as 2 μM/m (Paulmier and Ruiz-Pino, 2009), will produce highly variable environmental conditions with even minor fluctuations as to the size and depth of the oxycline. As oxygen is a major controlling factor on anammox activity (Jetten, *et al.*, 2009), the periphery of the OMZ is likely to produce a challenging environment for anammox bacteria compared with the stable core. However, recent research suggests that the definition of anammox bacteria as being strict anaerobes may (at least in some cases) be erroneous and some organisms may be able to operate in higher O_2 concentrations (Woebken, *et al.*, 2007). Thus, it was hypothesised that, if micro-diversity within the Scalindua genus exists in the ETNP OMZ, the upper and lower boundaries of the OMZ would potentially demonstrate the most significant shift in the anammox community. Organisms within the Scalindua genus may have adapted to become more tolerant to these regions of variable O_2 concentrations and hence would outcompete their less aerotolerant relatives, resulting in a shift in the anammox community between these

regions. It was further hypothesised that these distinct groups, if they existed, would likely be evenly distributed throughout the lateral extent of the OMZ and would vary with depth in conjunction with variations to the vertical extent of the OMZ (*e.g.* a shallower or deeper oxycline).

However, the variability associated with the OMZ's oxycline is also correlated with variation in the concentration of inorganic nutrients (Nicholls, *et al.*, 2007). As the availability of DIN is obviously a major limiting factor to anammox rates (Jetten, *et al.*, 2009), changes in the anammox community may also be observed representing increases and decreases in the concentrations of NO_2^- and NH_4^+ with the potential for different anammox organisms to demonstrate a higher or lower affinity for these substrates. However, as the dominant environmental variable in OMZ environments, it was hypothesised that O_2 concentrations and the extent of the upper and lower oxyclines would be the greatest controlling factor on anammox community structure in the ETNP OMZ.

1.3.2.2. Utilisation of Organic Nitrogen Substrates

1.3.2.2.1. Rationale

Anammox organisms are classically defined as strictly autotrophic bacteria (Van de Graaf, *et al.*, 1996; Güven, *et al.*, 2005) using solely inorganic substrates for metabolism. Despite this, anammox from laboratory bioreactors have been reported to be potentially able to oxidise propionate to CO_2 with either NO_2^- or NO_3^- as an electron acceptor (Güven, *et al.*, 2005) and a new, novel anammox organism, "*Ca.* Anammoxoglobus propionicus", has been suggested to be responsible for this process (Kartal, *et al.*, 2007b). Furthermore, anammox have also been reported to have close metabolic associations with nitrifiers (Dalsgaard, *et al.*, 2003) and methane oxidisers (Strous and Jetten, 2004; Luesken, *et al.*, 2011; Zhu, *et al.*, 2011). Clearly, the metabolism of anammox organisms and their ecological interactions and significance are not as simple or succinct as previously believed.

Prior to this project, serendipitous findings by Trimmer and Purdy (2012) showed that in the Arabian Sea OMZ, amine groups from allythiourea (ATU) were oxidised with NO_2^- to produce N_2 in a 1:1 ratio, similar to that of the anammox reaction. However, it was unknown whether this oxidation occurred directly (*i.e.* by anammox,

therefore making them mixotrophic) or via a syntrophic partnership (*e.g.* with organisms excreting inorganic compounds from the catabolism of organic substrates).

1.3.2.2.2. Specific Hypotheses

This project aims to attempt to discover more concrete evidence for such a process from a molecular standpoint (*i.e.* investigating *in-situ* microorganisms instead of just process measurements) and to look for potential syntrophic partners which may have a close spatial correlation with known anammox organisms. Investigation into the potential for organic pathways to the anammox reaction (henceforth named "organammox") would be conducted on samples collected from the Medway Estuary and ETNP OMZ.

It was hypothesised that if organammox existed as a process it would likely occur in one of two ways:

a) A syntrophic reaction with heterotrophic organisms which would metabolise organic compounds such as methyl amines or urea, potentially releasing amine groups as NH_4^+ which would then be utilised by anammox bacteria.

b) A heterotrophic anammox reaction in which anammox bacteria directly metabolise DON compounds, potentially incorporating organic carbon into biomass, in a unique and unprecedented reaction pathway in anammox bacteria.

If the first hypothesis is true, then anammox bacteria will be observed to react positively (in terms of activity and bacterial growth) to the introduction of organic substrates such as methyl amines and urea but would not be observed to incorporate organically derived carbon into biomass (though some organically derived carbon may be observed in anammox biomass via a feedback loop through CO_2 produced by anammox bacteria's potential syntrophic partners). Furthermore a close spatial correlation between anammox bacteria and syntrophic organisms would likely be observed, *e.g.* the formation of flocs as suggested by (Woebken, *et al.*, 2007; Quan, *et al.*, 2008).

Alternatively, if the second hypothesis is true, anammox communities amended with suitable organic compounds would potentially be seen to incorporate organic carbon into their biomass. As organammox bacteria would require high concentrations of DIN as a source of NO_2^- (and so it is likely that suitable inorganic NH_4^+ would also be available in such a system) it is unlikely that organammox would have evolved as a reaction to only cleave amine groups from organic molecules unless some other competitive advantage was also gained (*e.g.* the acquisition of carbon for anabolism). However, as the existence of such a process has yet to be confirmed, any hypotheses as to how it would occur are purely speculative and are based on little previous evidence.

In the Medway Estuary, it was hypothesised that, if organammox is present in this environment, it would be more prevalent upstream where organic loadings from agriculture and associated land uses are greater. However, DOM is present along the entirety of the estuary and as such the potential for organammox could be observed throughout the extent of the estuary.

Initial evidence for organammox was first observed in the Arabian Sea OMZ and so it was hypothesised that the ETNP OMZ would also contain anammox communities capable of this process. It was hypothesised that the potential for organammox would be greatest towards the top of the OMZ, in horizons of high NO_2^- concentrations, where the availability of this substrate would not be a limiting factor. However the potential for such a process to exist throughout the vertical extent of the OMZ could not be ignored. Furthermore, if evidence for organammox was found within the ETNP OMZ, then it was hypothesised that this process would be observed spatially throughout the OMZ, at least at the equivalent depth at which it was initially discovered.

1.3.3. Site Description and Rationale for Selection

1.3.3.1. Medway Estuary

The River Medway is a tributary of the Thames, flowing through the counties of West and East Sussex and Kent in south-eastern England. The river flows for 260 km from its source at Turners Hill, West Sussex to where it joins the Thames Estuary at Sheerness, Kent and is the largest river basin in Kent, covering a drainage

area of 1,800 km^2 (The Environment Agency, 2013). The lower Medway is navigable and the lock at Allington defines the tidal extent of the estuary (The Environment Agency, 2013). The upper river flows mainly through clay (Weald Clay), sandstone (Lower Greensand) and chalk (Upper and Lower Chalk) sedimentary formations whereas the tidal extent of the river flows through clay (Wadhurst Clay and London Clay) sediments (The Environment Agency, 2005; The Environment Agency, 2013). The prevalence of less permeable clay sediments makes the river liable to flooding in times of increased rainfall (The Environment Agency, 2005).

The main land use in the Medway river basin is agricultural though the lower extent of the estuary is heavily urbanised. The intensive use of fertilisers associated with increased agriculture has resulted in a high concentration of NO_3^- in the river leading to the Medway catchment and the underlying groundwater to be classified as a Nitrate Vulnerable Zone (NVZ) by The Environment Agency (The Environment Agency, 2005).

Tidal, Medway sediments are predominantly fine silts and clays with a high water-retention rate (42-66%) which gradually changes to more sandy sediments towards the mouth of the estuary (Wharfe, 1977). Medway Estuary sediments have also been reported to contain high levels of organic matter and exhibit a pronounced seasonal shift in their chemical properties due to seasonal algal blooms (Wharfe, 1977). The decomposition of such high concentrations of organics provides an ample source of NH_4^+ for the anammox reaction and bioturbation by the river's fauna (*e.g.* annelids) ensures that NO_2^- is also readily bioavailable (Rooks, *et al.*, 2012). Wharfe (1977) also described highly reduced redox potentials below a depth of 1 cm in tidal sediments, indicating anoxic (or substantially sub-oxic) conditions below this depth. However, redox potentials were higher (more positive) towards the downstream extent of the estuary suggesting higher concentrations of oxygen in these sediments (Wharfe, 1977).

Previous studies have shown that anammox organisms are both present and active in Medway Estuary sediments. Nicholls and Trimmer (2009) measured anammox rates along the estuary from Wouldham (salinity 5) to Stoke (close to where the estuary meets the Thames Estuary). Anammox rates were recorded ranging from 3.09% of

N_2 production to 10.93% of N_2 production (Nicholls and Trimmer, 2009). Rooks, *et al.* (2012) however recorded anammox rates as high as 32% of N_2 production at Medway Bridge Marina with rates decreasing to 6% of N_2 production at the marine extent of the Medway Estuary. No measurements were available for the most freshwater extent of the estuary. Anammox bacteria have also been shown to be present in the Medway Estuary, with FISH analysis indicating the presence of *Ca.* Scalindua organisms comprising up to 8% of the total prokaryotic community (Rooks, *et al.*, 2012). However, no study has investigated the presence and diversity of anammox bacteria in the estuary using sequencing techniques.

As such the Medway Estuary was chosen as a suitable location to investigate the hypotheses outlined in section 1.3.2.1.2.1, facilitated by the fact that the estuary was easily accessible. Please refer to section 2.1.1.1 for specific site locations investigated during this study and chapter 5 for a more detailed introduction in the biogeography of anammox organisms in estuarine environments.

1.3.3.2. ETNP OMZ

The ETNP OMZ is one of the four, major OMZs which exist permanently in the World's oceans (Podlaska, *et al.*, 2012). The ETNP OMZ typically extends from 0°-25° N and from approximately 80° W to 160° W, off the coast of Central America (Paulmier and Ruiz-Pino, 2009). OMZs are characterised by regions of very low oxygen concentrations, caused by intense upwelling of nutrients promoting blooms of primary producers at the surface, although there is currently no consensus as to what (*i.e.* a specific O_2 concentration) defines the extent of the OMZ, largely due to previous difficulties with accurately measuring low O_2 concentrations (Paulmier and Ruiz-Pino, 2009). Nevertheless, OMZs consistently demonstrate a steep oxycline (with an O_2 gradient of approximately -2 μM/m) from the surface, a stable OMZ core and then a more gradual oxycline below the OMZ core, returning to oceanic O_2 concentrations (Paulmier and Ruiz-Pino, 2009). These oxyclines produce unique redox and geochemical gradients which are important to our understanding of geochemical cycling (Stewart, 2011).

The ETNP OMZ is the largest of the permanent OMZ, covering an area of $12.4 \pm 1 \times 10^6$ km^2 (Paulmier and Ruiz-Pino, 2009). Despite the size of the ETNP OMZ, it has

not previously been characterised as thoroughly as other global OMZs (particularly the Arabian Sea and ETSP OMZs), however previous research suggests that the ETNP OMZ typically demonstrates a shallow upper oxycline (at approximately 40-100 meters BSL) with the core extending from 200-1000 m BSL (Yamagishi, *et al.*, 2007; Karstensen, *et al.*, 2008; Paulmier and Ruiz-Pino, 2009; Podlaska, *et al.*, 2012).

To date, no direct measurements of anammox rates and activity are available for the ETNP OMZ though anammox has been shown to be a significant contributor to N_2 loss from other OMZs, contributing up to 48% of the total N_2 loss (Kuypers, *et al.*, 2005; Thamdrup, *et al.*, 2006; Hamersley, *et al.*, 2007). Previous studies, using ladderane lipid (Rush, *et al.*, 2012) and 16S rRNA (Podlaska, *et al.*, 2012) analyses, have however shown that anammox bacteria are present in the ETNP OMZ. The presence of anammox bacteria in the ETNP OMZ and ubiquity of the anammox reaction in other OMZs suggests that anammox may also an important role in N cycling in the ETNP. This made the ETNP a suitable site for the investigation of the hypotheses outlined in section 1.3.2.1.2.2. However the ETNP OMZ was the only OMZ which was available for study during the timescale of this project.

1.3.4. Outline of the Project

This project was divided into three main components. Initially it was necessary to develop robust and effective methodologies for the investigation of anammox bacteria from the environment (see chapters 3 and 4). Secondly, the diversity and distribution and potential controls of these organisms were investigated (using the methods developed during the first part of the investigation) in the Medway Estuary and ETNP OMZ (see chapter 5). Thirdly, the potential for organammox was investigated in microcosm experiments from these two environments (see chapter 6).

2. Methodology

2.1. Sampling Strategy

2.1.1. Site Locations

The hypotheses presented in section 1.3 were investigated at two different locations; The Medway Estuary, Kent, UK and the ETNP OMZ (Figure 2.1). These two locations represented two different environmental gradients; a salinity gradient (Medway Estuary) and an oxygen gradient (ETNP OMZ).

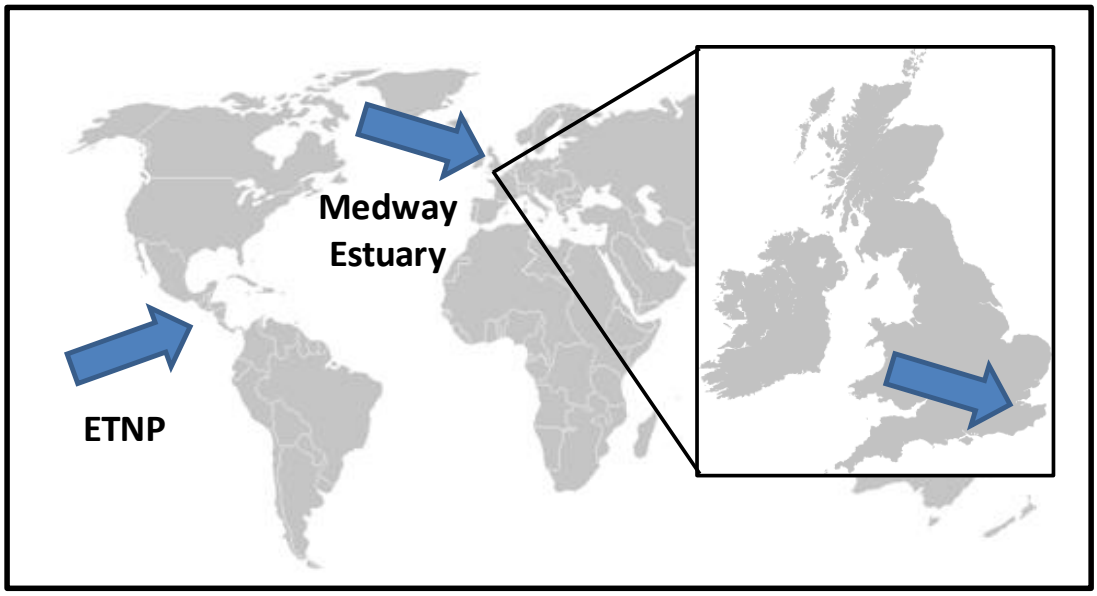
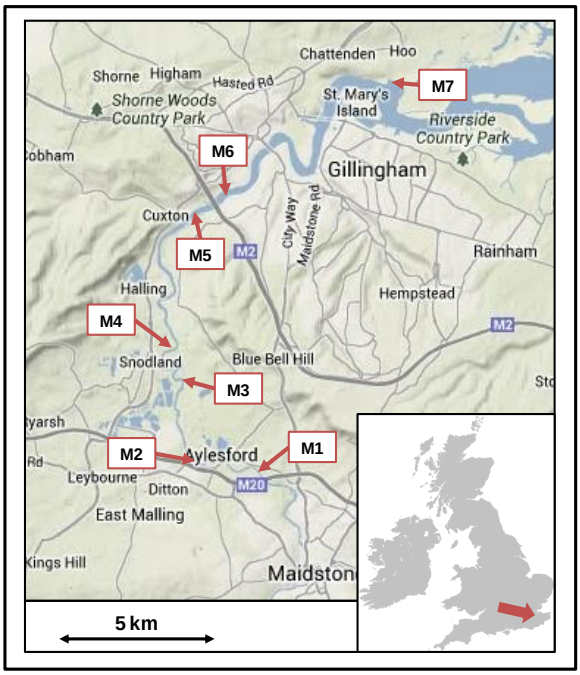


Figure 2.1: Map showing approximate locations of Medway Estuary and ETNP sampling locations.

2.1.1.1. Medway Estuary

Environmental samples were collected from the Medway Estuary at seven sites along the estuary (Figure 2.2). These sampling locations were chosen to represent the entire range of salinities experienced along the estuary, *i.e.* from freshwater to saline environments. Samples were collected at different dates during this project. Samples used for initial investigations into the suitability of different PCR primers in specifically targeting anammox organisms were collected from site M6 on 23/02/2010 (see section 3.2). Samples were collected from all 7 sites on 15/03/2012 in order to investigate anammox diversity along the extent of the estuary (see chapter 5). Samples for a pilot study into the potential for organammox were collected on 15/02/2011 (see section 2.4 chapter 6). Further experiments were also conducted on

Medway Estuary sediment collected on 25/07/2011 and 15/03/2012 to investigate the potential for this process (see section 2.4 and chapter 6). All sediment samples collected for organammox experiments were collected at site M6 apart from those which were collected on 15/03/2012 which were collected from sites M1 and M6 (see Table 2.6).

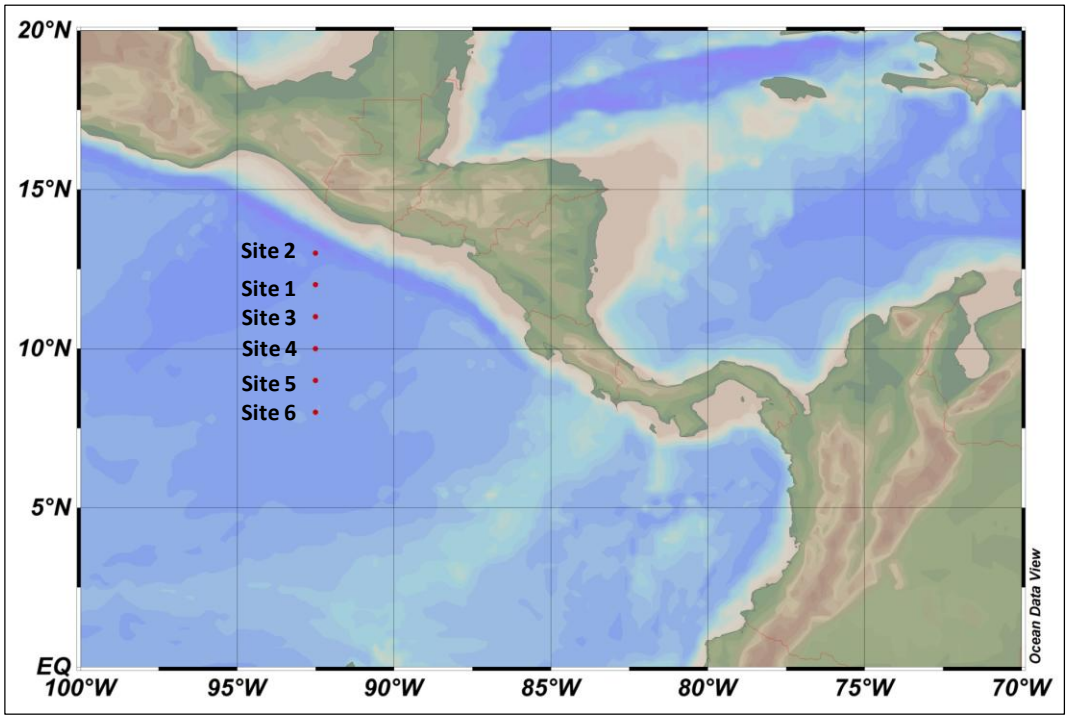


Site	Name		Latitude (° N)	Longitude (° E)
M1	Allington Lock		51° 17.83'	00° 29.87'
M2	Aylesford Priory		51° 18.18'	00° 28.05'
M3	Burham Marshes/Snodland		51° 19.29'	00° 27.64'
M4	Holborough Marshes/Works		51° 20.02'	00° 27.10'
M5	Cuxton Marina		51° 21.90'	00° 27.90'
M6	Medway Bridge Marina	1	51° 22.15'	00° 28.16'
		2	51° 22.07'	00° 28.16'
M7	Hoo Marina		51° 24.50'	00° 33.20'

Figure 2.2: Map showing approximate locations of sites along the Medway Estuary. Map was annotated from www.google.co.uk/maps. The attached table shows the exact locations and common names for each site.

2.1.1.2. ETNP OMZ

Samples were collected from the ETNP OMZ between 04/12/2011 and 28/01/2012 during NERC cruise D373. Samples were collected from 6 locations over a North-South longitudinal transect (Figure 2.3).



Site	Latitude (° N)	Longitude (° E)	Start Date	Depth to Bottom (m)
1	12.009	-92.503	11/12/2011	3882
2	12.998	-92.505	17/12/2011	4035
3	10.999	-92.501	23/12/2011	2671
4	9.997	-92.505	30/12/2011	3731
5	9.000	-92.500	05/01/2012	3611
6	8.000	-92.502	11/01/2012	3335

Figure 2.3: Map showing location of 6 sampling sites from ETNP cruise (D373). Map was drawn using Ocean Data View (ODV). Table shows the locations of sites sampled during ETNP Cruise. Start date is the date that sampling commenced at each site and depth to bottom is the depth of the water column at each site.

2.1.2. Sampling

2.1.2.1. Medway Estuary

All samples were collected in triplicate at the low-water mark. Samples were collected using 5 cm^3 corers, extracting the top 5 cm of sediment and transported back to the laboratory. The top 5 cm of sediment had previously been shown to demonstrate a high activity of anammox (Trimmer, *et al.*, 2003; Nicholls and Trimmer, 2009 and personal communication) with anoxic conditions experienced within the first ~1 cm. Samples were homogenised and porewater was extracted from these samples by centrifugation at 3000 rpm for approximately 15 minutes after

which the supernatant was filtered through a 0.2 μm filter and stored at -20°C. Sediment was transferred to 50 ml falcon tubes and stored at -20°C until extraction.

Samples which were to be used for SIP experiments were collected as outlined in section 2.4.2 (see below).

2.1.2.2. ETNP OMZ

Environmental samples were collected from 6 depths at each site (except site 2) in triplicate using a deck-controlled CTD with Niskin rosette attachment. Samples at site 2 were collected from 100 m, 200 m, 540 m and 750 m; site 4 were collected at 50 m, 60 m, 250 m, 436 m, 700 m and 860 m; site 5 were collected at 40 m, 67 m, 200 m, 450 m, 600 m and 719 m (see chapter 5). Triplicate samples for each depth were collected from three different Niskins which had been fired at the same time. 3 l of seawater from each Niskin was pre-filtered on collection using 200 μm filters and then filtered through Supor®-200 0.2 μm, ø 47 mm, polycarbonate filters (Pall, USA) with gentle vacuuming. Individual filters were placed in 2 ml cryotube vials and flash frozen in liquid nitrogen before being stored at -80°C.

SIP samples were collected as outlined in section 2.4.2 (see below).

2.2. Nucleic Acid Extraction

2.2.1. Extraction

DNA and RNA were extracted from environmental samples and SIP microcosms using the method outlined in Purdy (2005). This methodology was used as it has been shown to consistently provide high yields and purity of extracted nucleic acids (Purdy, *et al.*, 1996; Purdy, 2005) and allowed for the extraction of either total nucleic acids, only DNA, only RNA or separate DNA and RNA from one environmental sample. This methodology has also been found to be able to effectively extract nucleic acids from environments with low bacterial abundances (Zanardini, *et al.*, Unpublished; and personal communication with E. Monagahan, University of Warwick).

2.2.1.1. Medway Sediment

Nucleic acids were extracted from 0.5 - 0.7 g of homogenised estuarine sediment. Sediment was weighed in autoclaved 2 ml screw capped microcentrifuge tubes, with 0.5 g of DNA/RNA free glass beads, using a clean, sterile spatula and extracted as outlined by Purdy (2005).

2.2.1.2. ETNP Samples

For marine substrates, nucleic acids were extracted from either half of or a full polycarbonate filter. Due to low bacterial abundances from marine SIP incubations it was necessary to use the whole filter during extraction in order to obtain suitable quantities of DNA for downstream analyses. Filters were removed from storage and cut in half using a sterile scalpel and forceps. Cut filters were directly inserted into autoclaved 2 ml screw-capped microcentrifuge tubes with 0.5 g of DNA/RNA free glass beads and lysis buffer (see Purdy, 2005). DNA was then extracted as outlined in (Purdy, 2005). Using this method, the entire filter was seen to disintegrate/dissolve during lysis, thus negating the need to scrape, or wash cells from the filter as described in other extraction methods from aquatic samples.

2.2.2. Purification

It was occasionally necessary to perform a further purification step before sample analysis and downstream processing. This was particularly needed whilst processing sediment samples as it was discovered, even after extraction, that there were still large quantities of impurities (*e.g.* humic compounds) which hindered further processing (*e.g.* PCR).

DNA was purified using a Poly-ethylene-glycol (PEG) precipitation method adapted from Selenska and Klingmuller (1991) and Arbeli and Fuentes (2007). 0.2 volumes of a 5M NaCl solution and 1 volume of a 30% PEG_{6000} (Sigma-Aldrich) solution was added to DNA samples to be purified. Samples were incubated at 4°C overnight and then centrifuged for 30 minutes at 13,000 rpm. The resulting pellet was then washed twice with a 70% ethanol solution, again centrifuging for 30 minutes at 13,000 rpm between each wash. The supernatant was removed and the pellet left to

dry naturally for 15-30 minutes. The pellet was then re-suspended in 30-50 µl of a 10 mM Tris pH 7.5 solution.

2.2.3. Quantification

Extracted nucleic acids were quantified either by agarose gel electrophoresis with comparison against commercial ladders (Bioline, UK and Invitrogen, UK) or using a Nanodrop ND-1000 Spectrophotometer.

Typically 5 µl of aqueous, nucleic acid solution was mixed via gentle pipetting with 2 µl of loading dye and loaded onto 1% agorose gels (Bio-rad, UK). Agarose gels were run for between 30 minutes to 1 hour with a constant voltage of ~5 V cm^{-1} in 1x TAE buffer.

Aqueous nucleic acid samples were analysed via a Nanodrop ND-1000 Spectrophotometer as per the instructions which came with the instrument.

2.3. Environmental Analysis

2.3.1. Overview

The diversity and distribution of anammox organisms across environmental gradients was investigated using both a phylogenetic (16S rRNA) and functional (*hzo*) approach. Due to the difficulties associated with investigating anammox ecology (*e.g.* low abundance) it was necessary to develop a suite of PCR primers which would be able to effectively isolate and amplify target anammox sequences from the environment.

2.3.2. PCR

2.3.2.1. 16S rRNA

2.3.2.1.1. Primers

A range of PCR primers targeting anammox 16S rRNA genes were utilised during this project (Table 2.1). A full description of the primers used to detect anammox 16S rRNA genes during this study and their validation is given in chapter 3. PCR reaction protocols for each set of primers can be seen in Table 2.1.

	Primer Name	Primer Sequence	Specificity	Annealing Temperature (°C)	Approximate Amplicon Length (bp)	Reference
16S 1	Pla46F	5'-GGA TTA GGC ATG CAA GTC-3'	*Planctomycetales*	62	1344	(Neef et al., 1998)
	1390R	5'-GAC GGG CGG TGT GTA CAA-3'	Bacteria			(Zheng et al., 1996)
16S 2	Amx368F	5'-TTC GCA ATG CCC GAA AGG-3'	All Anammox	62	480	(Schmid et al., 2003)
	Amx820R	5'-AAA ACC CCT CTA CTT AGT GCC C-3'	Non-Scalindua Anammox			(Schmid et al., 2000)
16S 3	Amx368F	5'-TTC GCA ATG CCC GAA AGG-3'	All Anammox	62	480	(Schmid et al., 2003)
	BS820R	5' -TAA TTC CCT CTA CTT AGT GCC C-3'	*Ca. Scalindua spp.*			(Kuypers et al., 2003)
16S 4	Amx368F-GC	5'-CCG CCG CGC GGC GGG CGG GGC GGG GGC ACG GGG TTC GCA ATG CCC GAA AGG-3'	All Anammox	62	510	This Study
	Amx820R	5'-AAA ACC CCT CTA CTT AGT GCC C-3'	Non-Scalindua Anammox			(Schmid et al., 2000)
16S 5	An7F	5'-GGC ATG CAA GTC GAA CGA GG-3'	All Anammox	63	1380	(Penton et al., 2006)
	An1388R	5'- GCT TGA CGG GCG GTG TG-3'				
16S 6	An7F	5'-GGC ATG CAA GTC GAA CGA GG-3'	All Anammox	63	511	(Penton et al., 2006)
	518R	5'-ATT ACC GCG GCT GCT GG-3'				(Muyzer et al., 1993)
16S 7	Brod541F	5'- GAG CAC GTA GGT GGG TTT GT-3'	*Ca.* Scalindua spp.	60	719	(Penton et al., 2006)
	Brod1260R	5'-GGA TTC GCT TCA CCT CTC GG-3'				
16S 8	B540F	5'-GCT ACC GAA AGG GTT GCT AA	Non-Scalindua Anammox	55	669	This Study
	B1209R	5'-CCA TCG TTT ACG GCT AGG AC-3'				
16S 9	J697F	5'-AGG GTA AAG GCC TAC CAA GG-3'	*Ca.* Jettenia spp. and *Ca. Kuenenia spp.*	58	568	This Study
	J1265R	5'-CAA AAC CCC TCT ACC GAG TG-3'				
16S 10	K580F	5'-GCA AAA GCA CTT GTG GTC AA-3'	*Ca.*Kuenenia spp.	51	467	This Study
	K1047R	5'-CCC GTA CTC AAG CCC TGT AG-3'				

Table 2.1: Table showing list of 16S rRNA primers used in this study.

Primer Set	Hot Start	No. Cycles	Denaturation	Annealing	Elongation	Final Elongation
16S 1	95 °C for 2 min	a) 10	95 °C for 45 s	57 °C - 62 °C for 50 s*	72 °C for 1 min 22 s	72 °C for 5 min
		b) 20		62 °C for 50 s		
16S 2	95 °C for 2 min	30	95 °C for 45 s	62 °C for 50 s	72 °C for 1 min 22 s	72 °C for 5 min
16S 3	95 °C for 2 min	30	95 °C for 45 s	62 °C for 50 s	72 °C for 1 min 22 s	72 °C for 5 min
16S 4	95 °C for 2 min	30	95 °C for 45 s	62 °C for 50 s	72 °C for 1 min 22 s	72 °C for 5 min
16S 5	96 °C for 2 min	30	96 °C for 45 s	63 °C for 1 min	72 °C for 1 min	72 °C for 7 min
16S 6	96 °C for 2 min	a) 10	96 °C for 1 min	63 °C for 30 s	72 °C for 2 min 30 s	72 °C for 10 min
		b) 20	94 °C for 30 s		72 °C for 2 min	
16S 7	96 °C for 2 min	30	96 °C for 45 s	60 °C for 1 min	72 °C for 1 min	72 °C for 7 min
16S 8	96 °C for 2 min	a) 10	96 °C for 1 min	55 °C for 30 s	72 °C for 2 min 30 s	72 °C for 10 min
		b) 20	94 °C for 30 s		72 °C for 2 min	
16S 9	96 °C for 2 min	a) 10	96 °C for 1 min	58 °C for 30 s	72 °C for 2 min 30 s	72 °C for 10 min
		b) 20	94 °C for 30 s		72 °C for 2 min	
16S 10	96 °C for 2 min	a) 10	96 °C for 1 min	55 °C for 30 s	72 °C for 2 min 30 s	72 °C for 10 min
		b) 20	94 °C for 30 s		72 °C for 2 min	

Table 2.2: Table showing PCR reaction conditions for PCRs targeting anammox 16S rRNA genes. All PCRs were performed using a hot start at 96°C. Primer set numbers collaborate with those shown in Table **2.1**. The initial 10 cycles of PCR reaction 16S 1 utilised a touch-up PCR methodology, rising from 57°C to 62°C in increments of 0.5°C per cycle.

2.3.2.1.2. Promega Reagents

PCR was performed in either sterile 0.2 ml eppendorf tubes or 96 well PCR plates (Thermo-Scientific, UK). PCR was performed in 50 µl reactions using reagents and DNA Go-Taq polymerase (Promega, UK), dNTPs (Invitrogen, UK) and sterile, nuclease free water (Ambion, UK). The concentrations and volumes of PCR reagents are shown in Table 2.3. PCR was performed in an Eppendorf Mastercycler epgradient thermocycler using the specific conditions outlined in Table 2.2.

Reagent	Concentration	Volume	Final Concentration
PCR Reaction Buffer	5X	10 µl	1X
$MgCl_2$	25 mM	4 µl	2 mM
dNTPs	25 mM (each)	0.5 µl	0.25 mM (each)
Forward Primer	10 µM	2 µl	0.4 µM
Reverse Primer	10 µM	2 µl	0.4 µM
Bovine Serum Albumin (BSA)	100 µg/ml	2 µl	4 µg/ml
Taq DNA Polymerase	5 Units/µl	0.2 µl	1 Unit (0.02 Units/µl)
DNA	Variable	1 µl	-
ddH_2O	Pure	Up to 50 µl (28.3 µl)	-

Table 2.3: Table showing concentrations and volumes of PCR reagents for PCR carried out with Promega reagents. Volumes are for one reaction.

2.3.2.1.3. Bioline Reagents

PCR was also performed using reagents produced by Bioline, UK. One reaction contained: 12.5 µl Bioline My-Taq Red master mix, 1 µl of 10 µM of forward primer solution, 1 µl of 10 µM of reverse primer solution, 9.5 µl of sterile, nuclease free H_2O and 1µl of DNA solution to a final volume of 25 µl. PCRs were performed in an Eppendorf Mastercycler epgradient thermocycler using the specific conditions outlined in Table 2.2.

2.3.2.2. *hzo* Functional Genes

2.3.2.2.1. Primers

A range of primers targeting anammox *hzo* were used in this project (Table 2.4). A full discussion of these primers can be read in chapter 4.

Primer Set	Primer Name	Primer Sequence	Length (bp)	Reference
hzo 1	hzocl1F1	5'-TGY AAG ACY TGY CAY TGG-3'	470	(Schmid et al., 2008)
	hzocl1R2	5'-ACT CCA GAT RTG CTG ACC-3'		
hzo 2	hzocl1F1l	5'-TGY AAG ACY TGY CAY TGG G-3'	471	(Schmid et al., 2008)
	hzocl1R2	5'-ACT CCA GAT RTG CTG ACC-3'		
hzo 3	HZO4F	5'-TTG ART GTG CAT GGT CTA WTG AAA G-3'	1037	(Hirsch et al., 2011)
	HZO1R	5'-CTC TTC NGC AGG TGC ATG ATG-3'		

Table 2.4: Table showing PCR primer sequences used during the investigations into the use of hzo as a molecular marker of anammox.

2.3.2.2.2. Protocol

PCR was performed with primers targeting *hzo* as outlined in section 2.3.2.2 and Table 2.3 except that, for primers hzocl1F1 & hzocl1R2 and hzocl1F1L & hzocl1R2 (hzo 1 and hzo 2 in tables), 2.5 µl of 25 mM $MgCl_2$ (*c.f.* Table 2.3) was added per reaction to a final concentration of 1.25 mM as performed by Schmid, *et al.* (2008). PCR was performed in an Eppendorf Mastercycler epgradient thermocycler using the conditions outlined in Table 2.5.

Primer Set	Initial Denaturation	No. Cycles	Denaturation	Annealing	Elongation	Final Elongation
hzo 1	94 °C for 5 min	30	94 °C for 1 min	50 °C for 1 min	72 °C for 1 min 30 s	72 °C for 10 min
hzo 2	94 °C for 5 min	30	94 °C for 1 min	53 °C for 1 min	72 °C for 1 min 30 s	72 °C for 10 min
hzo 3	94 °C for 5 min	a) 35	94 °C for 1 min	53 °C for 1 min	72 °C for 2 min	72 °C for 10 min
		b) 1	94 °C for 1 min	53 °C for 1 min	N/A	

Table 2.5: Table showing PCR protocols used with primers targeting anammox hzo. All PCRs were initiated with a hot start at 94 °C. Primer set numbers collaborate with those in Table **2.4**.

2.3.2.3. Analysis

The presence or absence of amplifiable product was determined via electrophoresis on 1% agarose gels as described in section 2.2.3. PCR product was quantified either by visual comparison of agarose gels or using a Nanodrop ND-1000 Spectrophotometer (section 2.2.3).

2.3.2.4. PCR Purification

PCR products required for downstream processing (*e.g.* sequencing) were purified to remove impurities which may remain after PCR (*e.g.* unbound dNTPs). PCR products were purified either using a PCR Purification Kit (Qiagen, UK) or were extracted from 1% agarose gels using a Gel Extraction Kit (Qiagen, UK).

2.3.3. Cloning of Environmental Sequences

Small clone libraries of anammox DNA amplicon sequences were prepared in order to determine the specificity of PCR primers and to ensure that anammox gene

sequences could be successfully extracted from the environment prior to more comprehensive analyses using 454 pyrosequencing.

PCR products were purified as outlined in section 2.2.2. Purified PCR products were cloned into pGEM-T Easy (Promega, UK) vectors and transformed into JM109 competent *E. coli* cells (Promega, UK) as described in the pGEM-T Easy cloning manual (Promega, UK).

Transformed cells were grown overnight on LAXI (LB agar containing 100 μg ml^{-1} ampicillin, 2.5 μM IPTG and 80 μg ml^{-1} X-Gal) plates at 30°C (Sambrook, *et al.*, 1989). White colonies were isolated and screened for the correct sized insert using M13 vector-based primers. M13 PCR products were submitted for sequencing at either The Natural History Museum, London, UK or GATC, Germany via Sanger sequencing technologies. Typically between 5 and 10 clones were submitted per sample for sequencing. Amplicons were sequenced using both the forward and reverse primers. Forward and reverse motifs were manually checked for errors and trimmed in SeqManII (DNAStar, USA) and assembled into contigs. Motifs which could not form contigs (*e.g.* due to poor sequencing results) were discarded.

2.3.4. Phylogenetic Analysis

Environmental sequences were analysed in MEGA 5.1 (Tamura, *et al.*, 2011). Sequences were aligned to sequences from reference databases (e.g. NCBI, SILVA) using MUSCLE (Edgar, 2004) and phylogenetic differences inferred using a Neighbour-Joining methodology calculated using the p-distance model. 1000 bootstrap replications were calculated as a test of phylogeny.

2.3.5. 454 Pyrosequencing

2.3.5.1. Sample Submission and Sequencing

Selected DNA samples from the Medway Estuary and ETNP were submitted for 454 pyrosequencing analysis to investigate the diversity of anammox organisms in these environments. Samples were amplified using primers Amx368F & Amx820R to target anammox 16S rRNA gene sequences (see chapter 3 for full discussion of

primers) as outlined in section 2.3.2. PCR products were purified as outlined in section 2.3.2.4.

Purified PCR products were submitted to either The Research and Testing Laboratory, Lubbock, Texas, USA or DHVLA, Surrey, UK for standard amplicon pyrosequencing with tagged primers suitable for titanium chemistries (Dowd, *et al.*, 2008; Oakley, *et al.*, 2012).

2.3.5.2. Processing Raw Data

Raw 454 pyrosequencing data were analysed using a custom bioinformatics pipeline, originally designed by Oakley, *et al.* (2012), using the PERL scripting language, to quality control and analyse amplicon 454 pyrosequencing reads. These scripts were further developed during this project to suit the particular requirements of this study.

2.3.5.2.1. Quality Control and Data Clean-up

A number of assumptions (Huse, *et al.*, 2007; Kunin, *et al.*, 2010), as to what qualifies as erroneous or poor quality read, were used to remove such sequences from 454 pyrosequencing data; namely the absence of the primer sequences in the read sequence, the presence of ambiguous bases (Ns) and the length of sequenced reads. The application of these assumptions to the quality control of 454 pyrosequencing data included the use of a novel, innovative technique, Read Length Incremental Clustering (RLIC), developed as part of this thesis, for removing arbitrary decision making from read length trimming. A full discussion of the development of RLIC and the quality control of 454 pyrosequencing data is presented in section 5.3.

2.3.5.2.2. Defining OTUs and Clustering

OTUs were created from sequencing data which had passed the above quality control using CD-HIT-EST (Li, *et al.*, 2001; Niu, *et al.*, 2010; Fu, *et al.*, 2012). OTUs were defined at a similarity cut-off of 95%. Low abundance OTUs (*i.e.* those containing small numbers of reads), in general, were omitted from analysis of the diversity of anammox organisms as these were deemed to have a high potential for containing poor quality reads. However, these low abundance clusters were included in the

analysis of ETNP 454 pyrosequencing data in order to investigate the validity of this assumption (see section 5.4.3.3).

2.3.6. Measuring Diversity and Distribution

Rarefaction curves and OTU distributions and frequencies were calculated using MOTHUR (Schloss, *et al.*, 2009) and R. The output from CD-HIT provided reference sequences for each OTU. These sequences were aligned with other anammox sequences and phylogenetic distances calculated as outlined in section 2.3.4. These phylogenetic relationships were used to assign taxonomies to specific OTUs.

CCA plots for OTU abundances across samples were calculated in order to test for the degree of similarity between these samples. CCA statistics were calculated in R using the VEGAN package (Dixon, 2003). This was correlated with geochemical data collected during sampling (for ETNP data) or provided by The Environment Agency, UK (for Medway data). Exact data for each sampling point along the Medway estuary was not available and so data was interpolated and extrapolated linearly using MATLAB v8.1.0.604 (R2013a), The MathWorks Inc. (2013). Mantel tests were conducted using the VEGAN package (Dixon, 2003) and plotted with CCA plots in R.

2.4. SIP

2.4.1. Overview

Stable Isotope Probing (SIP) was used to investigate the potential for anammox organisms to utilise organic substrates and demonstrate a greater metabolic diversity than previously reported.

Experiment	Location	Site	Date	Incubations	Concentration	Time
Pilot	Medway	M6	15/02/2011	13C DMA	10 mM	4 weeks
				13C Urea	10 mM	
				No Addition	N/A	
Exp 1	Medway	M6	25/07/2011	12C DMA	10 mM	4 weeks
				13C DMA	10 mM	
				12C Urea	10 mM	
				13C Urea	10 mM	
				12C DMA	20 mM	
				13C DMA	20 mM	
				12C Urea	20 mM	
				13C Urea	20 mM	
				12C TMA	10 mM	
				13C TMA	10 mM	
				No Addition	N/A	
Exp 2	Medway	M1	15/03/2012	12C DMA	10 mM	4 weeks
				13C DMA	10 mM	
				12C Urea	10 mM	
				13C Urea	10 mM	
				No Addition	N/A	
		M6		12C DMA	10 mM	
				13C DMA	10 mM	
				12C Urea	10 mM	
				13C Urea	10 mM	
				No Addition	N/A	
Exp 1	ETNP	S4: 436 m + 700 m	03/01/2012	12C TMA	100 µM	3 weeks
				13C TMA	100 µM	
				12C Urea	100 µM	
				13C Urea	100 µM	
				12C TMA + NO2-	100 µM + 25 µM	
				13C TMA + NO2-	100 µM + 25 µM	
				12C Urea + NO2-	100 µM + 25 µM	
				13C Urea + NO2-	100 µM + 25 µM	
				No Addition	N/A	
		S5: 450 m + 600 m	09/01/2012	12C TMA	100 µM	2 weeks
				13C TMA	100 µM	
				12C Urea	100 µM	
				13C Urea	100 µM	
				12C TMA + NO2-	100 µM + 25 µM	
				13C TMA + NO2-	100 µM + 25 µM	
				12C Urea + NO2-	100 µM + 25 µM	
				13C Urea + NO2-	100 µM + 25 µM	
				No Addition	N/A	

Table 2.6: Table showing exact locations, dates and substrate concentrations of each SIP experiment carried out during this study. All incubations were set up in triplicate. Concentrations are the final concentration in each microcosm. For ETNP SIP experiments, identical experiments were set up at two depths at each site as indicated in the table. For more details of how SIP microcosms were set up please refer to the main text.

2.4.2. Microcosm Set-up

2.4.2.1. Medway Estuary

Microcosms were set up with sediment collected from sites 1 (Allington Lock) and M6 (Medway Bridge Marina) from the Medway Estuary. 5 cm^3 cores were collected at low water mark in triplicate. Cores were transported back to the laboratory on ice and transferred to clean 50 ml serum bottles. 5 ml of filtered (0.45 μm) porewater (collected from sediment collected at the same time as sediment cores) was added to each microcosm with an amount of labelled organic substrate (see section 2.4.3 below). Serum bottles were sealed with butyl stoppers and crimped. The headspace gas in each sealed serum bottle was flushed with oxygen free nitrogen (BOC, UK) for approximately 20 minutes, using a surgical needle, in order to ensure microcosms were anaerobic. A second needle was also inserted into the serum bottle to allow expelled gas to escape; this was removed prior to removing the inflow in order to prevent backflow of oxygenated air and in order to create a slight positive pressure. Microcosms were incubated in the dark for approximately four weeks after which microcosms were stored at -20°C prior to DNA extraction.

2.4.2.2. ETNP OMZ

Water was captured at specific depths (see Table 2.6) using a deck –controlled CTD with a Niskin rosette attachment. On deck, water was extracted from Niskins in triplicate directly into 1 L serum bottles using PVC tubing. Tubing was inserted into the bottom of the serum bottle and allowed to fill up 3 times to maintain anaerobic conditions. The tubing was removed allowing for the serum bottle to be overfilled and a butyl stopper carefully inserted so as to prevent oxygenated air entering the serum bottle. Serum bottles were crimped with aluminium seals. A 4 ml headspace of He was introduced into each serum bottle. Microcosms were amended with organic substrates as outlined in section 2.4.3. Microcosms were incubated at 4°C for approximately 2 weeks at which point they were filtered through 0.2 μm poly-carbonate filter (Whatman, USA) using a vacuum pump and flash frozen in liquid nitrogen and stored at -80°C prior to DNA extractions.

2.4.3. Substrates

Microcosms were amended with ^{12}C Urea, ^{13}C Urea, ^{12}C DMA, ^{13}C DMA, ^{12}C TMA or ^{13}C TMA (Sigma-Aldrich, UK). ^{13}C labelled substrates contained >99.9% labelled carbon. Varying concentrations of these substrates were added to different experiments. For a full list of SIP experiments and microcosms set up see Table 2.6.

2.4.4. DNA Extraction

DNA was extracted from SIP microcosms as outlined in section 2.2. The protocol was followed as described in Purdy (2005) except that the 120 mM Na_2HPO_4 wash step was replaced with 150 mM K_2HPO_4 in order to remove RNAs. This was done as a precaution to avoid the possibility that RNAs may act as a carrier to DNA during fractionation.

2.4.5. SIP

2.4.5.1. Centrifugation

SIP was performed on extracted DNA as outlined in (Neufeld, *et al.*, 2007). 2-5 µg of DNA from each sample was mixed with a 7.163 M CsCl solution and gradient buffer (0.1 M Tris, 0.1 M KCl and 1 mM EDTA) to a final density of 1.725 g ml^{-1} and volume of ~5.5 ml. Densities were measured using a AR200 digital refractometer (Reichert, USA). CsCl mixtures were added to ultracentrifuge tubes and centrifuged in a Beckman Coultier Optima L-90K Ultracentrifuge for 48 hours at 44,100 rpm. Centrifuged samples were fractionated immediately.

2.4.5.2. Fractionation

Centrifuged DNA samples were fractionated into 12 equal fractions as described in Neufeld, *et al.* (2007). The density of each fraction was measured using a AR200 digital refractometer (Reichert, USA) and gradients for each sample were calculated in order to ensure that density gradients had formed correctly during centrifugation and had remained intact during fractionation. DNA was obtained from each fraction via precipitation with PEG_{6000} (Selenska and Klingmuller, 1991) as outlined in section 2.2.2 except that 1 µl of 20 mg/ml glycogen (Roche UK) was also added as

this had been reported to aid the precipitation of low quantities of DNA (Neufeld, *et al.*, 2007).

2.4.6. PCR

PCR was performed on fractionated DNA as outlined in section 2.3.2. 1-2 µl of aqueous DNA (depending on the quantity of DNA) was amplified using primers Amx368F & Amx820R (Table 2.1).

hzo genes were also amplified from fractionated DNA using a nested PCR approach. DNA was initially amplified using primers HZO4F & HZO1R. 1 µl of product from this first reaction was then further amplified using primers hzocl1F1l & hzocl1R2. PCR product was not purified between the first and second round amplifications as, due to the low quantities of DNA, even after the first round of amplification, purification was deemed detrimental to further amplification. Furthermore, purification did not seem necessary to produce clean, PCR product of the correct size, from the second round amplification (Figure 4.2).

2.4.7. DGGE

DGGE was used to investigate the potential enrichment of ^{12}C and ^{13}C DNA fractions and to investigate community change of anammox bacteria across these fractions. Fractionated DNA was amplified for anammox 16S rRNA genes with primers Amx368F-GC & Amx820R (Table 2.1). DGGE was performed as described in (Schafer and Muyzer, 2001). Denaturing gradients were formed between 20% and 80% urea and 18 µl of PCR product loaded to each gel. Denaturing gels were run for 16 hours at a constant voltage of 100 V. Gels were stained with ethidium bromide for 30 minutes with constant shaking and then washed for 30 minutes in pure water. Gels were observed using a UV-illuminator.

2.4.8. Clone Libraries

Clone libraries were constructed (Sambrook, *et al.*, 1989) from SIP DNA fractions using both 16S rRNA bands obtained from DGGE and PCR product obtained from *hzo* specific PCR primers as outlined in section 2.3.3.

Notable bands observed during DGGE analysis were excised from polyacrylamide gels using a sterile scalpel and a UV trans-illuminator. Gel slices were transferred to autoclaved microcentrifuge tubes and washed with 200 µl of nucleic acid free water and incubated for 2 hours at room temperature. Samples were centrifuged and the supernatant removed. Slices were then re-washed with 50 µl of the same water, centrifuged briefly and incubated overnight at 4°C. 1 µl of the resulting supernatant was then re-amplified via PCR with primers Amx368F & Amx820R.

PCR product was purified using a commercial PCR Purification Kit (Qiagen, UK) and then cloned into the pGEMT-Easy vector (Promega, UK) as per kit instructions Cloned plasmids were transformed into JM109 competent *E. coli* cells (Promega, UK) and single colonies isolated. Colonies were screened for containing the insert and then sequenced via Sanger sequencing (Natural History Museum, UK or GATC, Germany) using both the M13F and M13R primers. Forward and reverse reads were checked and trimmed manually in SeqMan II (DNASTAR, USA) and assembled into contigs. Contigs were aligned against reference sequences and phylogenetic relationships calculated in MEGA4 and MEGA5 (Tamura, *et al.*, 2007; Tamura, *et al.*, 2011).

2.4.9. 454 Pyrosequencing

Individual fractions were submitted for pyrosequencing of anammox 16S rRNA genes and analysed as outlined in section 2.3.5.

3. Development of Robust and Effective Methods for Investigating Anammox Organisms and Ecology *In Situ*

3.1. Introduction

Microbial ecology, the "study of the interactions of (micro)organisms and their environments" (Brock, 1966), is a challenging subject. The microbial ecologist is presented with a number of unique challenges, namely:

a) The issue of *in vitro* versus *in vivo* studies; whereby observations made *in vitro* are not necessarily comparable with those from the environment.

b) That the potential power of *in vitro* experiments is far greater than *in situ* studies (*e.g.* the ability to test an organism's response to individual conditions, the possibility of whole genome and transcriptome sequencing and the ability to isolate and purify specific proteins from a known organism and test their function and activity).

c) The problem of scale.

d) That many ecological methods and theories have been devised from observations of macro-fauna and flora and thus are not necessarily transferrable to the microbial realm (*e.g.* the concept of species).

Nevertheless, as our knowledge of our planet and global environment increases and the role of microorganisms within geochemical cycles are found to be increasingly important, microbial ecology presents itself as a crucial discipline within the life and environmental sciences. Thus it is imperative that effective methods for investigating microbial ecology be devised and developed. However, with robust scientific methodologies including the definition of stringent and testable hypotheses, many crucial advances in this field can (and have) been made using a combination of *in vitro* and *in situ* measurements.

Despite this, the challenges presented to the microbial ecologist wishing to investigate anammox organisms (or indeed the anaerobic nitrogen cycle) are even greater. To date, anammox bacteria have resisted attempts to isolate them in pure culture and as such all described anammox organisms are only "*Candidatus*" species'

(Jetten, *et al.*, 2009; Li, *et al.*, 2010b). The doubling time of anammox bacteria has been reported to be between 11 and 20 days under optimum conditions (Jetten, *et al.*, 2009) and this slow growth rate appears to be the major factor limiting isolation (Hirsch, *et al.*, 2011). Hence, most investigations into anammox ecology have been restricted to culture-independent methods (Li, *et al.*, 2010b) which has culminated in a lack of physiological and molecular information about these organisms (Amano, *et al.*, 2007).

Some studies have been successful in growing enrichment cultures of these bacteria which typically comprise of between 70-99% (Gori, *et al.*, 2011; van de Vossenberg, *et al.*, 2012) anammox cells. From such studies it has been possible to sequence the genomes for two anammox bacteria; *Ca.* Kuenenia stuttgartiensis (Strous, *et al.*, 2006) and *Ca.* Scalindua profunda (van de Vossenberg, *et al.*, 2012) although the coverage of *Ca.* K. stuttgartiensis was not complete and the 5 contigs did not overlap. A metagenomic approach was also taken in reporting key genes from *Ca.* Brocadia fulgida from a 70% enrichment culture (Gori, *et al.*, 2011). Enrichment cultures have also allowed for investigations into key functional proteins of the anammox reaction such as hydrazine oxidoreductase (HZO, Shimamura, *et al.*, 2007) and nitrite reductases (NirS and NirK, Li, *et al.*, 2011; Hira, *et al.*, 2012). However, despite the usefulness of such studies to our knowledge of anammox, the findings from such research (*e.g. de novo* genome assembly) based on enrichment cultures must always be used with caution as the possibility of including genes and proteins from non-target or anammox-related organisms is high. Furthermore the acquisition of anammox enrichment cultures is far from trivial with numerous different reactors being used to create cultures in batch which typically take up to two years to produce a significant proportion of anammox cells (Jetten, *et al.*, 2009; van de Vossenberg, *et al.*, 2012).

Anammox activity rates, in terms of N_2 produced, can be measured, relative to denitrification rates, *in situ* using an adaption to the isotope pairing technique (IPT) (Nielsen, 1992). The addition of incubations with $^{15}NH_4^+$ (with and without NO_2^-) and $^{15}NO_3^-$ (with and without NH_4^+) and the subsequent measurement of $^{28}N_2$,$^{29}N_2$ and $^{30}N_2$ allows for the relative rates of denitrification and anammox to be calculated (Thamdrup and Dalsgaard, 2002). However, even this technique may not give an

accurate representation of anammox rates as it does not (and cannot) account for the role of dissimilatory nitrate reduction to ammonia (DNRA) which could potentially produce $^{30}N_2$ from the anammox reaction after incubation with only $^{15}NO_3^-$ (Kartal, *et al.*, 2007a). There is debate within the literature of whether DNRA is a significant contributor to the N-cycle (Dong, *et al.*, 2009; Dong, *et al.*, 2011). Furthermore, it has been suggested that anammox may have the ability to carry out a DNRA-like reaction themselves and can convert NO_3^- to NO_2^- and eventually NH_4^+ (Francis, *et al.*, 2007; Kartal, *et al.*, 2007a). However, Dalsgaard, *et al.* (2012), although they were unable to measure the effect of DNRA directly, were able to model the potential effect of this process on N_2 measurements and found this "false denitrification" signal to be negligible. Nevertheless, these discrepancies are likely to underestimate the role of anammox to global N_2 production.

Anammox organisms are typically low abundance in the environment despite apparently being of great significance in terms of global N_2 production. Cell counts in the literature range from $1.9 \pm 0.8 \times 10^4$ cells ml^{-1} in the Black Sea (Kuypers, *et al.*, 2003) to $10.1 \pm 1.5 \times 10^4$ cells ml^{-1} in the Peruvian OMZ (Hamersley, *et al.*, 2007) though it must be noted that these were determined using different methodologies and so may not be directly comparable. In pelagic environments, anammox bacteria typically represent much less than 5% of the total bacterial population and this is likely to be less in more diverse environments (*e.g.* sediments). It is impossible to accurately determine microbial numbers using FISH from sediments as fluorescent stains also bind to organic compounds which cannot be completely removed from a sample (Song and Tobias, 2011). Furthermore, bacterial cells which have adhered to such particles or have formed flocs cannot be accurately enumerated. This may also be an issue when performing cell counts on aquatic samples. FISH probes of unknown specificity and efficiency of hybridisation may also lead to inaccurate cell counts, exacerbating this problem.

The difficulties associated with the use of culture-dependent methods limits the type of investigations and experiments which can be conducted with these organisms. Therefore, in order to achieve the aims of this project (section 1.3) it was necessary to develop robust and efficient, culture-independent methods to investigate anammox ecology. Thus, a predominantly molecular approach was used to test the hypotheses

of this project (see section 1.3.2.1.2). PCR using primers specific for the anammox clade, amplifying the target genes above background levels, would allow for the qualitative and relative quantitative analysis of these organisms *in situ* and the determination of their diversity.

In recent years the power of molecular biological techniques has increased enormously. The development of "omics" approaches has increased our understanding of the systems which we, as scientists, wish to elucidate. With the onset of high-throughput sequencing technologies our ability to process and analyse such data has also developed making these techniques valuable tools if used in conjunction with good experimental procedure. However, due to low anammox cell densities in the environment, the use of general bacterial primers (or even those targeting a smaller, but still relatively broad clade *e.g.* targeting *Planctomycetes*) are likely to underestimate anammox diversity by returning only a small number of anammox-related sequences (Amano, *et al.*, 2007). It was therefore necessary to develop a specific and comprehensive suite of primers to target anammox genes which could be used reliably and with confidence to amplify anammox sequences from environmental samples via PCR.

To this end a range of PCR primers exist within the literature to amplify anammox 16S rRNA genes (reviewed in Li, *et al.*, 2010b; Song and Tobias, 2011). However such primers have been used with varying amounts of success (Penton, *et al.*, 2006; Li, *et al.*, 2010b) and have been shown to target 16S rRNA gene sequences from outside of the anammox group (Song and Tobias, 2011). If such PCR primers are non-specific and amplify non-target organisms, this may pose a problem for anammox ecology as the amplification of these non-target sequences may be far greater than that of target, anammox sequences, resulting in a dilution of the already weakened anammox signal.

3.2. Methodology

3.2.1. General Anammox 16S rRNA Primers

3.2.1.1. Overview

16S rRNA PCR primers, reported to be anammox specific (see table), were investigated as to their suitability for investigating anammox ecology. Primarily two candidate PCR assays were investigated: An7F & An1388R and Amx368F & Amx820R/BS820R.

3.2.1.2. Environmental Samples and PCR Controls

Primers were initially tested for specificity using a number of controls including environmental samples (Medway Estuary site M6, collected 23/03/2010), three environmental clones positively identified as *Ca.* Jettenia sp. *Ca.* Brocadia sp. and *Ca.* Scalindua sp. and the entire synthesised 16S rRNA gene of *Ca.* Kuenenia stuttgartiensis (GenScript, USA) and genomic DNA extracted from a pure culture of *Planctomyces maris* DSM-8797 (DSMZ, Germany) along with other DNA samples from the laboratory including archaeal DNA and DNA extracted from isolates of non-anammox bacteria. Site M6 was chosen as it had previously demonstrated high anammox rates and the presence of anammox bacteria (Rooks, *et al.*, 2012). DNA was extracted from environmental samples and *P. maris* cultures using the method outlined in Purdy (2005). PCRs were performed as outlined in section 2.3.2.1.

3.2.1.3. 454 Pyrosequencing

Following initial investigations as to the specificity of these primers in the laboratory, their suitability for measuring anammox ecology using high-throughput sequencing technologies (*i.e.* 454 pyrosequencing) was tested. PCR product was obtained in triplicate from environmental samples. Triplicate PCR products were pooled, purified (PCR Purification Kit, Qiagen, UK) and submitted for pyrosequencing (as outlined in section 2.3.5).

However, the amplicon produced by primers An7F & An1388R (~1380 bp) was too large to be used with 454 pyrosequencing (at time of sequencing maximum amplicon

size was around 800 bp)[*]. Hence this primer set was unsuitable for use with this technology. Therefore, in order to obtain a usable amplicon for pyrosequencing, amplification was attempted with the anammox-specific An7F primer and a universal bacterial reverse 16S rRNA primer. Potential reverse primers were initially checked for suitability by aligning them against anammox 16S rRNA sequences from NCBI and SILVA (Pruesse, *et al.*, 2007) and calculating primer sensitivity and specificity using ThermoPhyl (Oakley, *et al.*, 2011), BLAST (Altschul, *et al.*, 1990; Zhang, *et al.*, 2000) and ARB (Ludwig, *et al.*, 2004). Primer 518R (see Table 2.1) was highlighted as a suitable reverse primer as it targeted all known anammox 16S rRNA sequences and theoretical calculations showed that the primer could work thermodynamically with primer An7F. Primers An7F & 518R were used as a second round nested PCR on purified product from An7F & An1388R (see section 2.3 for full methods) as, although the addition of a universal reverse primer (*i.e.* 518R) would likely target organisms outside of the anammox clade, it was hoped that the specificity of the first round PCR would ensure that primarily anammox related sequences were obtained from the 2nd round reaction. This methodology produced a clear band of the correct size. Non-specific bands were also observed in this PCR but these were distinct from the band of correct size and the correct band was easily purified by gel extraction.

3.2.1.4. PCR Validation

Primers Amx368F & Amx820R were further validated as to their ability to be able to target non-Scalindua anammox organisms from environmental DNA. Unfortunately, due to the difficulties associated with investigating anammox organisms, such samples were not readily available. Nevertheless, a collection of samples were generously donated by several other research groups (see Table 3.1) which had been positively identified to contain a range of anammox genera. These samples were amplified using the suite of primers discussed above to test their specificity and their ability to amplify target DNA assessed on a simple presence/absence basis. All samples underwent a nested PCR approach with primers Pla46F & 1390R as a first round PCR. This PCR product was gel extracted and purified prior to the second round of amplification (see chapter 2.3.2.4). Samples which exhibited negative

[*] N.B. The amplicon of primers Amx368F & Amx820R/BS820R was already of a suitable size for use with pyrosequencing technologies.

results after this two-step PCR were then further optimised in order to ascertain that the lack of amplifiable product was indeed due to the inability of this primer set to amplify DNA from that sample and not due to other complications associated with PCR (*e.g.* impurities, low DNA yields *etc*). All samples which failed to be amplified after the nested PCR could not be amplified following this further optimisation.

Lead Researcher	Institution	Sample Location	Sample Description	Reference
Jakob Zopfi	University of Lausanne, Switzerland	Lake Neuchâtel, Switzerland	DNA from bulk soil	(Humbert et al., 2010)
Teruki Amano	Kyoto University, Japan	Haipong, Vietnam	DNA from sediment collected from Mangrove forest and shrimp pond sites	(Amano et al., 2011)
Josh Neufeld	University of Waterloo, Canada	Ontario, Canada	Cloned 16S rRNA genes extractd from groundwater samples	(Moore et al., 2011)

Table 3.1: Table of collaborations with other research groups who generously donated environmental DNA and positive controls which were used in the optimisation and validation of anammox specific PCR primers during this study.

	Sample	Quantity (ng/ul)	A260/A280	Description	Reported anammox genera
Zopfi Samples	Z1	132	1.68	DNA ext. from anammox enriched soil	Unknown
	Z2	82	1.68	DNA ext. from soil LnA	*Ca.* Brocadia spp.
	Z3	201	1.85	Plasmid DNA from enrichment culture	*Ca.* Jettenia sp.
Amano Samples	MF1	109	1.8	Mangrove Forest 1	All anammox (dominated by *Ca.* Scalindua spp.)
	MF2	157	1.7	Mangrove Forest 2	All anammox (dominated by *Ca.* Scalindua spp.)
	SP1	134	1.58	Shrimp Pond 1	All anammox (dominated by *Ca.* Kuenenia spp.)
	CH	67	1.63	Channel	All anammox (dominated by *Ca.* Scalindua spp.)

Table 3.2: Identity of individual samples donated by the Zopfi and Amano Laboratories (see Table 3.1) including sample names, used throughout this study.

To further investigate the ability of primers Amx368F & Amx820R to amplify anammox DNA from non-Scalindua genera, a small clone library was constructed using the samples mentioned above (Table 3.2). PCR product was purified using a commercial PCR Purification Kit (Qiagen, UK) and then cloned into the pGEMT-Easy vector (Promega, UK). Cloned plasmids were transformed into JM109 competent *E. coli* cells (Promega, UK) and single colonies isolated. Colonies were

screened for containing the insert and then sequenced via Sanger sequencing (Natural History Museum, UK) using both the M13F and M13R primers. Five colonies were sequenced for each of the seven samples. Forward and reverse reads were checked and trimmed manually in SeqMan II (DNASTAR, USA) and assembled into contigs. Contigs were aligned against reference sequences and phylogenetic relationships calculated in MEGA4 and MEGA5 (Tamura, et al., 2007; Tamura, et al., 2011).

3.2.2. Genera Specific Anammox 16S rRNA Primers

Genus specific 16S rRNA PCR primers, targeting intra-anammox diversity, were also investigated during this study. A suite of primers targeting individual genera, which could be used in conjunction with primers specific for the entire anammox clade, would be advantageous when investigating anammox diversity. Both primers from the literature and primers designed during this investigation were assessed as to their ability to target specific groups of anammox organisms (Table 2.1).

Thermodynamically viable primer pairs were obtained using BatchPrimer3 (You, *et al.*, 2008). Primers were then checked against target and non-target (*i.e.* anammox and non-anammox) 16S rRNA sequences from the SILVA database (Pruesse, *et al.*, 2007) using ThermoPhyl (Oakley, *et al.*, 2011). Potential primers were then further checked for specificity and nucleotide mismatches with target sequences using ARB (Ludwig, *et al.*, 2004). This was necessary as the position of mismatches (either towards the 5' or 3' ends or middle of the primer) and whether these mismatches represented true differences between the primer and reference sequences or were as a result of missing sequence data needed to be known in order to assess the suitability of each primer set. Such data were not provided by Thermophyl. Primer sets which appeared to show the desired specificity were tested empirically in the lab using the samples outlined in section 3.2.1.2.

3.3. Results and Discussion

3.3.1. Amplification Using Primers An7F & An1388R

3.3.1.1. Overview

Initially primers An7F & An1388R (Penton, *et al.*, 2006) were investigated as potential candidates for the specific amplification of anammox 16S rRNA genes. These primers were chosen as they had been reported to have a higher affiliation for anammox sequences than other 16S rRNA primers in the literature (Engstrom, *et al.*, 2009), had been shown to be able to amplify environmental anammox DNA (Moore, *et al.*, 2011) and produced the longest amplicon of any anammox specific primer set (Table 2.1) and thus were deemed more likely to produce more informative sequence data.

3.3.1.2. Primer Specificity

Investigations in the laboratory were promising and appeared to show that primers An7F & An1388R were suitable to be used as a tool for investigating anammox ecology (Figure 3.1). After optimisation, PCRs using these samples showed that these primers amplified all known anammox positive controls efficiently and also produced clear bands from DNA extracted from Medway Bridge Marina environmental samples, where anammox activity had previously been shown (Rooks, *et al.*, 2012). Furthermore, all negative PCR controls were not amplified by these primers with the exception of *P. maris* DNA. However, *P. maris* is one of the more closely related non-anammox *Planctomycetes* (based on 16S rRNA gene similarity) and therefore, if this primer set amplified the entire known anammox 16S rRNA gene diversity then it was reasonable to assume that sequences branching just outside of this clade may also be amplified.

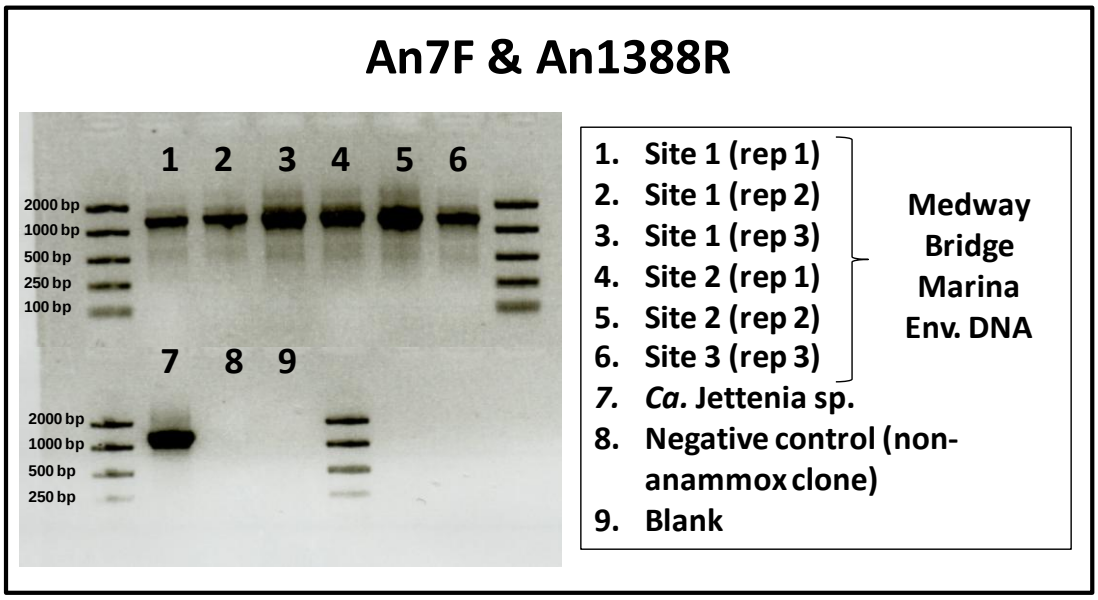


Figure 3.1: Agarose gel image of environmental and control DNA amplified by primers An7F & An1388R. PCR primers An7F & An1388R amplified DNA from both environmental samples known to contain anammox organisms and DNA isolated from anammox bacteria. DNA from cloned or isolated, non-anammox organisms (which were available in the laboratory) were not amplified by these primers.

However, despite satisfactory results using these primers in the laboratory, results from 454 pyrosequencing data were not as favourable. DNA extracted from Medway Estuary sites M6-1 and M6-2 was sequenced in order to investigate the specificity of these primers. From the two samples a total of 15,008 raw reads were obtained of which 3,401 passed quality control. Reads were clustered at 95% producing 510 potential OTUs. OTUs representing less than 0.1% of the total number of reads (*i.e.* < 4 reads/OTU representing 350 OTUs) were deemed insignificant and were omitted from downstream processing[*]. These 350 OTUs represented a total of 499 reads. It could be argued that such a large number of sequences may represent an important proportion of the population; nevertheless it was unlikely that these would alter the conclusions which would be drawn from these data. Reference sequences from each OTU which passed these criteria were aligned against the entire nucleotide database at NCBI using BLAST (Altschul, *et al.*, 1990; Zhang, *et al.*, 2000), the output of which was parsed in order to report only the highest scoring "hit" (based on e-values) for each.

[*] For a full description and discussion of the bioinformatics pipeline used please refer to section 5.3.

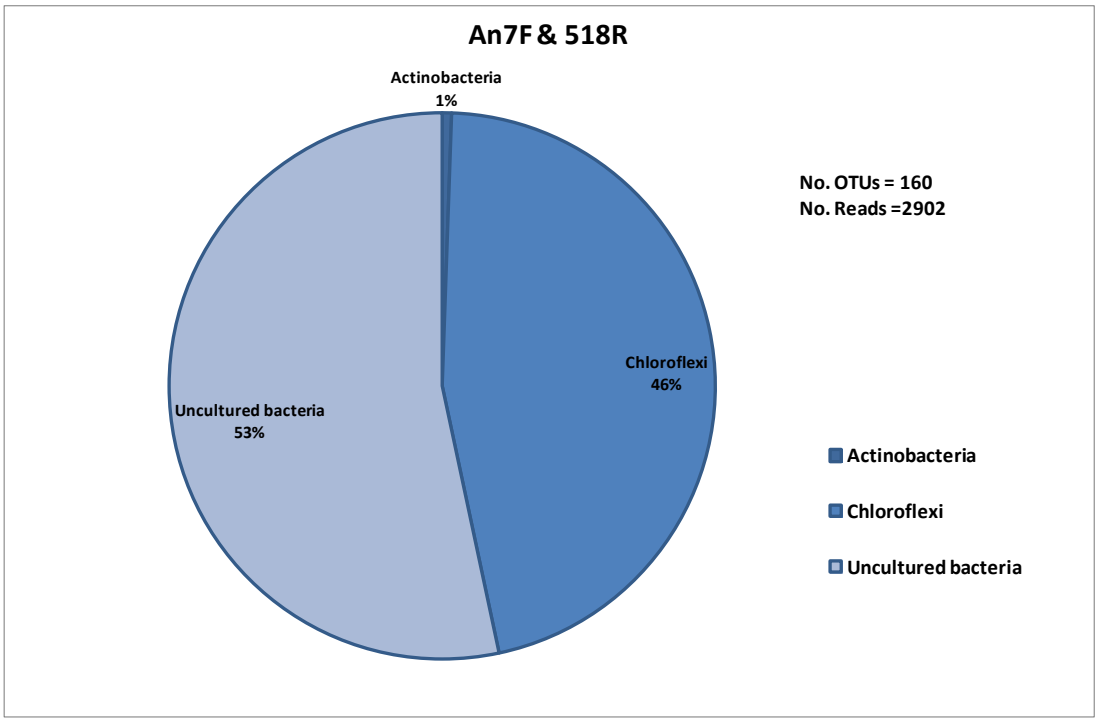


Figure 3.2: Percentage of 454 pyrosequencing reads, sequenced from amplicons using primers An7F & 518R associated with different bacterial taxa. See text for methodology. OTUs were defined at 95% similarity. A total of 510 OTUs were produced by this clustering cut-off. OTUs which represented <0.1% of the total number of reads (n=3401) have been omitted. Reference sequences for each of the remaining 160 OTUs were aligned against the NCBI nucleotide database using BLAST (Altschul, et al., 1990). BLAST hits were parsed to return the top hit (based on e-values) for each sequence if the percentage identity was greater than 50%. None of the sequences was found to be directly related to an anammox organism. The vast majority of BLAST hits returned "uncultured bacteria" sequences (53.34%) none of which were sequenced from studies especially looking for anammox bacteria though several were sequenced from environments where anammox have previously been found (e.g. the Black Sea, anoxic marine sediments, rice paddy fields, oil reservoirs).

As is evident from Figure 3.2, none of the 160 OTUs were directly associated with anammox sequences and represented a wide distribution of different micro-organisms suggesting that this primer set was not suitable for targeting and amplifying anammox 16S rRNA from the environment. 53.34% of the reads were associated with "uncultured bacterial" sequences from the NCBI database. It was not deemed necessary to investigate these sequences in depth though the environment from which they had been sequenced was recorded. None of the sequences were obtained from studies directly looking for anammox organisms in the environment, though several studies had investigated environments where anammox bacteria had previously been observed (*e.g.* oil reservoirs, the rhizosphere of rice crops and Black Sea sediments and free-floating aggregates). However, as none of the sequences were identified as anammox organisms, it is clear that this methodology was not

suitable for isolating and amplifying anammox 16S rRNA sequences from the environment. Rarefaction curves for the data (Figure 3.3) were plateauing in both samples for both the entire 510 OTUs and the 160 OTUs which passed the >0/1% cut-off. Therefore it can be assumed that the depth of sequencing was suitable to target the entire diversity of sequences amplified by the primers and so the lack of anammox related sequences cannot be explained by a poor sequencing effort. Even if the 350 OTUs omitted from this analysis had all represented anammox organisms, which is unlikely, this would not have affected these conclusions as the majority of sequences would still have been identified as belonging to non-anammox organisms. Hence, it was concluded that primers An7F & An1388R/518R were not suitable to investigate anammox 16S rRNA diversity. Furthermore, the depth of 454 pyrosequencing appeared to highlight potential short-fallings within the primers which were not observed whilst testing primers empirically either *in vitro* or *in silico*. Thus reinforcing the need for a specific and efficient primer set for the detection and isolation of anammox 16S rRNA sequences.

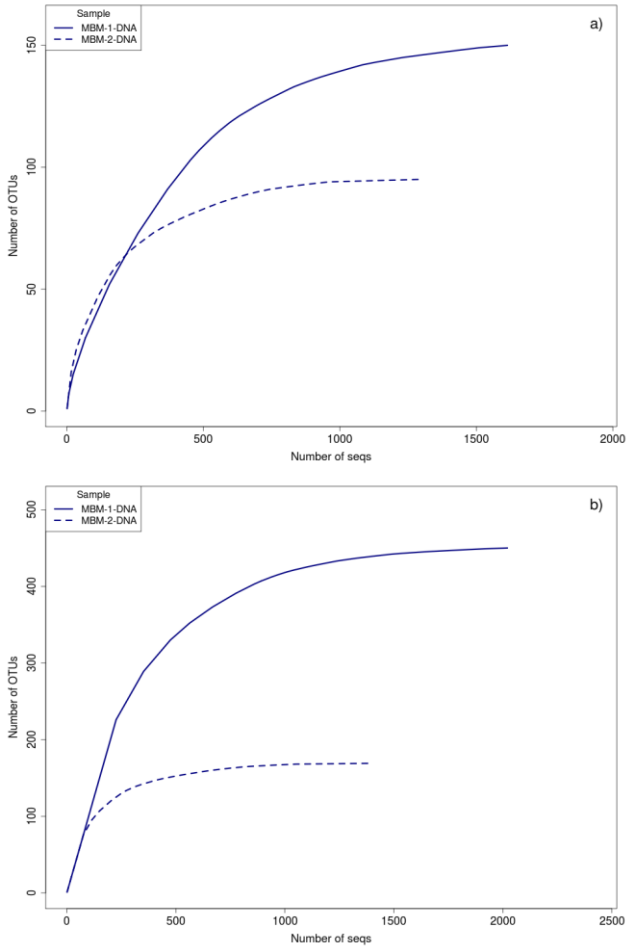


Figure 3.3: Rarefaction curves for Medway Bridge Marina data pyrosequenced using primers An7F & 518R. a) only those OTUs which satisfied the >0.1% of total reads criteria and b) all OTUs. Clustering was performed at 95% similarity. As can be seen, all curves can be seen to begin to plateau. Therefore it can be assumed that this sequencing effort represents most of the diversity amplified by this primer set.

3.3.2. Amplification using Primers Amx368F & Amx820R/BS820R

3.3.2.1. Overview

Due to the lack of specificity of primers An7F & An1388R/518R in the amplification of anammox 16S rRNA genes from the environment, further primer sets were investigated as to their suitability in specifically targeting this group of organisms. A suitable primer set would specifically target the entirety of the known anammox diversity, showing little or no bias towards specific genera within the family *Brocadiaceae*, be able to reliably and efficiently amplify environmental DNA and be of a suitable length for 454 pyrosequencing. Of these primer sets, primers Amx368F (Schmid, *et al.*, 2000) & Amx820R (Schmid, *et al.*, 2003) or BS820R

(Kuypers, *et al.*, 2003) emerged as the best candidates for an efficient and specific primer set for targeting anammox 16S rRNA diversity via PCR.

3.3.2.2. Primer Specificity

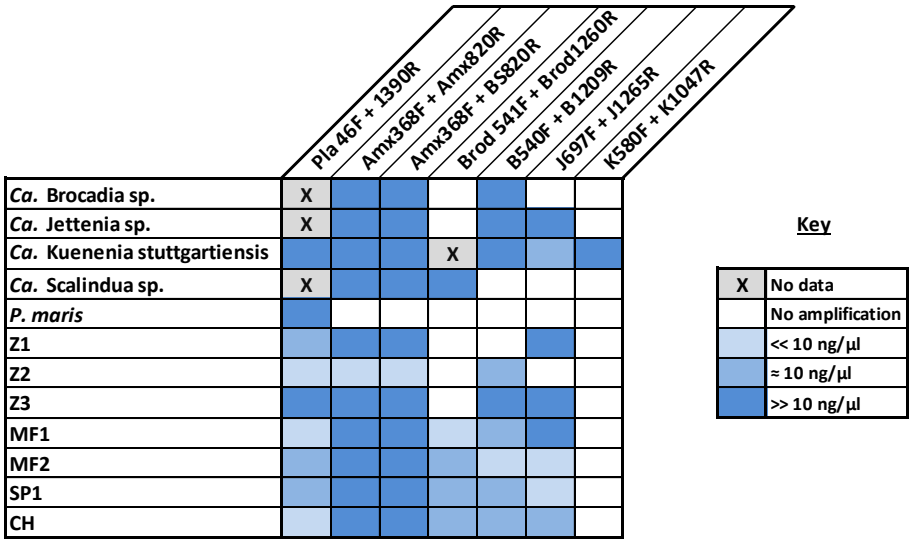


Figure 3.4: Heatmap showing specificity of 16S rRNA primers used during this study. All anammox samples were amplified effectively by primers Amx368F & Amx820R and Amx368F & BS820R indicating that they are capable of amplifying DNA from all known anammox genera. Genus specific primers also showed suitable specificity. Primers Brod541F and Brod1260R were shown to only amplify samples known to contain Ca. Scalindua spp. Primers B540F & B1209R appeared to amplify DNA from the Brocadia, Jettenia and Kuenenia genera whereas primers J697F & J1265R showed similar specificity but appeared unable to amplify Ca. Brocadia spp. DNA. Primer K580F & K1047R, designed to be specific for Ca. Kuenenia spp., have only been shown to be able to amplify DNA from Ca. K. stuttgartiensis and have not been shown to amplify environmental DNA.

These primers continually produced a large, single band from both positive controls and environmental DNA (see Figure 3.4, Figure 3.5 and Figure 3.6) and did not amplify DNA from organisms known to be outside of the anammox clade (*e.g. P. maris*). However, in samples where the abundance of anammox 16S rRNA genes may have been low, it was necessary to perform a nested PCR (see chapter 2) with primers Pla46F (Neef, *et al.*, 1998) & 1390R (Zheng, *et al.*, 1996) as suggested by Humbert, *et al.* (2010). The size of the amplicon from these primers (~480 bp) was also suitable for use with 454 pyrosequencing technologies available at the time of optimisation.

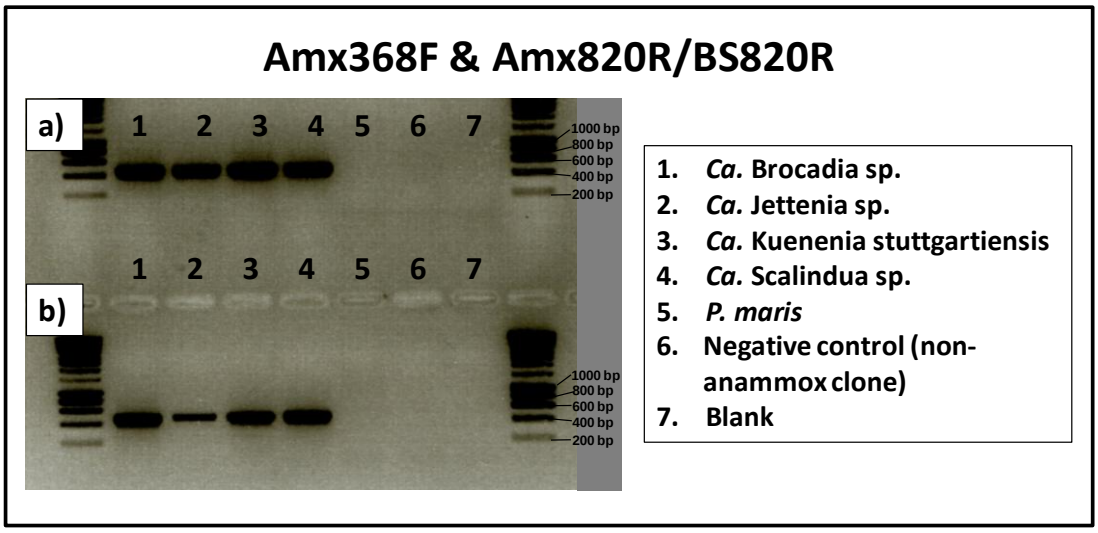


Figure 3.5: Agarose gel image of PCR products obtained from anammox control DNA using a) Amx368F & Amx820R and b) Amx368F & BS820R. Both primer sets amplified DNA from all anammox positive controls and did not amplify DNA from any of the negative controls (i.e. from non-anammox organisms).

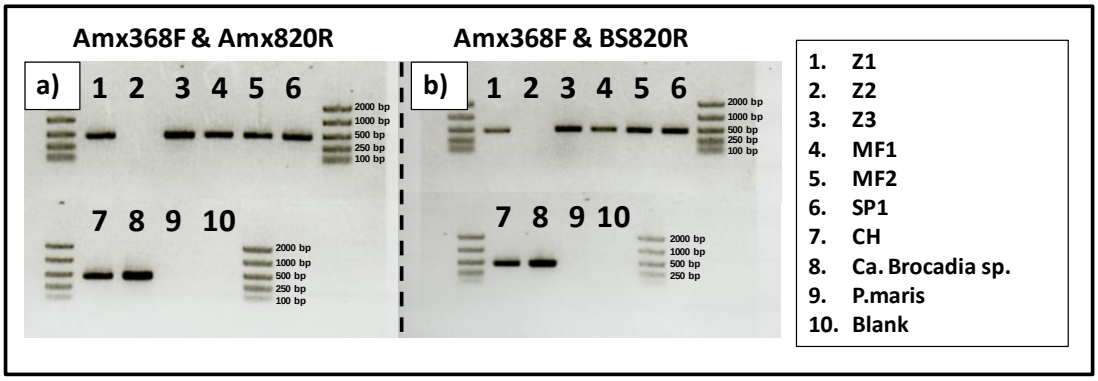


Figure 3.6: Environmental DNA samples amplified by primers a) Amx368F & Amx820R and b) Amx368F & BS820R. Sample names are described in Table 3.2. All samples were amplified by both primer sets with the exception of sample Z2. Sample Z2 was amplified by both of these primer sets after purification of this sample, however this had not been undertaken at the time that these images were taken.

However, the results of this investigation did not corroborate with those reported from the literature concerning the specificity of Amx368F & Amx820R/BS820R. Similar studies, suggest that primers Amx368F & BS820R specifically target *Ca.* Scalindua spp. whereas primers Amx368F & Amx820R target *Ca.* Brocadia spp. and *Ca.* Kuenenia spp. (Humbert, *et al.*, 2010; Moore, *et al.*, 2011). Primers Amx820R and BS820R are identical except for 3 base pair changes in the last 5 bases at the 5' end (see Table 2.1). Nevertheless, PCR amplification with these primer sets showed no difference in their specificity; both primer sets successfully amplified all four anammox positive controls (namely *Ca.* Brocadia sp., *Ca.* Jettenia sp., *Ca.* Kuenenia

stuttgartiensis and *Ca.* Scalindua sp.) and PCR product was obtained using both primer sets from the same environmental samples.

It had previously been discovered (see above) that apparently positive results from initial primer specificity investigations did not necessarily lead to satisfactory results in the acquisition of high-throughput sequencing data, as the greater sensitivity of the latter tended to reveal inadequacies in the specificity of the primers. Samples collected from M6-1 and M6-2 were amplified using primers Amx368F & Amx820R/BS820R in triplicate and sequenced using 454 pyrosequencing (see section 2.3.5), in a repeat of the investigation outlined in section 3.3.1, to assess whether these primers were more suitable for use with this sequencing technology.

After quality control of pyrosequencing data obtained using primers Amx368F & Amx820R a total of 20,255 reads (out of 53,762) remained. These were clustered into OTUs at a threshold of 95% similarity. 30 OTUs were defined, 6 of which contained single sequences. OTUs representing less than 0.1% of the total number of reads (i.e. less than 20 reads) were omitted leaving 15 OTUs for downstream analysis. Rarefaction analysis (Figure 3.7) indicate that the depth of sequencing was probably sufficient to report the entire diversity observed within these samples using this primer set. Reference sequences from each OTU were aligned against the NCBI nucleotide database using BLAST (Altschul, *et al.*, 1990; Zhang, *et al.*, 2000). The top hit for each sequence (based on e-value scores) was selected in order to determine the best indicator of the identity of each OTU.

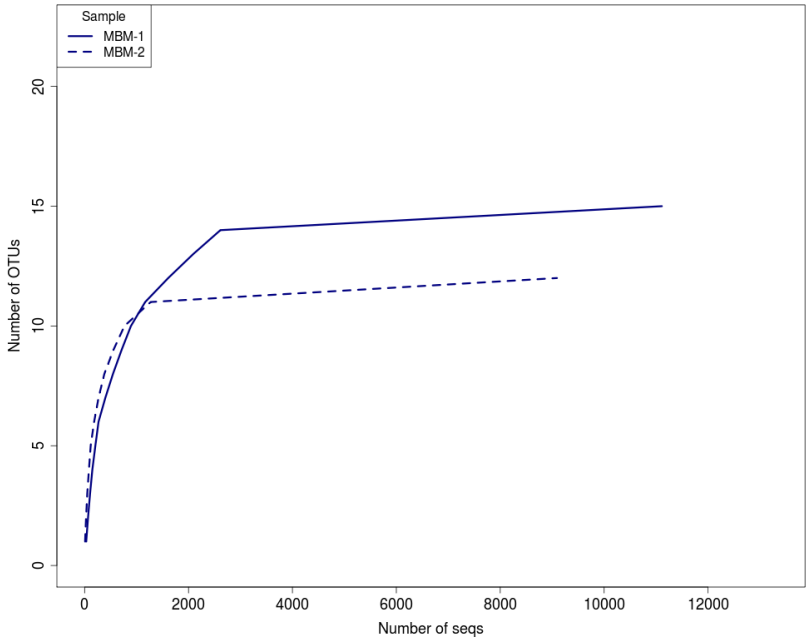


Figure 3.7: Rarefaction curves for 454 pyrosequencing data sequenced using primers Amx368F & Amx820R. Rarefaction curves show that, despite a low number of OTUs in each of the two samples, the entire diversity within these samples appears to have been reported and no significant OTUs are absent from the analysis.

As is evident from Figure 3.8, all OTUs were identified as most closely related to anammox organisms. 99.85% of the sequences were associated with *Ca.* Scalindua spp. with most of the sequences (89.22%) being most closely related to *Ca.* Scalindua marina. One OTU (n=31 reads) was identified as belonging to a non-Scalindua anammox organism, namely *Ca.* Brocadia caroliniensis.

Pyrosequencing data obtained with primers Amx368F & BS820R were analysed as above. Of a total of 31,678 reads 13,326 passed quality control. Clustering at a threshold of 95% similarity produced 25 OTUs of which 13 contained less than 0.1% of the total number of reads (*i.e.* 14 reads) representing a total of 31 reads. These were omitted resulting in 12 OTUs for downstream analysis.

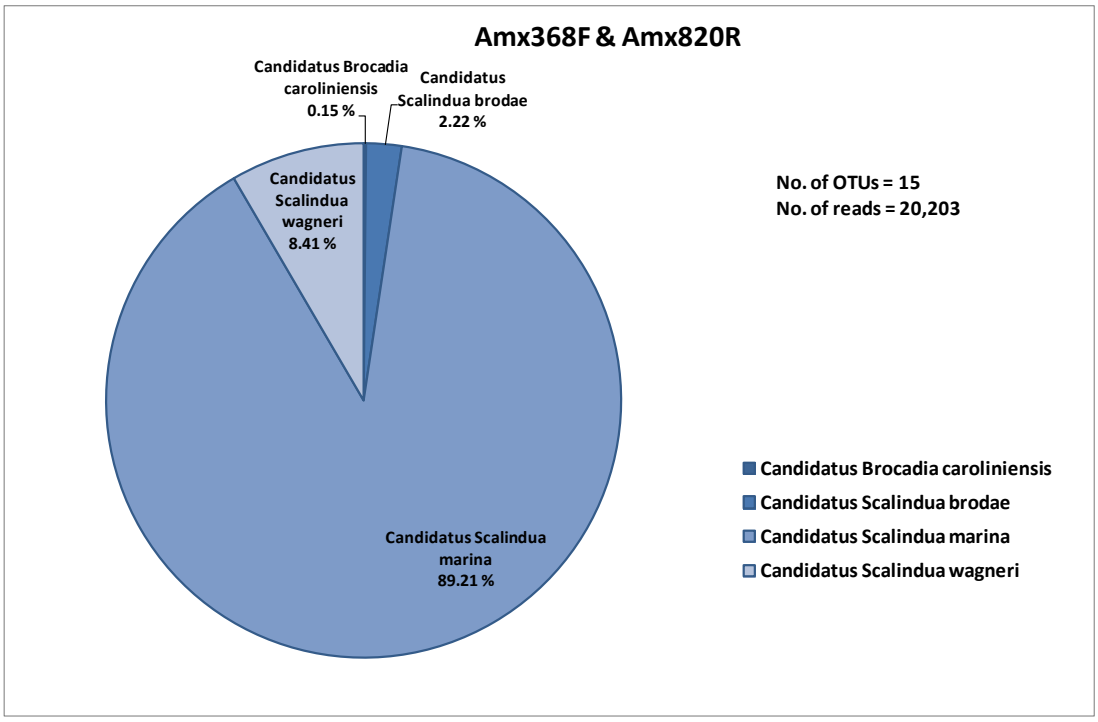


Figure 3.8: Pie chart showing the proportion of reads from 454 pyrosequencing data, sequenced using primers Amx368F & Amx820R, which were identified as different anammox organisms. OTUs were defined at a cutoff of 95% similarity and OTUs representing less than 0.1% of the total read abundance (n=20,255). 15 OTUs passed these criteria representing 20,203 reads. Reference sequences for each OTU were aligned against the NCBI nucleotide database using BLAST (Altschul, et al., 1990; Zhang, et al., 2000). Nucleotide sequences of uncultured organisms from NCBI were omitted from this BLAST query. BLAST results were parsed to report the top hit (i.e. lowest e-value) for each query and this hit was used to identify the OTU. 99.85% of the total reads were representative of Ca. S. spp. with only one OTU (n=31 reads) representing a non-Scalindua organism (namely Ca. B. caroliniensis).

Similar to data obtained with primers Amx368F & Amx820R, primers Amx368F & BS820R reported a population dominated by *Ca.* Scalindua spp. representing 99.87% of the total number of reads (Figure 3.9). Only one OTU was not identified as a *Ca.* Scalindua spp. but as an anaerobic ammonium-oxidising Planctomycete, however this sequence (Genbank accession number AB281489.1), was obtained from a marine environment which, according to the literature, is dominated by *Ca.* Scalindua spp. and thus this sequence may indeed represent this genus. This OTU however did not represent a significant proportion of the population encompassing only 0.14% of the total number of reads (n=18 reads).

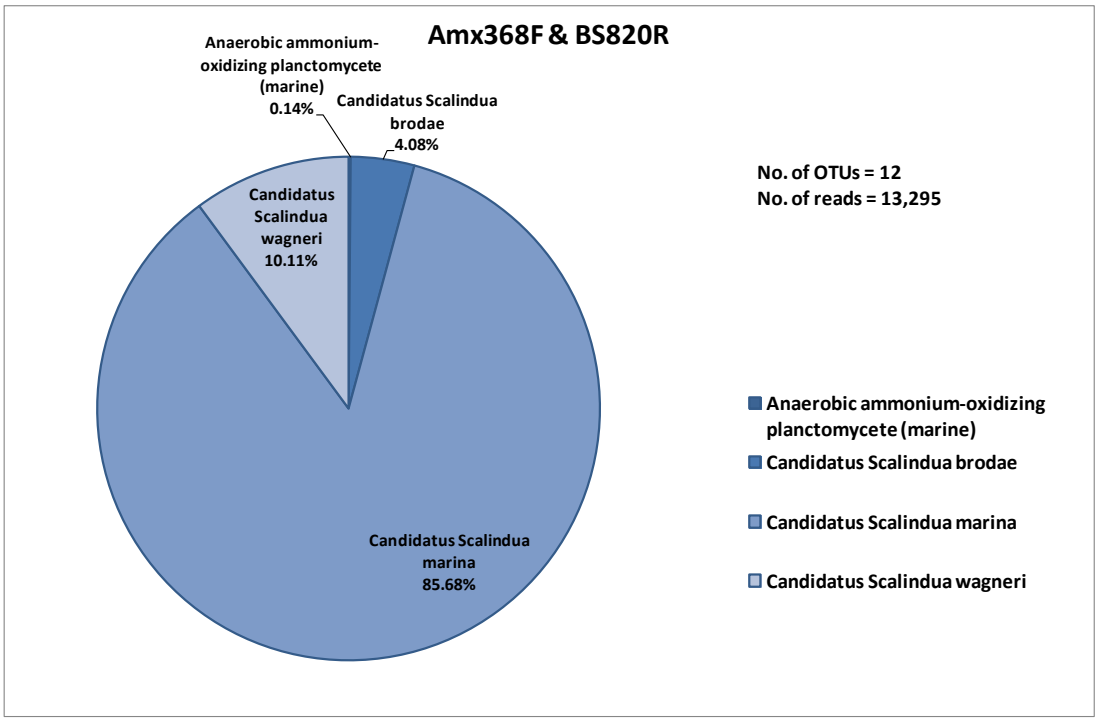


Figure 3.9: Pie chart showing the proportion of reads from 454 pyrosequencing data, sequenced using primers Amx368F & BS820R, which were identified as different anammox organisms. OTUs were defined at a cutoff of 95% similarity and OTUs representing less than 0.1% of the total read abundance (n=13,326). 12 OTUs passed these criteria representing 13,295 reads. Reference sequences for each OTU were aligned against the NCBI nucleotide database using BLAST (Altschul, et al., 1990; Zhang, et al., 2000). Nucleotide sequences of uncultured organisms from NCBI were omitted from this BLAST query. BLAST results were parsed to report the top hit (i.e. lowest e-value) for each query and this hit was used to identify the OTU. 99.87% of the reads were identified as Ca. Scalindua spp. with the majority of these identified as Ca. S. marina (85.68%). One OTU (n=18 reads) was identified as an anaerobic ammonium-oxidising planctomycete though this sequence was obtained from a marine environment and therefore is likely to also represent a Ca. Scalindua sp.

The results of this investigation appear to indicate that there is little difference between primers Amx368F & Amx820R and Amx368F & BS820R in terms of their specificity and amplification efficiency. Data from both primer sets produced a similar number of OTUs and were dominated by *Ca.* Scalindua spp. Non-Scalindua anammox genera represented a very small proportion of the total number of reads (<0.2% of reads) and may have even been absent in primers Amx368F & BS820R. Furthermore, although the absolute proportions of different Scalindua species was different, in both primer sets *Ca.* Scalindua marina represented the vast majority of Scalindua organisms (85-89%) with *Ca.* Scalindua wagneri being the second most abundant (8-10%) and *Ca.* Scalindua brodae being the least abundant (<4%). It is difficult to determine the relative diversity of these samples and primers (which was not the aim of this part of the study) without the presence of proper controls and

replication however it seems evident that there is little difference in either the specificity or efficiency of these two primer pairs. This is in direct contradiction with the literature which suggests that these primer sets, though specific to the *Brocadiales*, target different genera within this order namely *Ca.* Scalindua spp. and non-Scalindua organisms (Humbert, *et al.*, 2010; Moore, *et al.*, 2011).

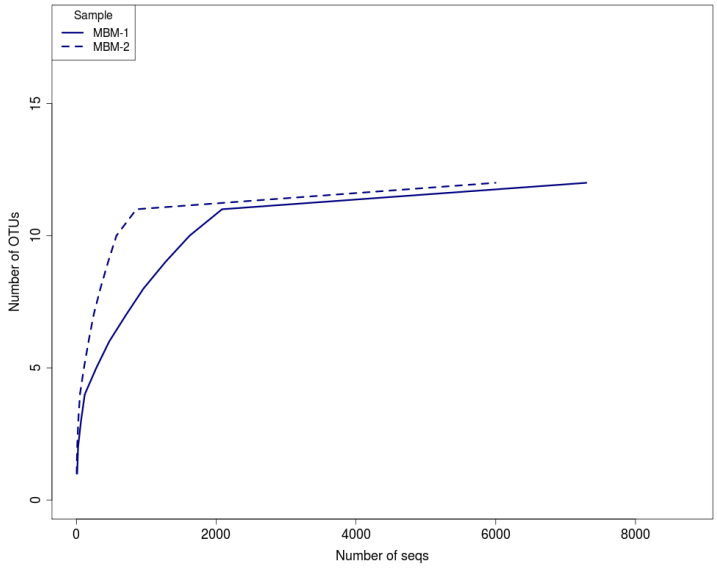


Figure 3.10: Rarefaction curves for 454 pyrosequencing data sequenced using primers Amx368F & BS820R. Rarefaction curves show that, despite a low number of OTUs in each of the two samples, the entire diversity within these samples appears to have been reported and no significant OTUs are absent from the analysis.

However, a new problem presented itself with the optimisation of these primer sets and analysis of 454 pyrosequencing data. Although these primers did not report the presence of any non-anammox organisms (at least after quality control) and therefore can be deemed to target only anammox organisms, nor did they detect significant non-Scalindua DNA. Primers Amx368F & Amx820R did however report one OTU which was positively identified as a non-Scalindua organism (namely *Ca.* Brocadia caroliniensis). Nevertheless, it was impossible to determine with the present data whether this apparent lack of non-Scalindua related sequences was as a result of bias within the primers or due to the actually diversity of anammox bacteria within the samples. These primers had successfully amplified cloned 16S rRNA gene fragments identified as non-Scalindua anammox but had not been shown to amplify such genera from environmental DNA. Such a question would need to be clarified before these primers could be utilised effectively to pursue the overall aims of this project.

3.3.2.3. Primer Validation

In order that the ability of primers Amx368F & Amx820R to amplify non-Scalinuda DNA from environmental samples might be assessed, it was necessary to obtain DNA or samples from environments which had previously been shown to contain such organisms. Such samples had been generously donated by several other research groups (see Table 3.1) which had been positively identified to contain a range of anammox genera. These samples were amplified using the suite of primers discussed above to test the specificity of each and their ability to amplify target DNA assessed on a simple presence/absence basis. All samples underwent a nested PCR approach with primers Pla46F & 1390R as a first round PCR. This PCR product was gel extracted and purified prior to the second round of amplification (see section 2.3.2.4). Samples which exhibited negative results after this two-step PCR were then further optimised in order to ascertain that the lack of amplifiable product was indeed due to the inability of this primer set to amplify DNA from that sample and not due to other complications associated with PCR (*e.g.* impurities, low DNA yields *etc*). All samples which failed to be amplified after the nested PCR could not be amplified following this further optimisation and therefore did not amplify these samples due to the specificity of the primers in question.

As can be seen in Figure 3.4, all primers used within this investigation demonstrated the desired specificity. Primers Pla46F & 1390R amplified all samples with the exception of the positive controls for *Ca.* Brocadia sp., *Ca.* Jettenia sp. and *Ca.* Scalindua sp., however, these three samples are cloned inserts and the PCR primers used to clone them produced an amplicon which was shorter than that of primers Pla46F & 1390R. Furthermore DNA extracted from the *P. maris* culture was not amplified by any of the anammox specific primer sets suggesting that amplification by these primers was restricted to the anammox clade.

Both primer sets Amx368F & Amx820R and Amx368F & BS820R appeared to exhibit the same specificity within the anammox clade as they both amplified all positive control samples and all environmental samples. Furthermore the amplification efficiency of these two primer sets also appeared to be similar as the intensity of amplified bands appeared to be similar between them (*i.e.* if a band was weakly amplified by Amx368F & Amx820R the same was true using primers

Amx368F & BS820R). These primers produced PCR products from samples which had been reported to contain only non-Scalindua anammox genera indicating that these primers are not specific to *Ca.* Scalindua spp (Figure 3.4 and Table 3.2).

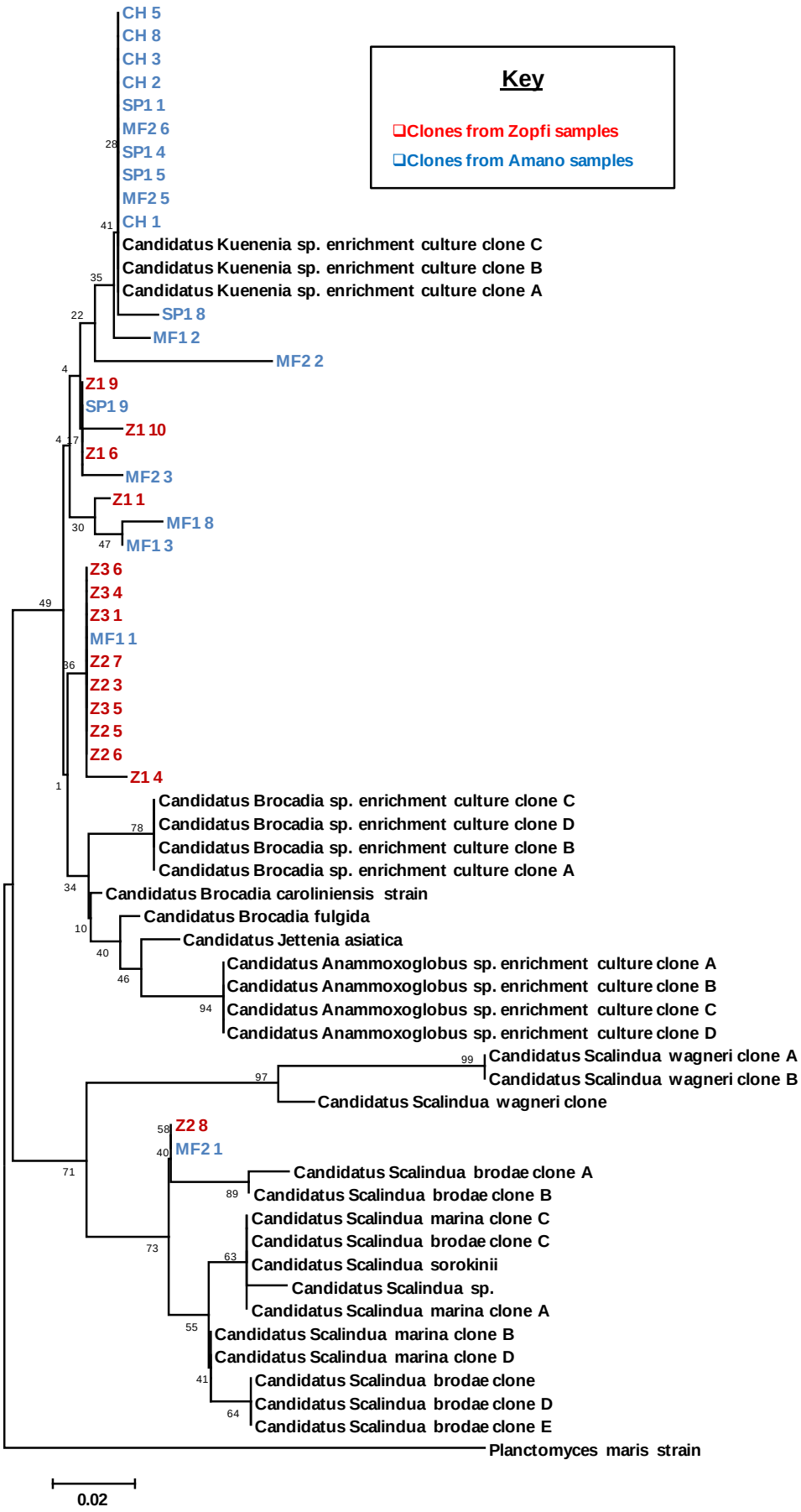

Key
Clones from Zopfi samples
Clones from Amano samples
CH 5
CH 8
CH 3
CH 2
SP1 1
MF2 6
SP1 4
SP1 5
MF2 5
CH 1
Candidatus Kuenenia sp. enrichment culture clone C
Candidatus Kuenenia sp. enrichment culture clone B
Candidatus Kuenenia sp. enrichment culture clone A
SP1 8
MF1 2
MF2 2
Z1 9
SP1 9
Z1 10
Z1 6
MF2 3
Z1 1
MF1 8
MF1 3
Z3 6
Z3 4
Z3 1
MF1 1
Z2 7
Z2 3
Z3 5
Z2 5
Z2 6
Z1 4
Candidatus Brocadia sp. enrichment culture clone C
Candidatus Brocadia sp. enrichment culture clone D
Candidatus Brocadia sp. enrichment culture clone B
Candidatus Brocadia sp. enrichment culture clone A
Candidatus Brocadia caroliniensis strain
Candidatus Brocadia fulgida
Candidatus Jettenia asiatica
Candidatus Anammoxoglobus sp. enrichment culture clone A
Candidatus Anammoxoglobus sp. enrichment culture clone B
Candidatus Anammoxoglobus sp. enrichment culture clone C
Candidatus Anammoxoglobus sp. enrichment culture clone D
Candidatus Scalindua wagneri clone A
Candidatus Scalindua wagneri clone B
Candidatus Scalindua wagneri clone
Z2 8
MF2 1
Candidatus Scalindua brodae clone A
Candidatus Scalindua brodae clone B
Candidatus Scalindua marina clone C
Candidatus Scalindua brodae clone C
Candidatus Scalindua sorokinii
Candidatus Scalindua sp.
Candidatus Scalindua marina clone A
Candidatus Scalindua marina clone B
Candidatus Scalindua marina clone D
Candidatus Scalindua brodae clone
Candidatus Scalindua brodae clone D
Candidatus Scalindua brodae clone E
Planctomyces maris strain
0.02

Figure 3.11: Phylogenetic tree showing identity of clones sequenced from environmental samples using primers Amx368F & Amx820R. For sample names see Table 3.2. All clones are annotated as the name of the sample followed by the clone identifier number. Red highlighted clones were obtained from samples donated by the Jakob Zopfi Laboratory, University of Laussane, Switzerland whereas those highlighted in blue were donated by the Amano Teruki Laboratory, Kyōto University, Japan. The tree was constructed using the Neighbour-Joining method using p-distances in Mega 5.10 (Tamura, et al., 2011). A bootstrap test (1000 replicates) was conducted on the phylogeny; branch annotations represent the percentage of replicate trees in which the associated taxa clustered together. Scale bar represents 2% sequence similarity. Clones were sequenced which were found to have a high degree of similarity with all known anammox genera therefore demonstrating the ability of these PCR primers to amplify DNA from the entire anammox diversity.

As can be seen from Figure 3.11, sequences were obtained from these samples using primers Amx368F & Amx820R which showed a high similarity with 16S rRNA sequences obtained from *Ca.* Brocadia spp., *Ca.* Jettenia spp., *Ca.* Kuenenia spp. and *Ca.* Scalindua spp. The majority of sequences belonged to either the Jettenia or Kuenenia genera. It is unlikely that the prevalence of sequences from these two genera represent bias within the sample as previous investigations on samples from the Medway Estuary indicate anammox populations dominated by *Ca.* Scalindua spp. (see above). Thus these primers are capable of amplifying anammox 16S rRNA genes in both *Ca.* Scalindua spp. and non-Scalindua dominated communities. As such it appears that the primers Amx368F & Amx820R are suitable to be used within this study to investigate anammox ecology as they are both specific to the order *Brocadiaceae* and have been shown to target all the known genera within this family.

3.3.3. Genera Specific Anammox Primers

Primers Brod541F & Brod1260R demonstrated the specificity reported in the literature in that they appeared to be specific to *Ca.* Scalindua spp. (Penton, *et al.*, 2006). Of the four anammox positive controls only DNA from *Ca.* Scalindua sp. was amplified (Figure 3.12). Also, of the environmental samples, no PCR product was obtained from samples Z2 or Z3 which were not reported to contain *Ca.* Scalindua spp. (Humbert, *et al.*, 2010) but did amplify product from MF1, MF2, SP1 and CH which had tested positive for this genera (Amano, *et al.*, 2011).

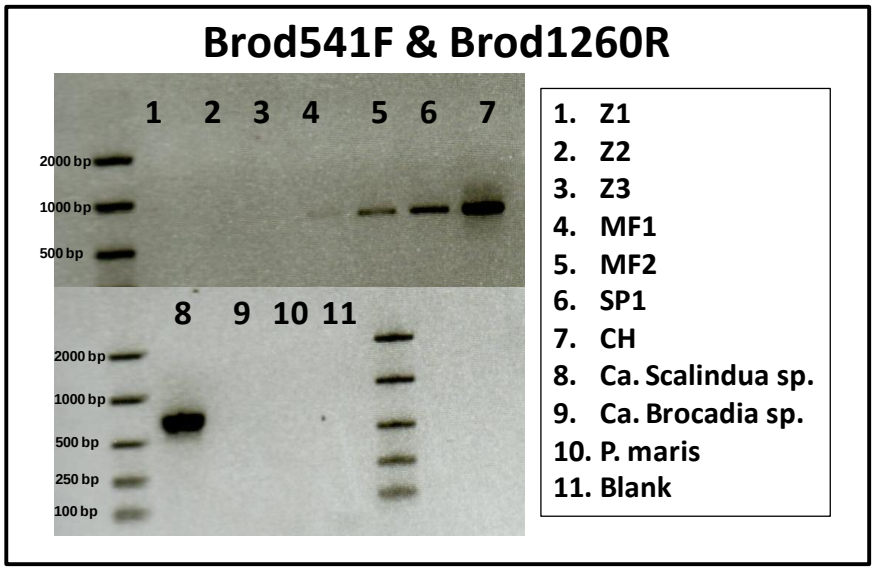


Figure 3.12: Environmental DNA samples amplified by primers Brod541F 7 Brod1260R. For sample names see Table 3.2. DNA from samples Z1-Z3 (which were not reported to contain Ca. Scalindua spp, DNA) were not amplified by these primers whereas samples MF1, MF2, SP1 and CH were amplified. Furthermore these primers amplified the Ca. Scalindua sp. positive control but did not amplify any of the other anammox clones or any non-anammox clones or isolates.

The genera specific primers designed during this study to target intra-anammox diversity also appeared to show a reasonably satisfactory specificity. Primers B540F & B1209R, designed to specifically target *Ca.* Brocadia spp., were unable to amplify *Ca.* Scalindua sp. or *P. maris* DNA but amplified all other anammox controls (Figure 3.4 and Figure 3.13). Furthermore they amplified product from all of the environmental DNA samples (Figure 3.4 and Figure 3.13). However this primer set does not appear to be specific to only *Ca.* Brocadia spp., but also targets Kuenenia and Jettenia genera. Both of these controls were strongly amplified as well as sample Z3 (plasmid DNA from an anammox enrichment culture which predominantly consisted of *Ca.* Jettenia spp).

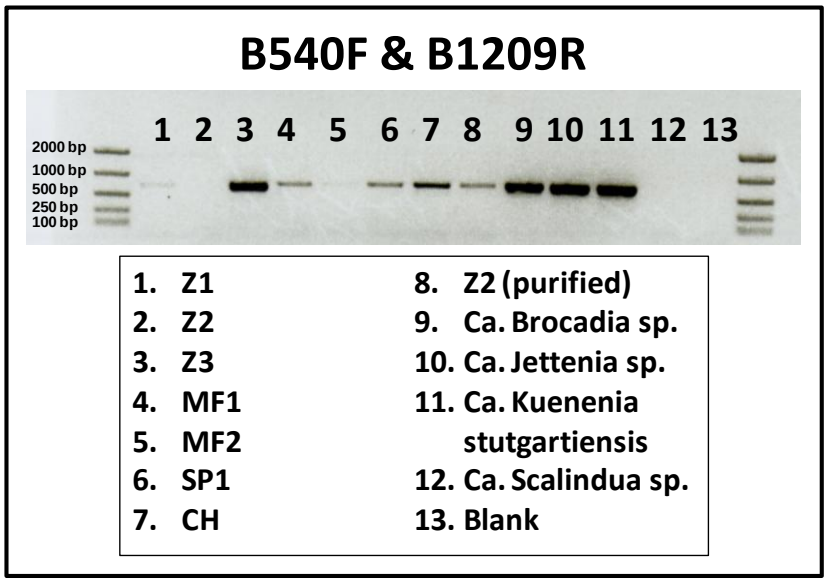


Figure 3.13: Environmental DNA and anammox controls amplified by primers B540F & B1209R. For sample names see Table 3.1. Primers amplified all environmental samples (Z2 was only amplified after purification of DNA c.f. lanes 2 and 8). All anammox controls were amplified except for Ca. Scalindua sp., suggesting these primers are specific to non-Scalindua anammox organisms.

Primers J697F & J1265R, designed to solely target *Ca.* Jettenia spp., also did not show the desired specificity. Although they were shown to not amplify DNA from *Ca.* Scalindua spp. or *Ca.* Brocadia spp. (amplifying neither the control sample or sample Z2) they did amplify *Ca.* K. stuttgartiensis DNA (Figure 3.14). However, the high similarity of the 16S rRNA gene between Brocadia, Jettenia and Kuenenia genera (Jetten, *et al.*, 2009) may make it difficult or even impossible to design primers which can specifically amplify these three genera independently. However, these primers do not appear to amplify the 16S rRNA gene of *Ca.* Scalindua spp. and so could be utilised as an effective molecular marker for non-Scalindua anammox organisms. Furthermore, as primers J697F & J1265R do not appear to amplify *Ca.* Brocadia spp. DNA, they could be used to indirectly identify the presence of this genus as, if PCR product was amplified from a sample by primers B540F & B1209R and not J697F& J1265R, then it could be concluded that the sample contained *Ca.* Brocadia spp. but not *Ca.* Jettenia spp. or *Ca.* Kuenenia spp.

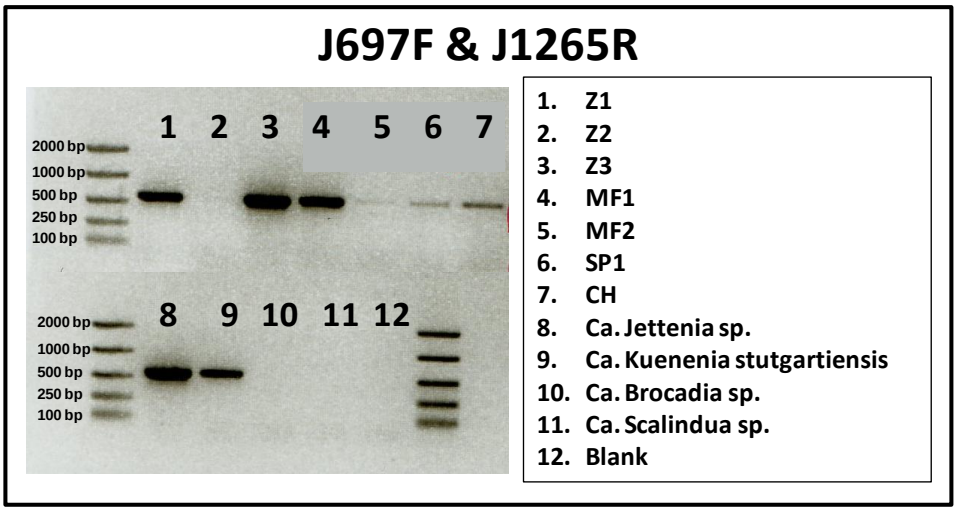


Figure 3.14: Environmental DNA and anammox controls amplified by primers J697F & J1265R. For sample names see Table 3.2. Primers amplified all environmental DNA samples and DNA from Ca. Jettenia sp. and Ca. Kuenenia stutgartiensis. Primers were not able to amplify Ca. Brocadia sp. or Ca. Scalindua sp. DNA.

Primers K580F & K1047R appear to be specific to the Kuenenia genus, only amplifying DNA from *Ca.* K. stuttgartiensis. However, to date, no product has been obtained from PCRs on environmental DNA using these primers. This may be due to inadequacies with this primer set (*e.g.* as to its specificity) or because no sample has been analysed which contained any target DNA. Samples MF1, MF2, SP1 and CH had been reported to contain *Ca.* Kuenenia spp. (Figure 3.11 and Table 3.2) but were not amplified by these primers. Therefore these primers may not target the whole Kuenenia genus but only a sub-division of it (*e.g.* only *Ca.* K. stuttgartiensis). However, it is unable to answer this question with the present data and it did not appear to be a sufficiently important one to warrant further analysis during this project.

3.4. Conclusion

The data presented in this study demonstrate that primers An7F & An1388R are not suitable for use as a phylogenetic marker for anammox organisms. Although the primers were capable of amplifying anammox DNA via PCR, the data also indicate that many non-anammox organisms were targeted by these primers. This was most evident in data obtained from 454 pyrosequencing which show that the majority of reads were associated with 16S rRNA genes from non-anammox organisms.

Due to the failings of these primers in their ability to specifically target only anammox 16S rRNA genes, it was necessary to investigate other primers in order to achieve the aims of this project. Primers Amx368F & Amx820R demonstrated the ability to both solely target anammox 16S rRNA genes and also to report the entire known diversity of anammox organisms. Pyrosequencing data obtained using these primers demonstrate that all reads (after quality control) were associated with anammox or anammox-related sequences. As such, these primers appear suitable for accurately reporting anammox phylogenetic diversity and are a suitable tool for investigating anammox ecology.

PCR primers targeting 16S rRNA genes specific to individual anammox genera were also investigated in this study. Primers Brod541F & Brod1260R were shown to be able to specifically target *Ca.* Scalindua spp. Primers were also designed which were show to be able to target non-Scalindua (*i.e.* B540F & B1209R) and *Ca.* Jettenia spp. and *Ca.* Kuenenia spp. (*i.e.* J697F & J1265R). The ability to use these intra-anammox specific primers within a nested PCR allows them to be used as a potentially powerful tool for anammox ecology. Though the power of this technique would be far less than other molecular ecological methods (*e.g.* high-throughput sequencing), the ability to use this suite of primers to obtain an approximate understanding of anammox diversity within an environmental sample (at least in terms of the presence/absence of specific anammox genera and hence an estimation of the richness of the sample) would be advantageous. The ability to obtain results from this method within a day makes it a more useful tool for investigating anammox ecology. However, conclusions drawn from this method would have to be used tentatively in anticipation of more extensive data in terms of the presence of anammox genera within the sample in question.

3.5. Summary

- The investigation of anammox ecology is a challenging subject and, in the absence of classic ecological methods, requires the development of a reliable and efficient suite of culture-independent techniques.

- The onset of high-throughput sequencing technologies, gleaning ever-increasingly greater amounts of data, may potentially reveal problems

with the specificity of molecular markers used to investigate anammox (and other bacteria) in the environment.

- PCR primers Amx368F & Amx820R appear to be the best candidates, of the primers investigated, to specifically amplify anammox 16S rRNA sequences from environmental samples and to be able to target the entire known anammox diversity.

4. Potential Use of Functional Genes as Molecular Markers for Anammox Diversity

4.1. Introduction

4.1.1. Overview

Following the successful development of an effective suite of anammox specific PCR primers targeting the 16S rRNA gene (see chapter3), the possibility of developing primers targeting functional molecular markers for the activity of anammox bacteria was also investigated. The ability to investigate one or more functional genes would not only allow for the investigation of anammox activity (through gene expression) within an environmental sample but also for the collaboration of measurements of diversity. That is, if the functional diversity within a particular gene represented the phylogenetic diversity of the entire organism, then the diversity of this gene could be used in conjunction with the 16S rRNA gene in order to gain more confidence about the diversity reported in the latter. Furthermore, it has been suggested that the conserved nature of the 16S rRNA gene may limit its effectiveness as a tool to measure anammox diversity and is not directly linked to function (Hirsch, *et al.*, 2011). However such a methodology would depend on the existence of a suitable gene, which was both unique to and ubiquitous in anammox bacteria, and the development of reliable primers which would target only the gene in question and, ideally, the entire diversity of that gene. To date, in the literature, three potential functional genes for the detection of anammox bacteria have been suggested; a nitrite reductase (*nirS*) (Li, *et al.*, 2011), hydrazine oxidoreductase (*hzo*) (Hirsch, *et al.*, 2011) and hydrazine synthase (*hzs*) (Harhangi, *et al.*, 2012).

4.1.2. Nitrite Reductase (*NirS*)

The reduction of nitrite is a crucial step in denitrification but is also utilised by many organisms to combat the toxic effect of NO_2^- accumulation within the cell (Klotz, *et al.*, 2008). *nirS* has been shown to be present in the genomes of *Ca.* K. stuttgartiensis and *Ca.* S. profunda (Strous, *et al.*, 2006; van de Vossenberg, *et al.*, 2012) and is thought to be an important step in anammox metabolism (Kartal, *et al.*, 2011a; van Niftrik and Jetten, 2012). As such *nirS* has been suggested to be an effective functional marker for anammox (Lam, *et al.*, 2009; Li, *et al.*, 2011). However *nirS*

genes are also found in numerous other organisms including denitrifiers (Smith, *et al.*, 2007; Hira, *et al.*, 2012), which can exist in the same environments as anammox bacteria. Thus, it would be potentially difficult to confirm with certainty that a particular *nirS* sequence obtained from the environment was from either an anammox or denitrifying bacteria. Furthermore, *nirK*, reported to be unique to denitrifiers has also been discovered in KSU-1, a potential anammox organism from a continuous-flow reactor (Hira, *et al.*, 2012). In addition, *nirS* genes were not been found in anammox bacteria in a number of studies investigating the genomes of *Ca.* Brocadia fulgida (Gori, *et al.*, 2011), KSU-1 (Hira, *et al.*, 2012) and *Ca.* Jettenia asiatica (Hu, *et al.*, 2012) and therefore *nirS* genes may not be ubiquitous in anammox organisms.

Our knowledge of *nir* genes in relation to anammox bacteria is not complete and hence use of this gene as a potential functional marker for anammox must be viewed with caution. In addition, the findings of Li, *et al.* (2011) can be disputed as the tree of *nirS* sequences presented in their study bases the division of anammox and non-anammox related sequences on unsupported branches which appear to be insignificantly different. Thus the conclusion that sequences from the environment can be classified as belonging to one of these two groups and that the diversity of "anammox" *nirS* represents anammox phylogeny is, at present, unsubstantiated. Furthermore, the identification of clades within this tree appears to be founded on only one sequence from a known anammox organism (*i.e. Ca.* K. stuttgartiensis). The anammox *nirS* sequences presented in the supplementary information presented by Lam, *et al.* (2009), though they appear to show a more evident phylogenetic difference from denitrifier *nirS* genes, are still defined based on similarity to only three, definitively anammox-related sequences, of which only one belongs with any certainty to an anammox bacteria. As such, it was decided not to pursue the use of *nirS* as a potential functional marker for anammox further for the purposes of this study.

4.1.3. Hydrazine Oxidoreductase (*hzo*) and Hydrazine Synthase (*hzs*)

Anammox has been shown to exhibit a novel and unique metabolism in terms of the oxidation of NH_4^+ (van Niftrik and Jetten, 2012). This pathway involves the synthesis (from NH_4^+ and NO) of hydrazine (N_2H_4) and its subsequent oxidation to

N_2. The ability to utilise N_2H_4 in this way has not previously been observed within a bacterial species (Jetten, *et al.*, 2009). Thus, the genes responsible for this process, hydrazine oxidoreductase (*hzo*) and hydrazine synthase (*hzs*), provide a unique possibility to be effective as molecular markers for anammox function and phylogeny.

The *hzs* gene encodes for a protein that synthesises N_2H_4 and the formation of the N-N bond (Kartal, *et al.*, 2011a; Harhangi, *et al.*, 2012). Initial research into the use of *hzs* as a phylogenetic marker appear promising (Harhangi, *et al.*, 2012), however the role of this gene within anammox metabolism was only confirmed towards the end of this project and as such it was not possible to investigate its use within the time-scale of this study.

The *hzo* gene has been more extensively studied than *hzs* as a molecular marker for anammox and so was a more suitable candidate for measuring diversity. It therefore warranted further investigation within this study. A number of primers targeting the *hzo* gene have been developed and used within the literature (Hirsch, *et al.*, 2011). Schmid, *et al.* (2008) reported that the diversity of *hzo*/HZO could be clustered into 3 clades though only "*hzo* Cluster 1" was suitable as a marker for anammox diversity as it contained sequences obtained from all 5 anammox genera and the phylogeny was consistent with that observed within the 16S rRNA gene. The diversity within this cluster was proposed to be related to two copies of the *hzo* gene (*hzoA* and *hzo*B) found in the genome of KSU-1 (Shimamura, *et al.*, 2007). However, Hirsch, *et al.* (2011) reported a greater phylogenetic diversity within this cluster than previously suggested, though maintained that it was still congruous with anammox diversity. Furthermore, up to nine putative copies of *hzo*/*hao* genes have been reported in the genome of *Ca.* K. stuttgartiensis, indicating further issues with the potential use of *hzo* as a functional marker for anammox (Klotz, *et al.*, 2008; Kartal, *et al.*, 2011a).

4.2. Methods

A selection of different primers has been used in the literature to detect *hzo* "cluster 1" (Table 2.4). Primer sets hzocl1F1 & hzocl1R2 and hzocl1F1l & hzocl1R2 were used as the literature suggested these where the most efficient primers for targeting the diversity of "*hzo* cluster 1" and the amplicon size (470 bp) would have been

suitable for 454 pyrosequencing if this was deemed necessary. DNA was amplified from sediment samples collected from Medway Estuary Site 6 on (23/03/2010).

Reference sequences were acquired from GenBank (NCBI) which were submitted by Dang, *et al.* (2010) and Li, *et al.* (2010b). Sequences from these studies were chosen as they also used PCR primers reported to only target "*hzo* cluster 1". For a full description of the methods used including specific PCR conditions and sampling locations see section 2.3.2.2.

4.3. Results

4.3.1. PCR Efficiency

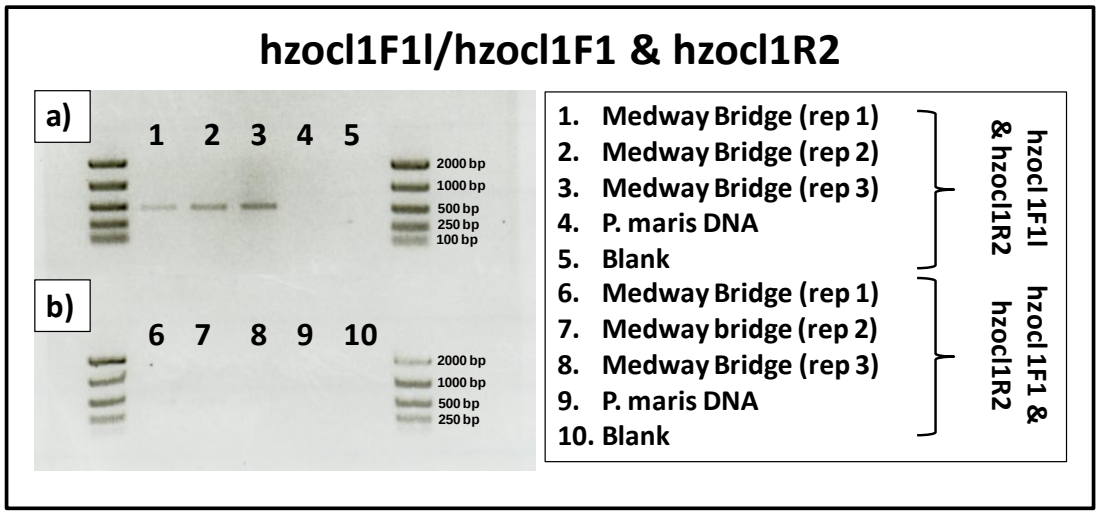


Figure 4.1: Agarose gel image of PCR product obtained using primers a) hzocl1F1l & hzocl1R2 and b) hzocl1F1 & hzocl1R2 from DNA extracted from Medway Estuary M6 sediment. PCR product from primers hzocl1F1l & hzocl1R2 was visible for all environmental replicates however no bands from primers hzocl1F1 & hzocl1R2 were evident. Negative and blank controls for both primer sets demonstrated no amplification. Increasing the intensity of UV light on this gel did show that PCR product had been obtained from primers hzocl1F1 & hzocl1R2 though the amount of product was minimal.

Difficulty was encountered when attempting to amplify *hzo* sequences using either primers hzocl1F1 & hzocl1R2 or hzocl1F1l & hzocl1R2 from DNA obtained from Medway Bridge Marina sediment. Clear bands of the correct size were difficult to obtain from samples which had previously produced positive results for the presence of anammox bacteria via 16S rRNA screening (see chapter 3). This was surprising considering that *Ca.* K. stuttgartiensis has been reported to have up to 9 copies of *hzo* (Klotz, *et al.*, 2008; Kartal, *et al.*, 2011a), two of which are targeted by the "*hzo*

Cluster 1" PCR primers (Hirsch, *et al.*, 2011), compared with 1 copy of 16S rRNA (Strous, *et al.*, 2006). Thus, it can only be concluded that these primers have a low efficiency of amplification, possibly due to numerous degeneracies in the primer sequence (Table 2.4) or due to a greater susceptibility to sample contamination due to a decrease in the efficiency of DNA polymerase with these primers. Nevertheless, in general, better amplification was observed using primer set hzocl1F1l & hzocl1R2 (Figure 4.1) and so these primers were used during further investigations.

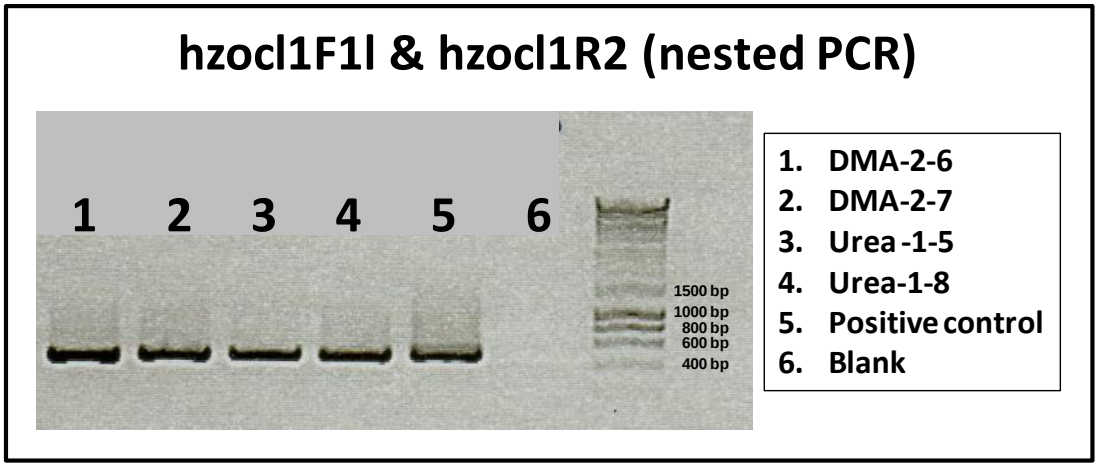


Figure 4.2: Agarose gel image of PCR product obtained from SIP fractions using hzo specific primers hzocl1F1l & hzocl1R2 as a nested PCR. Clear bands of the correct size were only obtainable after a nested PCR with a first round PCR using HZO4F & HZO1R. However, amplification efficiency of this nested PCR appeared to be far greater that the single round PCR (*c.f.* Figure **4.1**).

However, primer set hzocl1F1l & hzocl1R2 was unable to directly amplify *hzo* genes from SIP DNA fractions. Despite extensive optimisation, no visible PCR reaction product could be seen. However this is not surprising considering the low yields of DNA obtained after fractionation and precipitation. Therefore, a nested PCR approach was attempted using primers HZO4F & HZO1R before amplification with "*hzo* Cluster 1" specific primers hzocl1F1l & hzocl1R2. This methodology is similar to that outlined in Hirsch, *et al.* (2011) except that primer hzocl1F1l was used instead of hzocl1F1 based on the results obtained from the cloning of *hzo* genes from Medway Bridge Marina sediment (see above). This nested PCR provided a much greater amplification and higher yields efficiency than the one step PCR approach which facilitated cloning attempts (Figure 4.2).

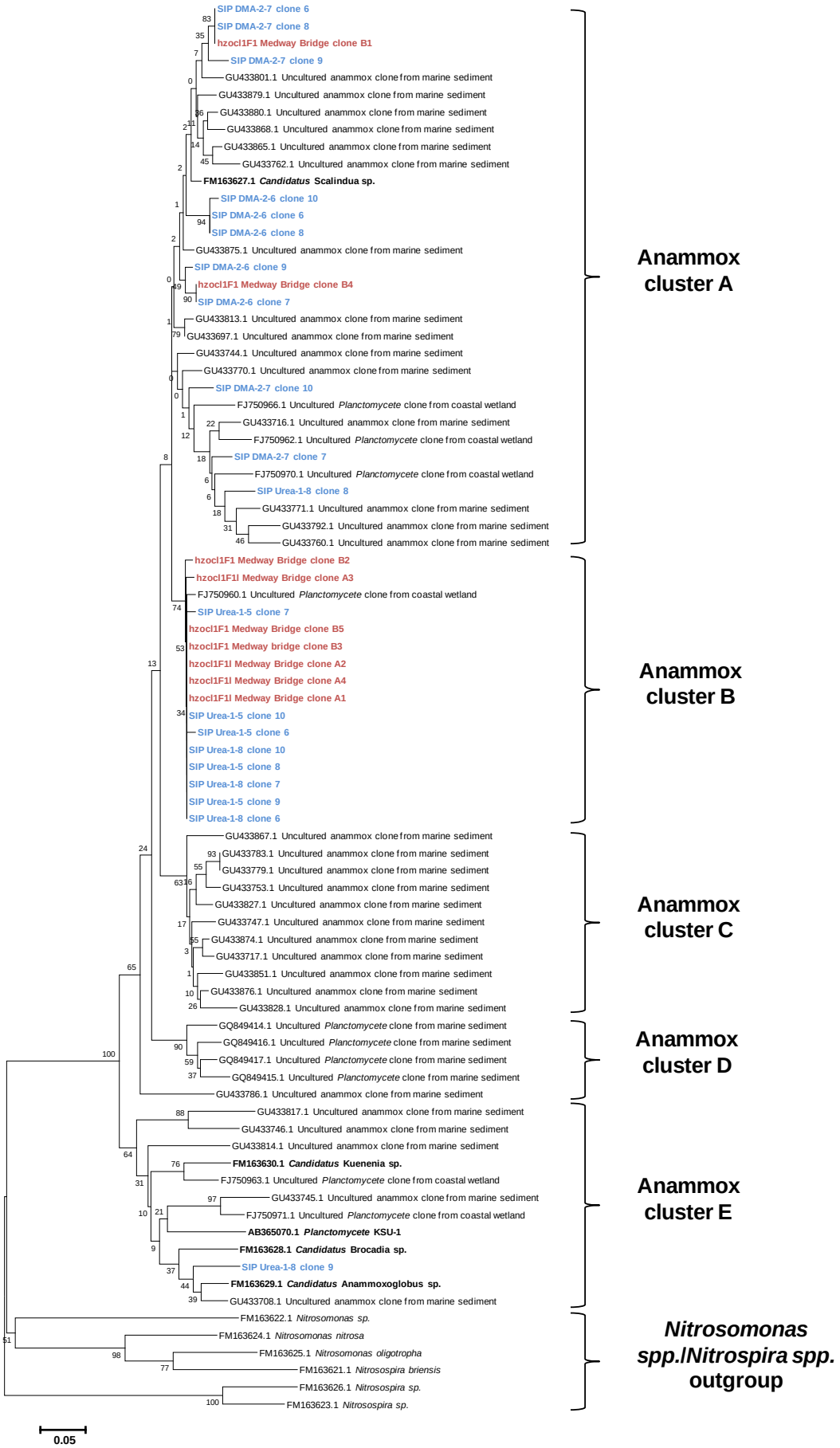

SIP DMA-2-7 clone 6
SIP DMA-2-7 clone 8
hzocl1F1 Medway Bridge clone B1
SIP DMA-2-7 clone 9
GU433801.1 Uncultured anammox clone from marine sediment
GU433879.1 Uncultured anammox clone from marine sediment
GU433880.1 Uncultured anammox clone from marine sediment
GU433868.1 Uncultured anammox clone from marine sediment
GU433865.1 Uncultured anammox clone from marine sediment
GU433762.1 Uncultured anammox clone from marine sediment
FM163627.1 Candidatus Scalindua sp.
SIP DMA-2-6 clone 10
SIP DMA-2-6 clone 6
SIP DMA-2-6 clone 8
GU433875.1 Uncultured anammox clone from marine sediment
SIP DMA-2-6 clone 9
hzocl1F1 Medway Bridge clone B4
SIP DMA-2-6 clone 7
GU433813.1 Uncultured anammox clone from marine sediment
GU433697.1 Uncultured anammox clone from marine sediment
GU433744.1 Uncultured anammox clone from marine sediment
GU433770.1 Uncultured anammox clone from marine sediment
SIP DMA-2-7 clone 10
FJ750966.1 Uncultured Planctomycete clone from coastal wetland
GU433716.1 Uncultured anammox clone from marine sediment
FJ750962.1 Uncultured Planctomycete clone from coastal wetland
SIP DMA-2-7 clone 7
FJ750970.1 Uncultured Planctomycete clone from coastal wetland
SIP Urea-1-8 clone 8
GU433771.1 Uncultured anammox clone from marine sediment
GU433792.1 Uncultured anammox clone from marine sediment
GU433760.1 Uncultured anammox clone from marine sediment
Anammox cluster A
hzocl1F1 Medway Bridge clone B2
hzocl1F1l Medway Bridge clone A3
FJ750960.1 Uncultured Planctomycete clone from coastal wetland
SIP Urea-1-5 clone 7
hzocl1F1 Medway Bridge clone B5
hzocl1F1 Medway bridge clone B3
hzocl1F1l Medway Bridge clone A2
hzocl1F1l Medway Bridge clone A4
hzocl1F1l Medway Bridge clone A1
SIP Urea-1-5 clone 10
SIP Urea-1-5 clone 6
SIP Urea-1-8 clone 10
SIP Urea-1-5 clone 8
SIP Urea-1-8 clone 7
SIP Urea-1-5 clone 9
SIP Urea-1-8 clone 6
Anammox cluster B
GU433867.1 Uncultured anammox clone from marine sediment
GU433783.1 Uncultured anammox clone from marine sediment
GU433779.1 Uncultured anammox clone from marine sediment
GU433753.1 Uncultured anammox clone from marine sediment
GU433827.1 Uncultured anammox clone from marine sediment
GU433747.1 Uncultured anammox clone from marine sediment
GU433874.1 Uncultured anammox clone from marine sediment
GU433717.1 Uncultured anammox clone from marine sediment
GU433851.1 Uncultured anammox clone from marine sediment
GU433876.1 Uncultured anammox clone from marine sediment
GU433828.1 Uncultured anammox clone from marine sediment
Anammox cluster C
GQ849414.1 Uncultured Planctomycete clone from marine sediment
GQ849416.1 Uncultured Planctomycete clone from marine sediment
GQ849417.1 Uncultured Planctomycete clone from marine sediment
GQ849415.1 Uncultured Planctomycete clone from marine sediment
Anammox cluster D
GU433786.1 Uncultured anammox clone from marine sediment
GU433817.1 Uncultured anammox clone from marine sediment
GU433746.1 Uncultured anammox clone from marine sediment
GU433814.1 Uncultured anammox clone from marine sediment
FM163630.1 Candidatus Kuenenia sp.
FJ750963.1 Uncultured Planctomycete clone from coastal wetland
GU433745.1 Uncultured anammox clone from marine sediment
FJ750971.1 Uncultured Planctomycete clone from coastal wetland
AB365070.1 Planctomycete KSU-1
FM163628.1 Candidatus Brocadia sp.
SIP Urea-1-8 clone 9
FM163629.1 Candidatus Anammoxoglobus sp.
GU433708.1 Uncultured anammox clone from marine sediment
Anammox cluster E
FM163622.1 Nitrosomonas sp.
FM163624.1 Nitrosomonas nitrosa
FM163625.1 Nitrosomonas oligotropha
FM163621.1 Nitrosospira briensis
FM163626.1 Nitrosospira sp.
FM163623.1 Nitrosospira sp.
Nitrosomonas spp./Nitrospira spp. outgroup
0.05

Figure 4.3: Phylogenetic tree containing *hzo* gene sequences Sequences obtained from Medway Bridge Marina environmental samples and from SIP fractions (see chapter 6 for context and full discussion). Medway Bridge sequences (red, n=9) were amplified using either PCR primers hzocl1F & hzocl1R2 or hzocl1F1l & hzocl1R2 (annotated in tree) whilst SIP sequences (blue, n=19) were amplified using a nested PCR using hzocl1F1l & hzocl1R2 followed by HZO4F & HZO1R. Reference sequences were obtained from GenBank from a mixture of clones from enrichments and environmental samples. Jiaozhou Bay (China) sequences were obtained from Dang, *et al.* (2010) and Mai Po Nature Reserve (Hong Kong) and South China Sea sequences from (Li, *et al.*, 2010b). Tree was constructed using the Neighbour-Joining method using p-distances in Mega 5.10 (Tamura, *et al.*, 2011). A bootstrap test (1000 replicates) was conducted on the phylogeny; branch annotations represent the percentage of replicate trees in which the associated taxa clustered together. Scale bar represents 5% sequence similarity. As can be seen from the tree, there is little consensus as to the phylogenetic diversity within the *hzo* gene. There appears to be a distinction between a potentially Ca. Scalindua spp. clade and other anammox organisms but this is based on unsupported branches and thus can be used with little confidence. The small amount of sequences from identified anammox organisms limits the usefulness of this data. As such it was concluded that the *hzo* gene was not a suitable marker for the phylogenetic diversity within anammox, at least with the current level of research. However, *hzo* gene sequences obtained using these PCR primers do appear to be distinct from *hao* and so may be suitable as a functional marker for anammox activity in an environment.

4.3.2. Phylogenetic Analysis of *hzo*

Phylogenetic analysis of *hzo* sequences from Medway Estuary M6 sediment (red highlights) is shown in Figure 4.3. The majority of clones obtained from Medway Estuary M6 sediments (7 out of 9) clustered within a distinct cluster, 'Cluster B' (bootstrap consensus=74%). From 16S rRNA analysis of these samples (Figure 3.8) it appeared that the only anammox organisms present in these samples were related to *Ca.* Scalindua spp. However these sequences clustered independently from the only known *Ca.* Scalindua sp. *hzo* sequence (FM163627.1), available at the time of constructing this tree, which was obtained from an enrichment culture. It must be noted that this reported lack of phylogenetic congruity can be used with little confidence as only one identified *Ca.* Scalindua sp. *hzo* sequence was found within the GenBank database. Indeed for all five characterised anammox genera, only one *hzo*, nucleotide sequence was available in GenBank (*i.e.* there were not even two *hzo* sequences which were identified as coming from the same organism or genus). Thus it is impossible to infer phylogenetic relationships with the data present due to the lack of positively identified sequences. Furthermore, the fact that all "uncultured" *hzo* sequences acquired from GenBank (which were used in calculating phylogeny to provide a greater statistical robustness to the analysis) were generated from only two studies (Dang, *et al.*, 2010; Li, *et al.*, 2010b) from three environments, all of which are from a marine setting, raises further questions as to the usefulness of this tree.

Marine environments (both coastal and open-ocean) are dominated by *Ca.* Scalindua spp. (Schmid, *et al.*, 2007) and therefore it is probably reasonable to accept that *hzo* sequences obtained from such environments would be heavily biased towards this genus.

4.4. Discussion

From the results of these investigations it appears as if the diversity of *hzo* does not match the phylogenetic diversity of the 16S rRNA gene. This contradicts the findings of Schmid, *et al.* (2008) who reported that the phylogeny of "*hzo* Cluster 1" was in good agreement with that of the 16S rRNA gene. Conversely, it has been argued that 16S rRNA does not accurately report phylogenetic diversity. Li, *et al.* (2010b) concluded that *hzo* primers were more specific and efficient than those of 16S rRNA targeting primers whereas Hirsch, *et al.* (2011) argued that the 16S rRNA gene underestimated anammox diversity based on a greater diversity obtained from *hzo* sequences which they concluded to be a closer representation of the true diversity. Wang and Gu (2013) further agreed with these statements. However, these claims are unreasonable as firstly, determination of the "true" microbial diversity, especially in complex environments, is impossible to attain due to the inability to accurately and comprehensively enumerate the number of individuals within the environment. Furthermore, it is also impossible to accurately determine bacterial "species" richness as the definition of what constitutes a "species" is neither clear nor universally implemented in bacteria and nevertheless, if a bacterial species could be accurately defined, each individual would have to be isolated from the community to positively determine which species it belonged to beyond doubt. As such measurements of the "true" diversity of a bacterial population cannot be obtained and so any investigations into bacterial ecology can only report a relative diversity. Secondly, with the present data available, it cannot be positively determined that diversity within the *hzo* gene represents phylogenetic diversity. Therefore it cannot be assumed that the higher measurement of diversity is more correct simply because it is the greater of the two as the assumption that it is overestimating diversity is equally as valid. In addition, regardless of whether the 16S rRNA gene is an accurate measurement of "true" anammox diversity, it remains the method by which the current interpretation of anammox taxonomy was established and as such, must be

used as the benchmark until anammox diversity is validly re-classified. Hirsch, *et al.* (2011) stated that "16S rRNA sequences provide taxonomic identification of bacteria but no evidence for functional capability. Thus the detection of *hzo* genes can be more promising to examine functional anammox bacteria in the environment". This statement is true providing that we can be confident of the role and identity of *hzo* and the quality of sequences obtained using primers to specifically target this gene. However, this may be confounded by multiple analogues of the *hzo* gene in the genomes of anammox bacteria, which may be functionally redundant.

It could be argued that 'Cluster A' and 'Cluster B' represent diversity within the genus Scalindua whereas the other three clusters represent that of other anammox organisms, however this would be in disagreement with the literature where *Ca.* Scalindua spp. *hzo*/HZO clearly cluster into one distinct clade within "*hzo* Cluster 1" (Schmid, *et al.*, 2008; Hirsch, *et al.*, 2011). In fact, these data do not support this claim as no sequences obtained from this investigation were positively identified as belonging to non-Scalindua anammox organisms. Only one sequence (Urea-1-8 clone 9) could be putatively assigned to an unknown, non-Scalindua anammox organism based solely on its association with 'Cluster E', which has been inferred to represent a sub-section of non-Scalindua anammox bacteria (Li, *et al.*, 2010b). As such, the ability of these primers to differentiate between *hzo* genes belonging to *Ca.* Scalindua spp. and non-Scalindua organisms cannot be asserted with confidence. Therefore, in the light of these data, it is difficult to confidently infer anammox diversity with the use of these primers.

Another issue with the use of *hzo* as a molecular marker for anammox is its similarity to *hao* genes found in aerobic ammonium oxidising bacteria (AOB). *hzo* and *hao* share a close functional and sequence similarity and it is difficult to definitively distinguish between HZO and HAO protein sequences (Klotz, *et al.*, 2008). In fact Klotz, *et al.* (2008) further speculated that HZO and HAO may be the same protein which oxidises different molecules depending on the upstream metabolic products of that particular organism (*i.e.* N_2H_4 in anammox and NH_2OH in AOB). However Shimamura, *et al.* (2007) managed to separate HZO and HAO like complexes from KSU-1 and demonstrated that the resulting HZO protein could

oxidise N_2H_4 but not NH_2OH thus suggesting a different function in at least some of the proteins translated from *hzo*/*hao* like genes.

4.5. Summary

- Three functional genes have been suggested in the literature as suitable for investigating the ecology of anammox organisms: *hzo*, *hzs* and *nirS.*
- *nirS* does not appear suitable as a phylogenetic marker due to its close relationship to *nir*K and discrepancies reported in the literature.
- *hzs* shows promise as a functional and phylogenetic marker for anammox organisms however requires further research to validate these claims.
- Despite contrary findings published in the literature, *hzo* does not appear to be a suitable molecular marker for measuring anammox diversity based on the data gathered during this investigation. However it may be suitable as a molecular marker for anammox functionality within an environment. Further research is required in order to elucidate the diversity of the *hzo* gene and its implications for anammox ecology.

5. Biogeography of Anammox Bacteria

5.1. Introduction

5.1.1. Overview

The anammox process is mediated by a unique suite of chemolithautotrophic bacteria (van Niftrik and Jetten, 2012). These bacteria belong to a monophyletic clade (Figure 1.3) which branch deeply within the phylum Planctomycetes (van Niftrik and Jetten, 2012). Recently, anammox have been designated as belonging to a new, distinct family, the *Brocadiaceae* (Song and Tobias, 2011), and order, the *Brocadiales* (Gori, *et al.*, 2011). Currently 5 genera of *Brocadiales* (Figure 1.3) have been identified from a range of environments; *Candidatus* Anammoxoglobus, *Candidatus* Brocadia, *Candidatus* Jettenia, *Candidatus* Kuenenia and *Candidatus* Scalindua (van Niftrik and Jetten, 2012). However, to date, anammox bacteria have not been successfully isolated or grown in pure culture and, as such, all anammox organisms have been designated as "*Candidatus*" organisms. The current phylogeny of anammox organisms is described by 16S rRNA sequence similarity. However, despite large phylogenetic differences both between anammox genera and other prokaryotes, the evolutionary history of these organisms is unclear (van Niftrik and Jetten, 2012).

The anammox reaction was first discovered in a wastewater treatment plant in the Netherlands (Mulder, *et al.*, 1995; van de Graaf, *et al.*, 1995; Van de Graaf, *et al.*, 1996) and it is from these environments that the majority of enriched anammox communities have been obtained. However anammox bacteria have since been found in a multitude of environments including marine sediments (Dalsgaard and Thamdrup, 2002; Engstrom, *et al.*, 2005; Amano, *et al.*, 2007; Rich, *et al.*, 2008; Engstrom, *et al.*, 2009), oxygen minimum zones in the World's oceans (Rysgaard, *et al.*, 2004; Hamersley, *et al.*, 2007; Woebken, *et al.*, 2008; Galan, *et al.*, 2009; Galan, *et al.*, 2012) and other anoxic marine waters (Dalsgaard, *et al.*, 2003; Kuypers, *et al.*, 2003; Kuypers, *et al.*, 2005; Woebken, *et al.*, 2007), particulate matter in oceanic anaerobic water columns (Hamersley, *et al.*, 2007; Jaeschke, *et al.*, 2007), estuarine sediments (Trimmer, *et al.*, 2003; Risgaard-Petersen, *et al.*, 2004; Dale, *et al.*, 2009; Nicholls and Trimmer, 2009), stratified lakes (Schubert, *et al.*, 2006; Hamersley, *et*

al., 2009; Yoshinaga, *et al.*, 2011), hot springs (Jaeschke, *et al.*, 2009a), petroleum reservoirs (Li, *et al.*, 2010a), marine sponges (Mohamed, *et al.*, 2010), terrestrial anaerobic soils (Humbert, *et al.*, 2010), wetlands (Penton, *et al.*, 2006; Humbert, *et al.*, 2012), hydrothermal vents (Byrne, *et al.*, 2008), groundwater (Clark, *et al.*, 2008; Moore, *et al.*, 2011), mangrove forests (Amano, *et al.*, 2011) and paddy fields (Sato, *et al.*, 2012). Indeed it has been suggested that anammox may be a ubiquitous process (Francis, *et al.*, 2007; Kartal, *et al.*, 2008) and may be found in any ecosystem with high amounts of dissolved inorganic nitrogen (DIN) and very low concentrations of O_2, *i.e.* less than 2 μM (Strous, *et al.*, 1997; Jetten, *et al.*, 2009).

5.1.2. Anammox and Global Nitrogen Cycling

Despite initial scepticism as to the importance of anammox in terms of the global nitrogen cycle (Zehr and Ward, 2002; Devol, 2003; Ward, *et al.*, 2009; Bulow, *et al.*, 2010), anammox has been shown to be a significant contributor to N_2 production (Francis, *et al.*, 2007; Russ, *et al.*, 2012). Anammox have been reported to account for as much as 50% of the total oceanic N_2 production (Devol, 2003) and have been reported to be responsible for 40% of N_2 production in the Black Sea (Kuypers, *et al.*, 2003), up to 35% in the Golfo Dulce, Costa Rica (Dalsgaard, *et al.*, 2003) and up to 67% in continental shelf sediments (Dalsgaard and Thamdrup, 2002). Theoretically, assuming that anammox and denitrification are the only routes to nitrogen loss from the anoxic environment, that anammox is a completely chemolithoautotrophic process which relies on denitrification for the production of NO_2^- and NH_4^+ (through the mineralisation of organic matter) and that the Redfield ratio (Redfield, 1934) holds true, anammox can only account for a maximum of 29% of N_2 production (Ward, *et al.*, 2007; Dalsgaard, *et al.*, 2012). Clearly this is not in agreement with data presented in the literature, presumably due to an over-simplistic view of the anammox process and the metabolic versatility of the organisms responsible for it. Furthermore, if these assumptions were true, anammox would be dependent on the presence and activity of denitrifying organisms which has not always been shown to be the case (Thamdrup, *et al.*, 2006; Galan, *et al.*, 2009). Dalsgaard, *et al.* (2012) argued that this statement does hold true in OMZs and that pockets of high denitrification were spatially separated from anammox which contributed to an overall ratio of anammox to denitrification of ~29:71. They also argued that the

small number of such regions expressing high denitrification rates meant that they had not been represented by the poor sampling resolution of previous investigations (see above), thereby skewing the data. However, if this were true then NH_4^+ would be seen to accumulate in the presence of denitrification which is not shown by the current data (e.g. Bulow, *et al.*, 2010). Furthermore, if anammox bacteria were solely dependent on denitrifiers for NO_2^- it would be unreasonable to presume that they would be spatially separated.

However, the relative importance of anammox to N_2 production has been reported to be much lower in non-pelagic environments. In coastal regions anammox have been found to be responsible for less than 10% of N_2 production (Thamdrup and Dalsgaard, 2002; Kuypers, *et al.*, 2006). Anammox rates in estuarine environments have also been shown to be low, with estimates of N_2 production by anammox bacteria ranging from <1-8% (Trimmer, *et al.*, 2003), 0-24% (Risgaard-Petersen, *et al.*, 2004) and 3.8-16.5% N_2 (Dale, *et al.*, 2009). Trimmer, *et al.* (2003) also showed that anammox rates were higher towards the marine end of the estuary and were much lower upstream. However, the large numbers of estuarine environments on Earth still make anammox an important component of the global N-cycle. Anammox rates from stratified lakes were also shown to be less than those of pelagic environments with ~13% of N_2 production attributed to anammox in Lake Tanganyika (Schubert, *et al.*, 2006) although Yoshinaga, *et al.* (2011) did report maximum values of N_2 production via anammox of 40% though this was only at 2 of 11 sites.

5.1.3. Diversity

Ca. Brocadia anammoxidans was the first anammox bacterium to be identified and named (Strous, *et al.*, 1999) and was enriched from the wastewater plant mentioned above. Other anammox bacteria were subsequently discovered in similar wastewater systems belonging to the other four anammox genera (Schmid, *et al.*, 2000; Schmid, *et al.*, 2003; Kartal, *et al.*, 2007b; Quan, *et al.*, 2008). Anammox bacteria were first discovered in the natural environment, namely the Black Sea, by Kuypers, *et al.* (2003) and identified as belonging to the Scalindua genus (*Ca.* S. sorokinii). Since then many investigations have been conducted on anammox bacteria in the natural environment (see above). Observations made from such studies appear to suggest

that anammox diversity can be divided into two distinct groups: *Ca.* Scalindua spp. which are specific to marine environments and non-Scalindua anammox organisms found in freshwater and brackish environments (Hirsch, *et al.*, 2011). However, most investigations into the diversity of anammox organisms have been limited due to the low resolution of acquired data and paucity of the number of sequences obtained (largely from clone libraries) which restricts the strength of subsequent conclusions (Humbert, *et al.*, 2010). Furthermore the methods for targeting specific anammox genes have been used with varying success (see chapter 3 and 4).

Despite some level of understanding in the diversity of anammox organisms in terrestrial and marine environments, the causes of this biodiversity and its environmental controls are relatively unknown (Amano, *et al.*, 2007; Hamersley, *et al.*, 2009). Kartal, *et al.* (2007b) state that each anammox species occupies a "well-defined (but yet unknown) niche" and that "different anammox species are rarely found in the same ecosystem". Dale, *et al.* (2009) report that different levels of anammox diversity and distribution are observed in different environments. However, few studies from the literature utilise the same methodologies in comparing anammox organisms in different environments and so conclusions drawn from such research must be used with trepidation. Due to the ecological importance and environmental significance of these organisms, the acquisition of a clear and comprehensive understanding of the diversity and distribution of anammox bacteria is imperative (Kuypers, *et al.*, 2006).

There has been suggestion that the diversity of anammox organisms is controlled by salinity, at least in estuarine environments (Dale, *et al.*, 2009; Hirsch, *et al.*, 2011), however experiments conducted on enriched, freshwater anammox organisms (Kartal, *et al.*, 2006) demonstrated that they were not affected by changes in salinity. It has been suggested that anammox bacteria are metabolically diverse (Strous, *et al.*, 2006; Kartal, *et al.*, 2007a; Jetten, *et al.*, 2009), though the extent of such metabolic diversity is poorly understood, which may account for at least some of the diversity and distribution of these organisms (see chapter 6 for full discussion). Anammox may have the ability to utilise organic substrates as sources of carbon and nitrogen or as an energy source (Kuypers, *et al.*, 2006; Trimmer and Purdy, 2012) and, if this were true, it would be feasible that the bioavailability of organic matter may affect

anammox diversity (the potential for organic pathways to the anammox reaction is covered in chapter 6).

5.1.4. Anammox in Estuaries and OMZs

Estuaries and oceanic OMZs have been shown to be important environments in global nitrogen cycling associated with the anammox process (see above) however we know little about the anammox organisms present in these environments. These two environments are doubly interesting as they allow for the investigation of anammox diversity over two contrasting environmental gradients; a salinity gradient, along an estuary and an O_2 gradient, in an OMZ.

OMZs, sometimes called anoxic marine zones (AMZs) (Ulloa, *et al.*, 2012) or oxygen deficient zones (ODZs) (Chang, *et al.*, 2012), have been shown to be important regions for nutrient cycling despite the fact that they are small in size (in relation to the entire ocean) and number (Paulmier and Ruiz-Pino, 2009). The most significant OMZs, in terms of size and N_2/N_2O loss, are the Eastern Tropical North Pacific (ETNP), Eastern Tropical South Pacific (ETSP), Arabian Sea, Bay of Bengal and the Benguela Upwelling (Woebken, *et al.*, 2008; Paulmier and Ruiz-Pino, 2009). OMZs occur in regions of massive upwelling of nutrients either from nutrient rich oceanic currents coming into contact with land masses (Kuypers, *et al.*, 2005), wind-driven upwelling (Karstensen, *et al.*, 2008) or the Ekman Effect (Bauer, *et al.*, 1991). This upwelling over-stimulates aerobic primary production (eutrophication) in surface waters which in turn draws up dissolved O_2 creating zones of low oxygen concentrations below (Woebken, *et al.*, 2008). OMZs typically demonstrate a steep oxycline (a decreasing O_2 gradient from the surface to the top of the OMZ) followed by a large anoxic core and a shallower oxycline at the bottom of the OMZ as dissolved O_2 content recovers (Paulmier and Ruiz-Pino, 2009; Rush, *et al.*, 2012; Ulloa, *et al.*, 2012). Such O_2 gradients create unique redox gradients where a diverse range of electron acceptors and donors are utilised by microorganisms (such as anammox bacteria) for metabolism (Stewart, 2011).

Estuaries have been shown to be important environments for the removal of nitrogen through N_2 and N_2O (Ogilvie, *et al.*, 1997; Trimmer, *et al.*, 2003; Nicholls and Trimmer, 2009) and are a major diffuse and point sources for anthropogenically

sourced nitrogen entering the water cycle (Dong, *et al.*, 2009). Estuarine environments demonstrate strong gradients of salinity and nutrients (Nogales, *et al.*, 2002). Such pronounced changes in salinity and salt-water intrusions give estuarine environments unique chemical and biological characteristics and thus are often ecologically diverse and environmentally important (Manahan, 2005).

This study aims to elucidate the diversity and distribution of anammox bacteria in estuarine and OMZ environments and to begin to investigate the environmental factors influencing the biogeography of these organisms as per the hypotheses outlined in 1.3.2.1.2. High-throughput sequencing techniques (454 pyrosequencing) were used in order to gain an extensive and comprehensive representation of these organisms which would be unprecedented in investigations of anammox ecology.

5.1.5. Measuring Diversity Using Next Generation Sequencing

The use of 454 pyrosequencing technologies (or indeed any high-throughput sequencing technology) presents unique challenges for metagenomic or amplicon sequencing applications (also known as massively parallel sequencing) compared with whole genome sequencing. In the latter, consensus data is obtained by aligning the data with either reference genomes or to itself, and thus the effect of erroneous bases or reads is ameliorated with increased coverage (Huse, *et al.*, 2007; Kunin, *et al.*, 2010). For example, if the genome coverage after sequencing was 20x, then the likelihood of a single base being miscalled 20 times is small and thus, the potential of a small number of errors, randomly distributed across the data, to effect the overall conclusions or outcome of the data is also small. However, with metagenomic or amplicon sequencing to investigate microbial diversity, these errors are more important as each individual read can represent a unique gene or taxa. Thus it is challenging to determine what constitutes an erroneous base or read and what constitutes true sequence diversity (Huse, *et al.*, 2007). Such a problem would be particularly debilitating when searching for single nucleotide polymorphisms (SNPs) or when investigating organisms that are phylogenetically closely related.

As such, the pipeline used in this study (chapter 2) is based on a number of assumptions as to what would qualify an erroneous read (or a read with significant errors in it) which are described in section 5.3 below. This method was chosen,

compared with other more widely used methods, *e.g.* QIIME (Caporaso, *et al.*, 2010) or Pyronoise (Quince, *et al.*, 2009), as the limits and effects of each step in this pipeline were better understood and it provided the ability to modify the various scripts in order to meet the specific requirements of this study. Furthermore, this pipeline produced good quality data for downstream analysis (see sections 5.4.2.1 and 5.4.3.2). Attempts were made in this investigation to validate some of these assumptions to ensure that good quality reads were not being removed with those of poorer quality. In addition a novel and innovative approach to the use of read lengths to determine good quality reads, Read Length Incremental Clustering (ReLIC), was developed. The rationale behind this method and the other assumptions used in this pipeline are discussed in section 5.3.

5.2. Methodology

5.2.1. Sampling Site Locations

Environmental samples were taken from 2 locations; the Medway Estuary (Kent, UK) and the ETNP OMZ. For a full description of sampling locations and strategies see chapter 2. All environmental samples were collected in triplicate.

5.2.2. DNA Extraction

DNA was extracted from ~0.5 g of wet sediment collected from the Medway Estuary using the protocol described in chapter 2 (Purdy, 2005). DNA was PEG precipitated as described in chapter 2 in order to remove impurities (Selenska and Klingmuller, 1991) and stored at -20°C in 50 μl of 10 mM Tris pH 7.5.

DNA was extracted from filtered samples collected from the ETNP OMZ as outlined in chapter 2. The filter was cut in half using a sterile scalpel blade and forceps and inserted into a 2 ml screw cap tube with ~0.5 g of sterile glass beads. DNA from filters was extracted as described in Purdy (2005) for other types of environmental samples. During this protocol the polycarbonate filter was observed to completely disintegrate/dissolve allowing for effective lysis of bacterial cells and negating the need to scrape off cells from the filter as discussed in other methods for the extraction of nucleic acids from aquatic environments. Samples were not PEG precipitated as only low quantities of DNA were obtained from these samples and

the DNA appeared to be of a suitable purity for downstream analysis and therefore this step was not necessary. DNA was re-suspended in 40 µl* of 10 mM Tris pH 7.5 and stored at -20°C.

5.2.3. PCR

PCR was performed on extracted DNA as outlined in chapter 2. DNA was first amplified with primers Pla46F & 1390R and product purified via gel extraction (Gel Extraction Kit, Qiagen, UK). A second round PCR using primers Amx368F & Amx820R was then utilised to specifically target anammox 16S rRNA genes.

5.2.4. 454 Pyrosequencing

Selected samples were submitted for pyrosequencing as outlined in chapter 2. PCR product obtained from Medway Estuary DNA was sequenced at the Research and Testing Laboratory (Lubbock, Texas, USA) and PCR product from ETNP OMZ DNA was sequenced by AHVLA (Surrey, UK). All PCR products were submitted in triplicate. Raw pyrosequencing data was analysed using the pipeline outlined in chapter 2. However, ETNP data were analysed both including and omitting OTUs which represented <0.1% of the total number of reads in order to assess the validity of the assumption that such OTUs are likely to represent poorer quality sequences or true rare taxa which would not affect investigations of relative anammox diversity and subsequent conclusions.

5.2.5. Environmental Data

Environmental data were obtained for both the Medway Estuary and ETNP OMZ in order to compare these data with anammox diversity data and investigate the potential factors affecting anammox diversity and distribution in these environments.

Medway Estuary data were obtained from the Environment Agency, UK. Data were available for the twelve years prior to this investigation from which yearly and monthly averages (with standard error measurements) were calculated. Twelve year averages for the month sampled (*i.e.* March) were interpolated and extrapolated to

* A lesser volume of 10 mM Tris was used to increase the final concentration of DNA in order to aid downstream analysis.

obtain measurements for the exact sites sampled (Figure 2.2) using Matlab v8.1.0.604 (R2013a), The MathWorks Inc. (2013).

ETNP OMZ data was obtained during NERC research cruise D373 (December 2011-January 2012). O_2 data was acquired using a CTD attached to the Niskin rosette which provided continuous data with every drop. Nutrient data was collected using a segmented flow auto analyser (Skalar, The Netherlands) from 60 ml of seawater collected anaerobically. These data were obtained by the research group of Professor Mark Trimmer, Queen Mary's University, London in conjunction with this study. Nitrogen deficits (N_{def}) were calculated as outlined in Gruber and Sarmiento (1997) from these nutrient data.

Environmental data were compared with 454 pyrosequencing data using Mantel tests and CCA analyses as outlined in chapter 2. Mantel tests were used to plot the effects of environmental variables on CCA plots.

5.3. 454 Pyrosequencing, Quality Control and Rationale

5.3.1. Overview

A number of assumptions (Huse, *et al.*, 2007; Kunin, *et al.*, 2010), as to what qualifies as erroneous or poor quality read, were used to remove such sequences from 454 pyrosequencing data.

5.3.2. Presence of Primer Sequences

Initially, the raw data were screened for reads which did not contain the forward or reverse primer sequences (Table 2.1). If a particular read did not contain the forward primer then it can be assumed that the sequence represents DNA from a source other than the pool of anammox 16S rRNA amplicons which were being investigated or that there had been significant errors during PCR or sequencing. Similarly, if the reverse primer sequence was not present, then this may be indicative that the read has been prematurely truncated or fragmented during PCR or sequencing or may represent a chimeric sequence (Brakenhoff, *et al.*, 1991; Kopczynski, *et al.*, 1994; Huber, *et al.*, 2004).

In the data collected during this investigation, the quality of the majority of reads (determined via investigation of Q scores and visual comparison) was greater to the start of the read (5' end) than the end (3' end). This contradicts the findings of Huse, *et al.* (2007), who did not observe any discernible or significant reduction in the quality along reads, however, the smaller amplicon length of their study (~200 bp), compared with the data presented in this investigation (~450 bp), may have resulted in a greater quality throughout the read. As such, a higher number of reads contained the forward primer sequence than the reverse primer sequence. It was possible to only screen for the forward primer, ignoring either the presence or absence of the reverse primer, to increase the number of reads kept after screening. This is not desirable as a higher proportion of sequences screened with only the forward primer, compared with both primers, contained significant errors, as discovered during downstream analysis. However, this may be the only course of action when dealing with large amplicons which are close to the current limits of such sequencing technologies. Fortunately, in this investigation, the depth of coverage was sufficient to be able to be stringent at this initial stage of quality control and still report the entire diversity observed in these samples (see Figure 5.3 and Figure 5.15). Therefore, reads which did not contain both the forward and reverse primer sequences were omitted from further analysis. Specific results as to the number of reads in each investigation which did not contain the primer sequence are presented in the relevant sections below.

5.3.3. Ambiguous Bases

The presence of ambiguous bases (Ns) have also been shown to be indicative of poor quality sequence data (Huse, *et al.*, 2007; Kunin, *et al.*, 2010). The data produced during this investigation typically had very low numbers of ambiguously called bases, with typically only between 0 and 1% of the reads having more than one 'N' present (see section 5.4). Nevertheless, any read which did contain ambiguous bases was omitted from further analysis as this could affect phylogenetic analyses.

5.3.4. Length Statistics

The length of individual reads has also been used as an indicator of poor quality sequences (Kunin, *et al.*, 2010). Each amplicon will have a particular length and so

deviation away from this length would suggest significant insertions or deletions (indels), perhaps indicative of poor quality sequence. True indels, *i.e.* arising from true sequence variation, are common in nature (Britten, *et al.*, 2003) though are more common in functional, protein coding genes than highly conserved ribosomal genes within closely related organisms. Therefore, a large deviation from the expected amplicon size when investigating 16S rRNA genes is likely to be due to considerable errors.

However, what constitutes "considerable deviation" is unclear and there has been little investigation in the literature to define this. Kunin, *et al.* (2010) were highly specific in how they trimmed read lengths, selecting read lengths which only exactly matched the expected amplicon size. Such an approach may be suitable when investigating pyrosequencing data from pure cultures, as in Kunin, *et al.* (2010), but is unlikely to be so when investigating complex microbial communities where different taxa will have different amplicon lengths. Therefore a range of read lengths must be selected which will remove reads of erroneous lengths but not those representing true, good quality sequences. However the literature contains no method for the designation of suitable length cut-offs and so read lengths are commonly chosen arbitrarily. Thus, during this investigation, a novel method, ReLIC, was developed in order to attempt to designate such cut-offs statistically.

Read length distributions were calculated for each dataset using the above data analysis pipeline. Distributions calculated from these data typically demonstrated an overall bell-shaped distribution which was negatively skewed. This negative skew is likely to represent the fact that read length was constrained by a minimum value (*i.e.* 0) and maximum value constrained by the current limitations of 454 pyrosequencing technologies (maximum read length is around 500-600 nucleotides). The mean read length (~460-470 nucleotides typically observed with these data) lies closer to the upper than the lower bound, resulting in a negatively skewed distribution. That is to say, if the data were not constrained by lower and upper bounds, it was expected that the data would be more evenly distributed about the mean with little or no skew. Frequently, distributions also demonstrated more than one peak, with a main peak approximately centred about the mean value. It was assumed that individual peaks represented different taxonomic groups and the distribution about the individual

apexes, however confirmation of this hypothesis, would require further investigation using datasets with known amounts of different taxa with varying lengths.

The minimum and maximum read lengths which represented "real" (*i.e.* not erroneous) sequences were investigated during this study. Sequences which had passed the quality control criteria discussed in sections 5.3.2 and 5.3.3 were trimmed incrementally down from the median value (in these datasets the mean and median were almost equal except for that the median represented a discrete value). The data were also trimmed independently up from the median value, creating two series of sequence files in FASTA format (*i.e.* one series representing the minimum read length up to the median length and the other encompassing read lengths from the median to the maximum observed read length). Each increment was then clustered independently as discussed in section 2.3.5.2.2. The increase of OTUs and reads between adjacent lengths was calculated and investigated. The ratio of the increase in OTUs compared with the increase in reads (R_{inc}) for each length was calculated using the equation:

$$R_{inc} = \frac{n_{OTU}}{n_{read}}$$

R_{inc} provided a numerical representation of the relative increase in the number of OTUs compared with reads. A low value (approaching 0) would indicate that new reads from that increment either clustered into existing clusters or that a large number of reads clustered into a new OTU, representing a new taxon, observed at that specific length. Values approaching 1 would indicate that the inclusion of a new read produced a new OTU, suggesting that this read contained significant errors as the sequence failed to align to other sequences within the dataset. Plots of these data (see Figure 5.2 and Figure 5.12) indicated that R_{inc} remained low either side of the median length, but appeared to reach a critical length value, increasing rapidly to 1. This critical value was deemed as a suitable cut-off where lengths below (or above for lengths above the median) this value were likely to represent erroneous sequence data. As such, reads outside of these critical length values were omitted from further analyses.

This is unlikely to be the optimum method for removing erroneous read lengths from such data and has the potential to still include poor quality data. Due to poor data

resolution (*i.e.* not every potential read length was represented in the data and often a specific length had no data available) the determination of an exact critical length value was impossible. Furthermore, a clear shift in R_{inc} from ~0 to 1 was not always observed, especially for lengths greater than the median (see Figure 5.2 and Figure 5.12). Negative values were also encountered, where an increase in the number of reads reduced the number of OTUs (due to the clustering algorithm used, see below) which also made the determination of a cut-off length value difficult. As such, poor quality reads, potentially producing misleading or erroneous results may still be included in these data. Regardless of the potential inclusion of poorer quality reads, RLIC remains to be a preferable method than choosing minimum and maximum read length values arbitrarily. Investigations into more accurate methods of defining maximum and minimum length cut-off values to develop the approach highlighted above, including the modelling of length distributions to determine critical values, are required; however time constraints prevented their inclusion in this thesis.

5.3.5. Defining OTUs and Clustering

454 pyrosequencing reads which had passed the above quality controls were clustered into OTUs as outlined in section 2.3.5.2.2. OTUs were defined at a similarity cut-off of 95%. In the literature 97% is often used as an indicator of species level diversity in 16S rRNA genes (Quince, *et al.*, 2008), however, this value is often misquoted from Stackebrandt and Goebel (1994). Regardless, a universal cut-off of 97% does not represent the diversity of bacterial genetic variation even if the concept of a bacterial species was well constrained. Furthermore, anammox organisms do not conform to such a strict diversity cut-off as the observed sequence variation is not consistent between different groups of anammox organisms (Jetten, *et al.*, 2009). Therefore, a cut-off distinguishing between *Ca.* Brocadia spp. may not necessarily be transferrable when investigating diversity within the genus *Ca.* Scalindua and therefore is not universally applicable to the anammox clade. Furthermore, to my knowledge, the literature does not contain any similar study on anammox diversity using next generation sequencing technologies. Thus, no similarity cut-off value was satisfactorily defined in the literature to determine individual OTUs from anammox 16S rRNA sequence data. As such, a cut-off value of 95% was chosen for this study as greater, similarity cut-off values tended to give

large numbers of OTUs which proved difficult to analyse whereas lower cut-off values did not appear to accurately portray anammox inter-genera diversity. A cut-off of 95% was also chosen as it consistently appeared to represent the entire anammox diversity of the sample through rarefaction analyses (Figure 3.3, Figure 3.7, Figure 5.3 and Figure 5.15).

Failure to absolutely and consistently define a "species" (or even higher taxonomic groups) in these data was not seen as an issue for analysis. This would only be an issue when defining absolute diversity, which is not possible in microbial ecology due to the vast numbers of organisms involved (each member of a community would have to be individual characterised and measured). As such it is only possible to measure relative diversity, for which, as long as the criterion which are used to define OTUs remain consistent throughout analysis, the specific cut-off used is irrelevant (providing that the entire diversity of the sample is represented).

Low abundance OTUs (*i.e.* those containing small numbers of reads) were generally omitted from analysis[*] of the diversity of anammox organisms as these were deemed to have a high potential for containing poor quality reads. Potentially, some of these low abundance clusters may have represented true, rare taxa which may be ecologically significant (Sogin, *et al.*, 2006; Huse, *et al.*, 2010; Kunin, *et al.*, 2010). However, the aims of this investigation (section 1.3) were to investigate the diversity and distribution of anammox organisms across environmental gradients. To achieve these aims the relative diversity of samples was measured as a representation of the true microbial diversity and so, the omission of rare taxa is not an issue as long as this was universally implemented across samples. The investigation of rare taxa is in itself a major problem for microbial ecology but was not necessary to achieve the aims of this study.

[*] These clusters were included in the analysis of ETNP data in order to test the validity of this assumption (see section 5.4.3.3).

5.4. Results

5.4.1. 454 Pyrosequencing Errors and Reproducibility

5.4.1.1. Positive Controls

DNA from a cloned 16S rRNA gene fragment, positively identified as belonging to a *Ca.* Brocadia sp., and technical replicates from ETNP sample S2-100m 2 were sequenced in triplicate using 454 pyrosequencing in order to ascertain the error associated with these sequencing runs. It must be noted however that the error rates in 454 pyrosequencing are likely to be variable and associated with individual reads and genes.

After quality control a total of 12,306 reads remained out of 14,771, with 5,041 reads from replicate 1, 4,077 reads from replicate 2 and 3,189 reads from replicate 3. Reads were clustered at similarity cut-off values of 97%, 95% and 93%. All three cut-off values produced identical results with all but one read (*i.e.* a singleton cluster) being assigned to the same OTU. BLAST analysis of representative sequences from this cluster showed that this cluster showed a 95-96% sequence similarity with other *Ca.* Brocadia spp. 16S rRNA gene sequences. As such, the 454 pyrosequencing technologies used in this project appeared to demonstrate a high degree of accuracy.

5.4.1.2. Technical Replication

Technical replicates to analyse the precision of 454 pyrosequencing were acquired by sequencing a single PCR product from an environmental sample (S2-100m 2) in triplicate. Sequencing produced 9,221 reads of which 8,081 survived quality control (replicate 1=2,877, replicate 2=2,872 and replicate 3=2,332 reads). Reads were clustered at a similarity cut-off value of 95% resulting in 8 OTUs. Mean read abundances (expressed as percentages) were calculated for each OTU with standard errors. Mean read abundance data can be seen in Figure 5.1. These data showed a high degree of precision (SE <0.23%) associated with the 454 pyrosequencing technologies used in this project with low standard errors associated with each OTU indicating a low degree of variance about the mean value.

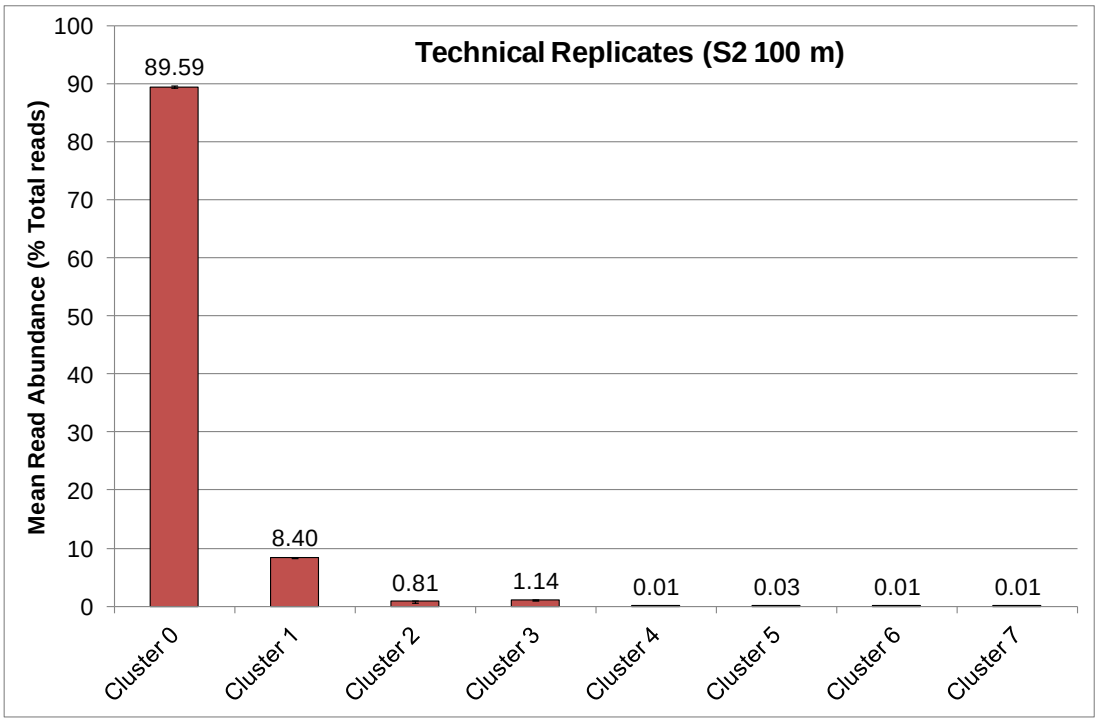


Figure 5.1: Histogram showing the mean percentage of reads from 454 pyrosequencing technical replicates. Technical replicates were obtained from sample S2-100 m. Clusters were defined at a similarity cut-off of 95%. Error bars represent ± 1 S.E.

5.4.2. Medway Diversity

5.4.2.1. Quality Control

A total of 179,439 reads were obtained from Medway Estuary 454 pyrosequencing data for the seven sites M1-M7 (including biological triplicates). Of these 84,379 contained both the forward and reverse anammox specific 16S rRNA primer sequences (*i.e.* Amx368F & Amx820R) with which PCR had been conducted.

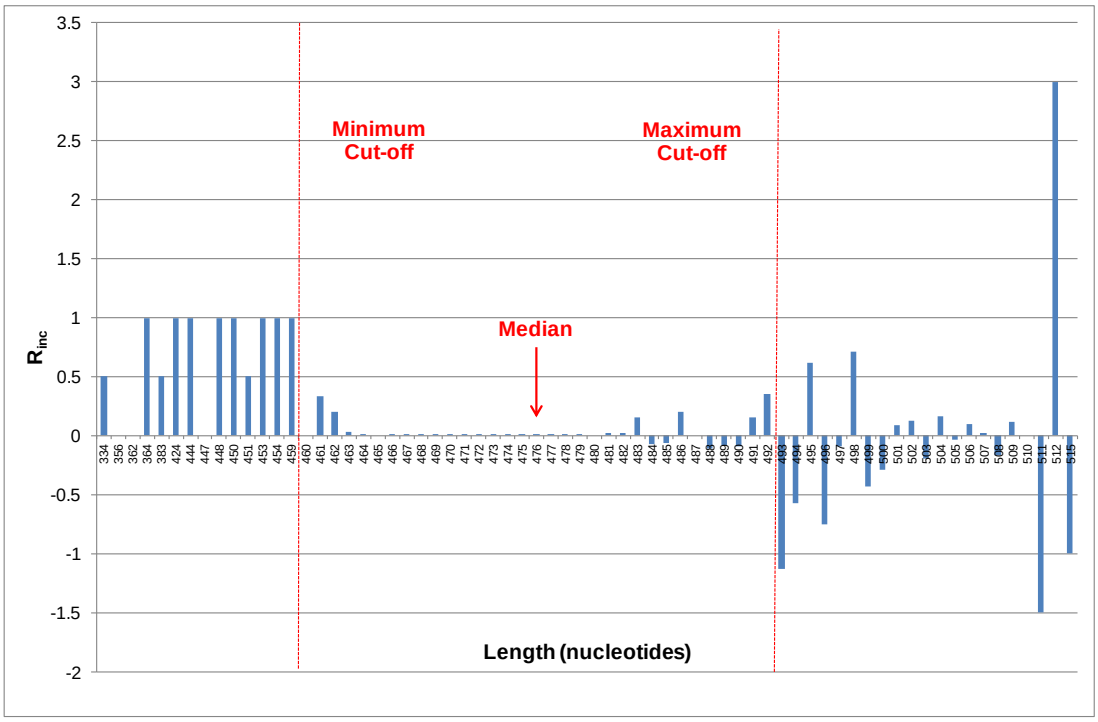


Figure 5.2: R_{inc} values for all 454 pyrosequencing data from Medway estuary sites 1-7. Only read lengths which were represented in the data are shown. Red hashed lines indicate chosen minimum and maximum cut-off values from which reads between these two values were used for downstream analysis.

The length distribution of these reads ranged from 334 to 515 reads with the median value at 476 bases. No ambiguous bases (Ns) were observed in any of the 84 k reads. ReLIC analysis and calculation of R_{inc} was carried out as described in section 5.3, the results of which can be seen in Figure 5.2. This analysis typically demonstrated a clear division between low (~0) and high (~1) R_{inc} values towards lower read lengths but this division was not as clear or definitive towards higher read lengths (see section 5.3 for full discussion). However, an approximate upper-bound cut-off value was estimated. As such reads were trimmed between lengths of 460 bases and 492 bases and only reads between these two cut-offs were included in further analysis. This produced 83,999 reads for downstream processing.

5.4.2.2. Clustering and Phylogenetic Analysis

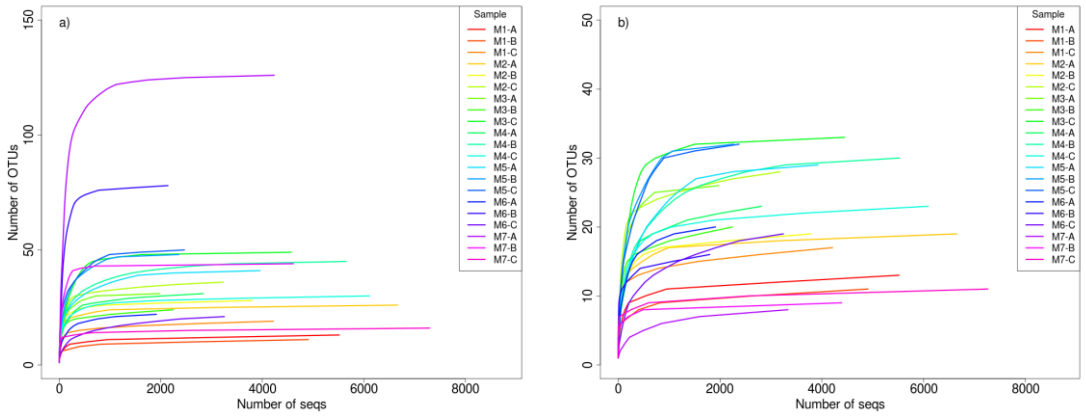


Figure 5.3: Rarefaction curves from Medway Estuary 454 data. Rarefaction curves are shown for a) all OTUs (n=294 OTUs) and b) only OTUs which represented >0.1% of the total number of reads (n=49 OTUs). Including all OTUs, most samples contained <50 OTUs except sample M6-B (n=78) and M7-A (n=126), potentially highlighting a large number of poor quality reads and OTUs in these samples. Rarefaction curves for each sample were shown to plateau, indicating that the sampling depth was sufficient to report the entire anammox diversity in these samples, as expressed by the specificity of these primers. Rarefaction curves drawn only from clusters which represented greater than 0.1% of the total number of reads (b) were also shown to plateau indicating that by omitting these clusters, relative measurements of anammox diversity were unlikely to be affected and conclusions drawn from both datasets would be comparable.

Reads which had passed quality control (n=83,999) were clustered as outlined in section 5.2.4 at a similarity cut-off of 9%. This produced 294 OTUs of which 245 represented less than 0.1% of the total number of reads and 48 of these were singleton clusters. These reads/clusters were omitted resulting in a total of 49 OTUs for further analysis representing 97.47% of the total number of reads (83,999) which had passed quality control. Rarefaction analysis of these data (Figure 5.3) indicates that in all samples, rarefaction curves were plateauing indicating that the entire anammox diversity, as represented by the specific primers used (*i.e.* Amx368F & Amx820R) had been reported and as such, the sequencing depth used had been suitable. Rarefaction curves were also plateauing when performing rarefaction analysis on only OTUs which represented greater than 0.1% of the total number of reads (Figure 5.3b), indicating that by omitting these OTUs, subsequent measurements of relative anammox diversity would not be significantly affected.

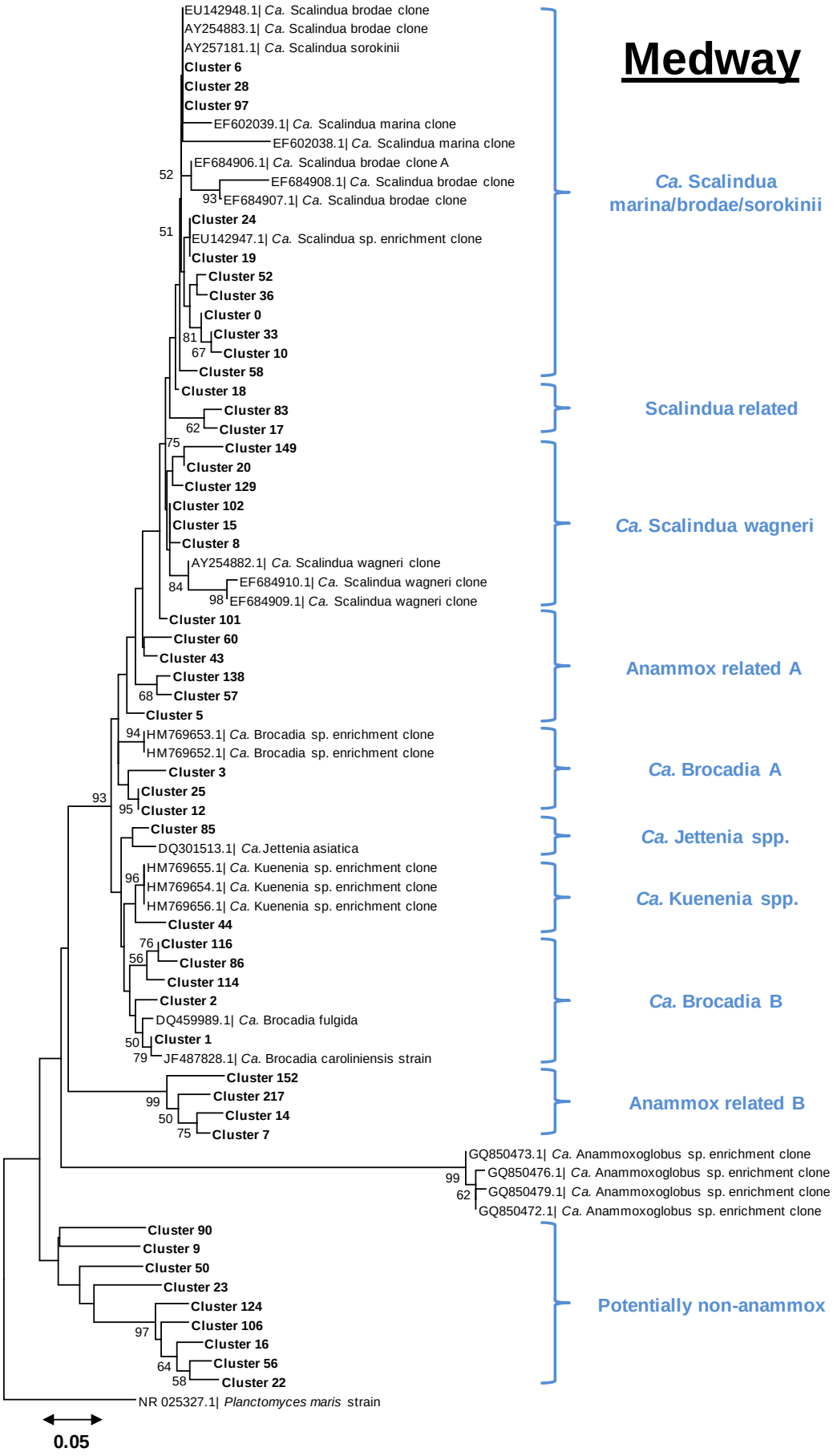

Medway
EU142948.1| Ca. Scalindua brodae clone
AY254883.1| Ca. Scalindua brodae clone
AY257181.1| Ca. Scalindua sorokinii
Cluster 6
Cluster 28
Cluster 97
EF602039.1| Ca. Scalindua marina clone
EF602038.1| Ca. Scalindua marina clone
EF684906.1| Ca. Scalindua brodae clone A
EF684908.1| Ca. Scalindua brodae clone
EF684907.1| Ca. Scalindua brodae clone
Cluster 24
EU142947.1| Ca. Scalindua sp. enrichment clone
Cluster 19
Cluster 52
Cluster 36
Cluster 0
Cluster 33
Cluster 10
Cluster 58
Cluster 18
Cluster 83
Cluster 17
Cluster 149
Cluster 20
Cluster 129
Cluster 102
Cluster 15
Cluster 8
AY254882.1| Ca. Scalindua wagneri clone
EF684910.1| Ca. Scalindua wagneri clone
EF684909.1| Ca. Scalindua wagneri clone
Cluster 101
Cluster 60
Cluster 43
Cluster 138
Cluster 57
Cluster 5
HM769653.1| Ca. Brocadia sp. enrichment clone
HM769652.1| Ca. Brocadia sp. enrichment clone
Cluster 3
Cluster 25
Cluster 12
Cluster 85
DQ301513.1| Ca. Jettenia asiatica
HM769655.1| Ca. Kuenenia sp. enrichment clone
HM769654.1| Ca. Kuenenia sp. enrichment clone
HM769656.1| Ca. Kuenenia sp. enrichment clone
Cluster 44
Cluster 116
Cluster 86
Cluster 114
Cluster 2
DQ459989.1| Ca. Brocadia fulgida
Cluster 1
JF487828.1| Ca. Brocadia caroliniensis strain
Cluster 152
Cluster 217
Cluster 14
Cluster 7
GQ850473.1| Ca. Anammoxoglobus sp. enrichment clone
GQ850476.1| Ca. Anammoxoglobus sp. enrichment clone
GQ850479.1| Ca. Anammoxoglobus sp. enrichment clone
GQ850472.1| Ca. Anammoxoglobus sp. enrichment clone
Cluster 90
Cluster 9
Cluster 50
Cluster 23
Cluster 124
Cluster 106
Cluster 16
Cluster 56
Cluster 22
NR 025327.1| Planctomyces maris strain
0.05
Ca. Scalindua marina/brodae/sorokinii
Scalindua related
Ca. Scalindua wagneri
Anammox related A
Ca. Brocadia A
Ca. Jettenia spp.
Ca. Kuenenia spp.
Ca. Brocadia B
Anammox related B
Potentially non-anammox

Figure 5.4: Phylogenetic tree of OTUs from Medway Estuary. Blue text indicates taxonomic classifications designated to OTUs. OTUs were defined at a similarity cut-off of 95%. A total of 294 OTUs were defined at this cut-off from 84,379 reads. Only OTUs which represented greater than 0.1% of the total number of reads are included in this tree (n=49 OTUs). Reference sequences were obtained from NCBI and include accession numbers. Sequences were aligned using MUSCLE (Edgar, 2004) and phylogeny inferred using a neighbour-joining methodology with a bootstrap test involving 1000 replications using MEGA 5 (Tamura, *et al.*, 2011). Branch lengths with bootstrap values greater than 50% are labelled.

Reference sequences for each OTU were aligned with anammox 16S rRNA sequences from NCBI as outlined in section 5.2. A neighbour-joining phylogenetic tree built from this alignment can be observed in Figure 5.4. OTUs were show to cluster with sequences from all the known anammox genera except *Ca.* Anammoxoglobus. A total of ten groups containing Medway OTUs were defined (Figure 5.4). In addition two, anammox related clades ('Anammox related A' and 'Anammox related B') were also observed though one of these clusters comprised of largely unsupported branch nodes. A further clade was also observed which appeared to cluster outside of the recognised *Brocadiales* clade and was therefore designated as 'Potentially non-anammox' OTUs. Table 5.1 indicates taxonomic identities assigned to each OTU obtained from these data.

Cluster	Taxonomic Group
Cluster 0	Other Scalindua
Cluster 1	*Ca.* Brocadia B
Cluster 2	*Ca.* Brocadia B
Cluster 3	*Ca.* Brocadia A
Cluster 5	Anammox related A
Cluster 6	Other Scalindua
Cluster 7	Anammox related B
Cluster 8	*Ca.* Scalindua wagneri
Cluster 9	Pot. Non-anammox
Cluster 10	Other Scalindua
Cluster 12	*Ca.* Brocadia A
Cluster 14	Anammox related B
Cluster 15	*Ca.* Scalindua wagneri
Cluster 16	Pot. Non-anammox
Cluster 17	Scalindua related
Cluster 18	Scalindua related
Cluster 19	Other Scalindua
Cluster 20	*Ca.* Scalindua wagneri
Cluster 22	Pot. Non-anammox
Cluster 23	Pot. Non-anammox
Cluster 24	Other Scalindua
Cluster 25	*Ca.* Brocadia A
Cluster 28	Other Scalindua
Cluster 33	Other Scalindua
Cluster 36	Other Scalindua

Cluster	Taxonomic Group
Cluster 43	Anammox related A
Cluster 44	*Ca.* Kuenenia
Cluster 50	Pot. Non-anammox
Cluster 52	Other Scalindua
Cluster 56	Pot. Non-anammox
Cluster 57	Anammox related A
Cluster 58	Other Scalindua
Cluster 60	Anammox related A
Cluster 83	Scalindua related
Cluster 85	*Ca.* Jettenia
Cluster 86	*Ca.* Brocadia B
Cluster 90	Pot. Non-anammox
Cluster 97	Other Scalindua
Cluster 101	Anammox related A
Cluster 102	*Ca.* Scalindua wagneri
Cluster 106	Pot. Non-anammox
Cluster 114	*Ca.* Brocadia B
Cluster 116	*Ca.* Brocadia B
Cluster 124	Pot. Non-anammox
Cluster 129	*Ca.* Scalindua wagneri
Cluster 138	Anammox related A
Cluster 149	*Ca.* Scalindua wagneri
Cluster 152	Anammox related B
Cluster 217	Anammox related B

Table 5.1: Table of assigned taxonomic groups to Medway clusters. Clusters defined at 95% similarity. Only OTUs representing > 0.1% of the total number of reads (84,379 reads) are included in this table.

5.4.2.3. Richness

Table 5.2 shows the richness (defined here as the number of different OTUs obtained after clustering at 95% similarity) of samples M1-M7* along the Medway Estuary. The total richness appeared to be lower at either end of the sampled range of the estuary *i.e.* towards the freshwater end (M1) and the marine end (M7); 20 different OTUs were observed at M1 and 21 at M7. Sites M2-M6 demonstrated a higher degree of richness ranging from 31 to 34 OTUs. The number of non-Scalindua OTUs decreased from M1/M2 (8/9 OTUs) to M7 (2 OTUs) whereas the number of OTUs related to *Ca.* Scalindua demonstrated the opposite trend, increasing from 9 OTUs at M1 to 18 different Scalindua OTUs at M6. M7 however contained fewer *Ca.*

* For a list of sampling locations and names please refer to chapter 2. Henceforth samples will be referred to by their sample codes M1-M7 which are in order of geographical distance from Allington Lock (M1) the most freshwater site sampled.

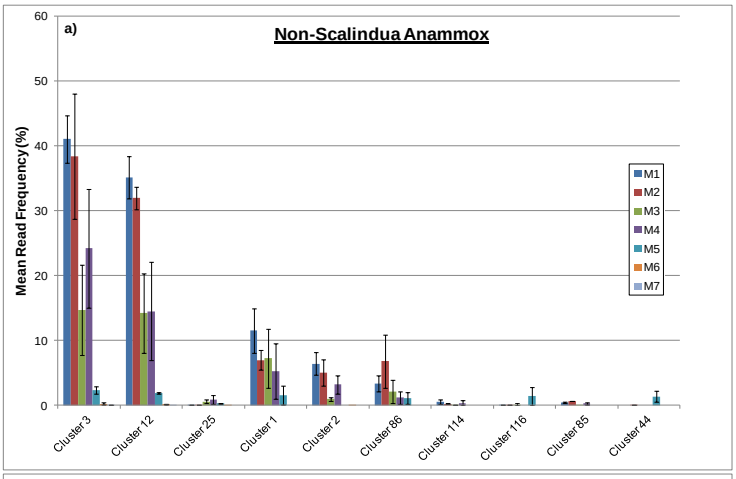
Scalindua OTUs (8 OTUs) than M6. OTUs related to 'Potentially non-anammox' and 'Anammox related B' clusters were only observed in samples M6 and M7. OTUs representing the 'Anammox related A' cluster were observed ubiquitously in all samples though demonstrated a higher richness towards the middle of the sampling range (M2-M5).

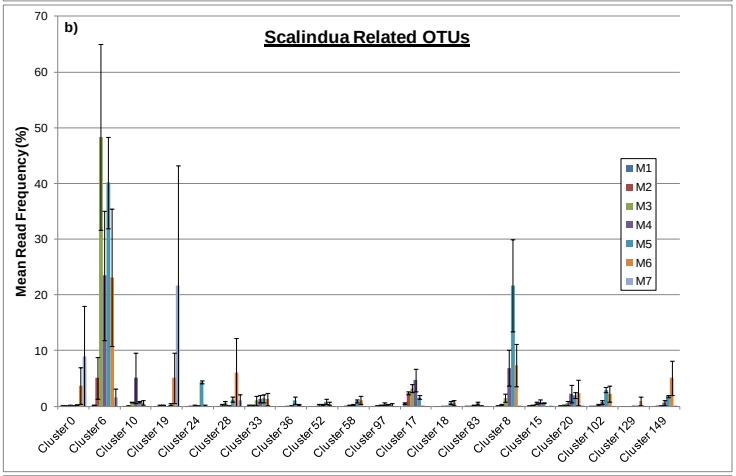
Taxa	M1	M2	M3	M4	M5	M6	M7
Ca. Brocadia A	3	3	3	3	3	3	2
Ca. Brocadia B	5	5	5	4	3	1	0
Ca. Jettenia	1	1	1	1	0	0	0
Ca. Kuenenia	0	1	0	0	1	0	0
Ca. Scalindua wagneri	4	5	5	6	6	5	4
Other Scalindua	3	9	11	10	11	11	4
Scalindua related	2	2	3	3	3	2	0
Amx related A	2	5	6	6	5	3	2
Amx related B	0	0	0	0	0	4	4
Pot. Non-anammox	0	0	0	0	0	4	5
Total	20	31	34	33	32	33	21

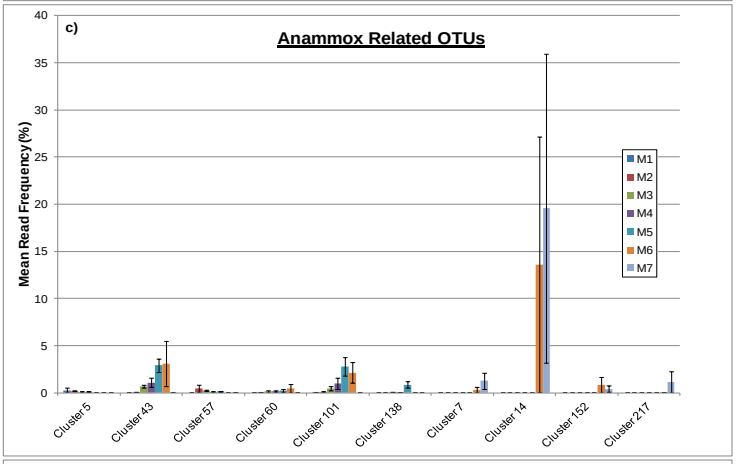
Table 5.2: Table showing sample richness from Medway Estuary sites per taxanomic group. Table shows the number of OTUs (95% similarity) in each sample (richness). Richness is calculated from the OTU abundances from triplicate samples and OTUs were counted even if they only appeared once across the three triplicates.

5.4.2.4. Diversity and Distribution

Figure 5.5 shows mean frequency distributions and standard errors for each OTU. Of the 10 non-Scalindua OTUs, only 5 (Clusters 3, 12, 1, 2 and 86) appeared to be important in reporting anammox diversity in these samples as the other 5 OTUs represented less than 1% of the reads in each samples (Figure 5.5a). Data from sites M6 and M7 contained only a small number of reads from these OTUs. Of the 5 OTUs representing greater than 1% of the total number of reads from each site, only Clusters 3 and 12 appeared to show any substantial difference between different sites. Both Clusters 3 and 12 had a higher proportion of reads assigned to them in M1 and M2 compared with M3-M7. However, in Cluster 3 the error bars for M2 and M4 are seen to overlap. A decreasing trend in the number of reads assigned to Clusters 1, 2 and 86 may be observed between M1 and M7 though this is not significant as determined by overlapping standard error bars (although no reads from M6 and M7 were associated with these clusters).


a)
Non-Scalindua Anammox
Mean Read Frequency (%)
60
50
40
30
20
10
0
M1
M2
M3
M4
M5
M6
M7
Cluster 3
Cluster 12
Cluster 25
Cluster 1
Cluster 2
Cluster 86
Cluster 114
Cluster 116
Cluster 85
Cluster 44


b)
Scalindua Related OTUs
Mean Read Frequency (%)
70
60
50
40
30
20
10
0
M1
M2
M3
M4
M5
M6
M7
Cluster 0
Cluster 6
Cluster 10
Cluster 19
Cluster 24
Cluster 28
Cluster 33
Cluster 36
Cluster 52
Cluster 58
Cluster 97
Cluster 17
Cluster 18
Cluster 83
Cluster 8
Cluster 15
Cluster 20
Cluster 102
Cluster 129
Cluster 149


c)
Anammox Related OTUs
Mean Read Frequency (%)
40
35
30
25
20
15
10
5
0
M1
M2
M3
M4
M5
M6
M7
Cluster 5
Cluster 43
Cluster 57
Cluster 60
Cluster 101
Cluster 138
Cluster 7
Cluster 14
Cluster 152
Cluster 217

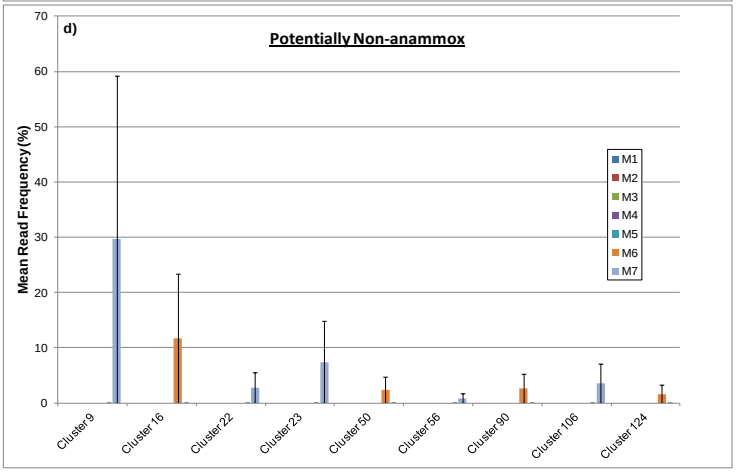

d)
Potentially Non-anammox
Mean Read Frequency (%)
70
60
50
40
30
20
10
0
M1
M2
M3
M4
M5
M6
M7
Cluster 9
Cluster 16
Cluster 22
Cluster 23
Cluster 50
Cluster 56
Cluster 90
Cluster 106
Cluster 124

Figure 5.5: Mean frequency distributions of Medway Estuary clusters. Frequencies are represented as the mean percentage of reads assigned to each cluster per sample for a) non-Scalindua anammox OTUs, b) Scalindua OTUs, c) anammox related OTUs and d) potentially non-anammox OTUs. All samples were sequenced in triplicate. Error bars represent ± 1 S.E. Only clusters representing >0.01% of the total number of reads (84,379) are included in these figures.

The frequency of OTUs associated with Scalindua organisms (Figure 5.5b) appeared to be more evenly distributed although most Scalindua related reads (48.42%) appeared to cluster within one OTU (Cluster 6). Despite large standard error bars, Cluster 6 appeared to show a difference between sites to the extremes of the sampling range (M1, M2 and M7) and sites M3-M6, with these middle sites exhibiting a greater number of reads associated with this OTU. Cluster 17 also appears to show higher read frequencies between sites M2 and M5. However, most OTUs appeared to show a trend in increasing frequencies towards the marine end of the estuary (*i.e.* towards M7) though only in clusters 58, 8, 102 and 149 were standard error bars (±1 SE) not seen to overlap.

Anammox related OTUs (Figure 5.5c) generally exhibited a low read abundance although clusters 43 and 101 represented up to ~2.5% of the total number of reads in some sites and appear to demonstrate an increase trend from M2 to M6. Cluster 14 contained the largest proportion of reads of the 'Anammox related' OTUs however reads were only assigned to this OTU from 1 M6 triplicate sample and 2 M7 triplicate samples resulting in large error bars.

Reads assigned to 'Potentially non-anammox OTUs' (Figure 5.5d) were only observed at sites M6 and M7. However, as is evident from Figure 5.5d, all of these OTUs are associated with large standard errors as these reads were not found in all of the triplicate samples from these sites.

Figure 5.6 shows the mean percentage of reads attributed to each taxonomic grouping as defined from Figure 5.4. The percentage of reads associated with non-Scalindua anammox organisms is greater upstream of the estuary and represent greater than 90% of the reads in sites M1 and M2. Non-Scalindua anammox reads are dominated by *Ca.* Brocadia spp. with only a few reads associated with *Ca.* Jettenia and *Ca.* Kuenenia spp. However, downstream sites (M5 and M6) are dominated by *Ca.* Scalindua spp. with ~80% of reads in M5 and ~60% in M6

identified as belonging to this clade. Sites M3 and M4 exhibit more equal proportions of non-Scalindua and Scalindua related reads. Data from M7 comprised of a large number of 'Anammox related' and 'Potentially non-anammox' reads. 'Anammox related' reads were observed within data from all sites though were more abundant downstream, particularly in M6 and M7.

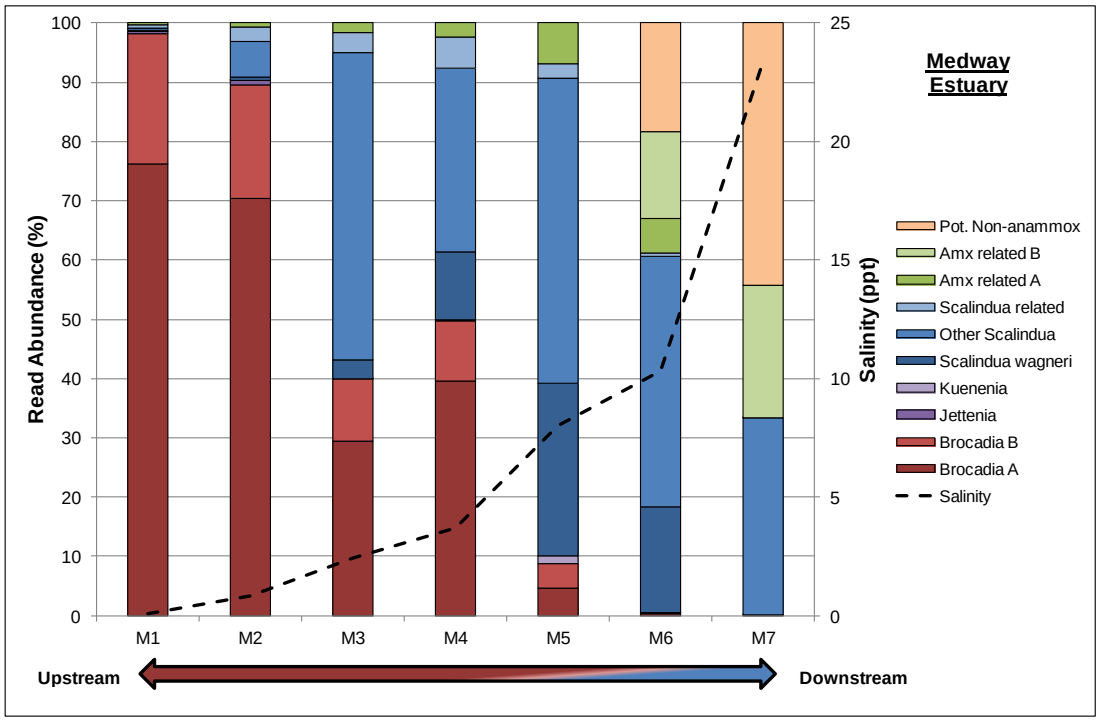


Figure 5.6: Mean Medway Estuary read abundance per sample, expressed as a percentage. Mean values for each OTU were calculated from triplicate samples. The mean abundance of OTUs which had been assigned to specific taxonomic groups (see Figure **5.4**) were combined as mean values and are presented in these figures. The black, dotted line shows the salinity (ppt) at each site which is extrapolated from twelve year average data (Environment Agency, UK) for the sampling month (March). Upstream sites M1 and M2 are dominated by non-Scalindua anammox genera whereas downstream sites have a greater proportion of OTUs associated with *Ca.* Scalindua spp.

Medway 454 pyrosequencing data were analysed in conjunction with Environment Agency data (see section 5.2.5) in order to attempt to elucidate the observed trends in terms of their environmental significance. An unconstrained CCA plot for these data can be seen in Figure 5.7. The majority of the diversity is represented along the *x*-axis (CA1) indicating a clear trend between upstream and downstream sites. There is also considerable diversity as represented by the *y*-axis (CA2) showing a trend in diversity between mid-estuary sites (M3, M4, M5) and sites towards either end of the sampling range associated with an increase in diversity towards the middle of the

estuary (see above and Table 5.2 and Figure 5.6). The trend in diversity between M1 and M7 was positively correlated with increasing salinity and conductivity (which is related to salinity) downstream and with increased DIN (NO_2^- + NO_3^- + NH_4^+) and PO_4^{3-} upstream*. These environmental factors appeared to be the strongest drivers of anammox diversity as expressed by these data. The trend towards greater diversity and richness towards the middle of the samples range (as expressed by the *y*-axis) may be correlated negatively with temperature, BOD, dissolved O_2 and pH though if this were true, the observed correlation was weak.

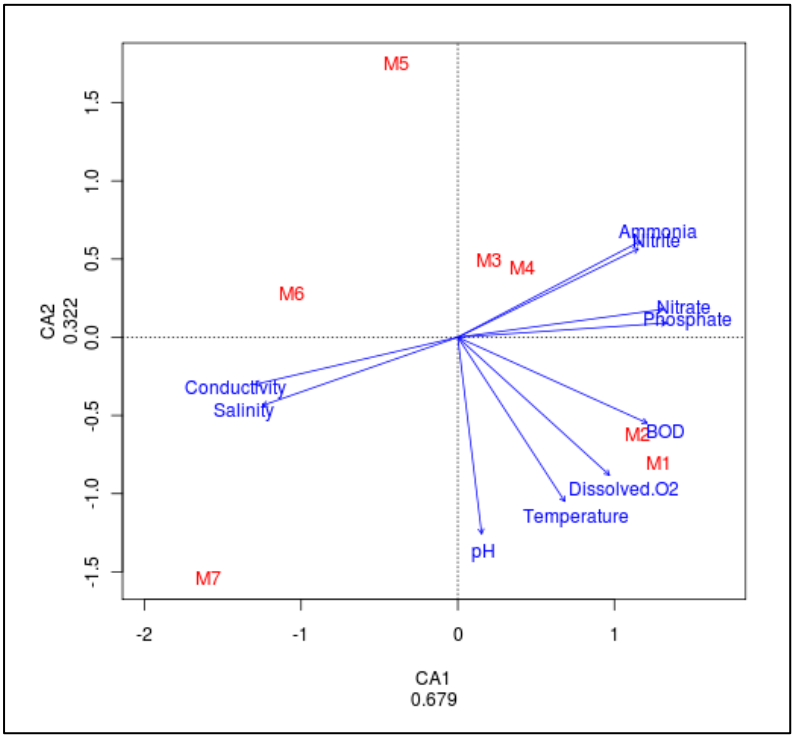


Figure 5.7: Unconstrained CCA plot for Medway Estuary 454 pyrosequencing data with environmental variables. CCA was constructed using all OTUs which represented greater than 0.1% of the total number of reads. A clear trend between sites upstream and downstream can be seen represented by CA1 (*x* axis). A trend also appears to be present between sites at the extremities of the estuary and sites towards the middle as represented by CA2 (*y* axis).

* Naturally a positive correlation with increasing salinity downstream corresponds to a negative correlation with salinity upstream.

5.4.3. ETNP Diversity

5.4.3.1. ETNP Characterisation

Profile data was collected at each of the six sites during sampling of the ETNP OMZ. Sites drew a longitudinal gradient between the latitudes ~8°N and 13°N (see Figure 2.3). Mean O2 concentrations across this gradient can be seen in Figure 5.8. The ETNP OMZ in this region was characterised by a shallow OMZ and a steep oxycline from the surface to approximately 50 m. The core of the OMZ (where concentrations remained ~0 µmol/kg) extended from approximately 300 m and 800 m. The top of the OMZ sporadically had horizons of slightly higher concentrations of O_2 up to ~10 µmol/kg. Below the core of the OMZ, a gradual oxycline was observed down to a depth of approximately 2,200 m where O_2 concentrations reached and maintained ~108 µmol/kg. The OMZ appeared to be slightly thicker to the north with the core extending to about a depth of 1000 m. This may be due to greater upwelling of nutrients at the coastal margin.

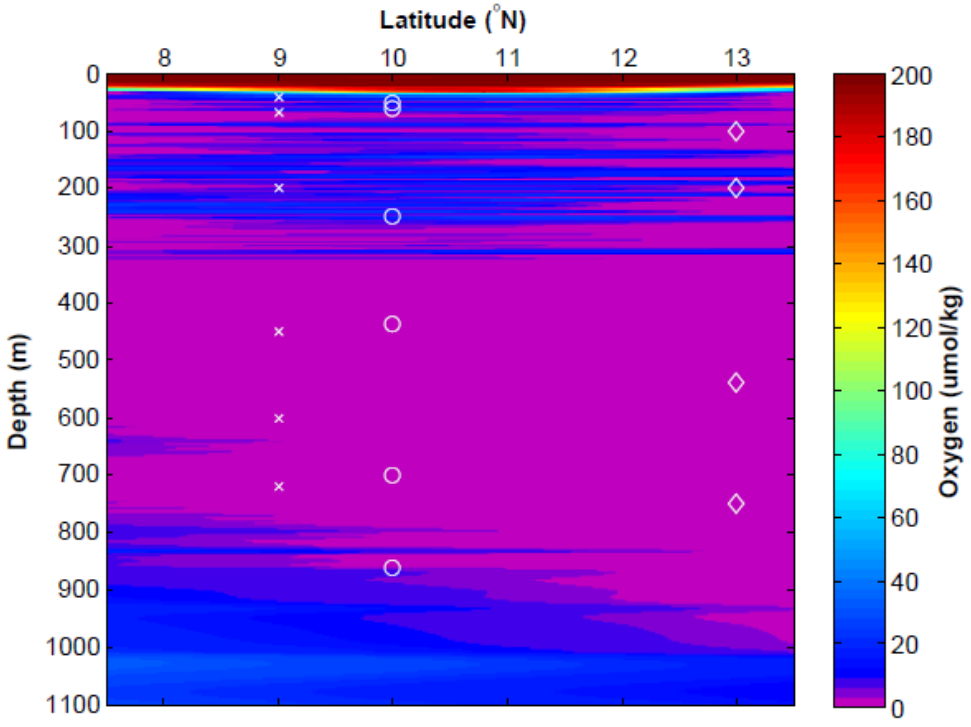


Figure 5.8: O_2 depth profile of ETNP OMZ. Plot calculated using mean O_2 (µmol/kg) data from every CTD at each station. The OMZ was characterised by a steep oxycline with the upper limit at about 50 m. The core of the OMZ is situated from approximately 300 m to 700 m. Sampling sites investigated in this study (*i.e.* 2, 4 and 5) are indicated by diamonds (site 2), circles (site 4) and crosses (site 5).

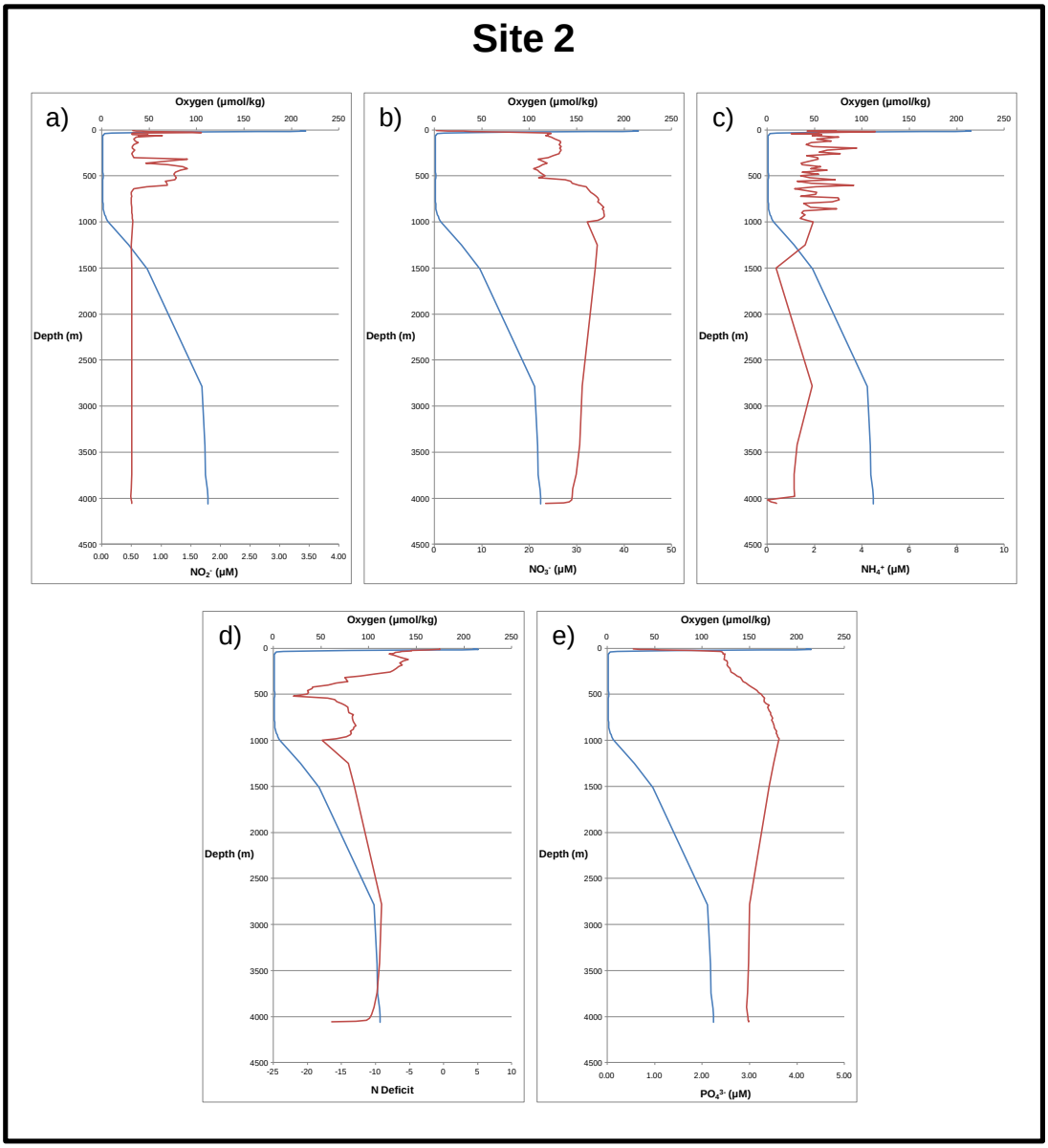


Figure 5.9: ETNP Profile data for Site 2. Profiles show data for a) NO2-, b) NO3-, c) NH4+, d) Nitrogen deficit and e) PO43-. Data for each graph is plotted in red and are plotted against O2 concentrations (blue).

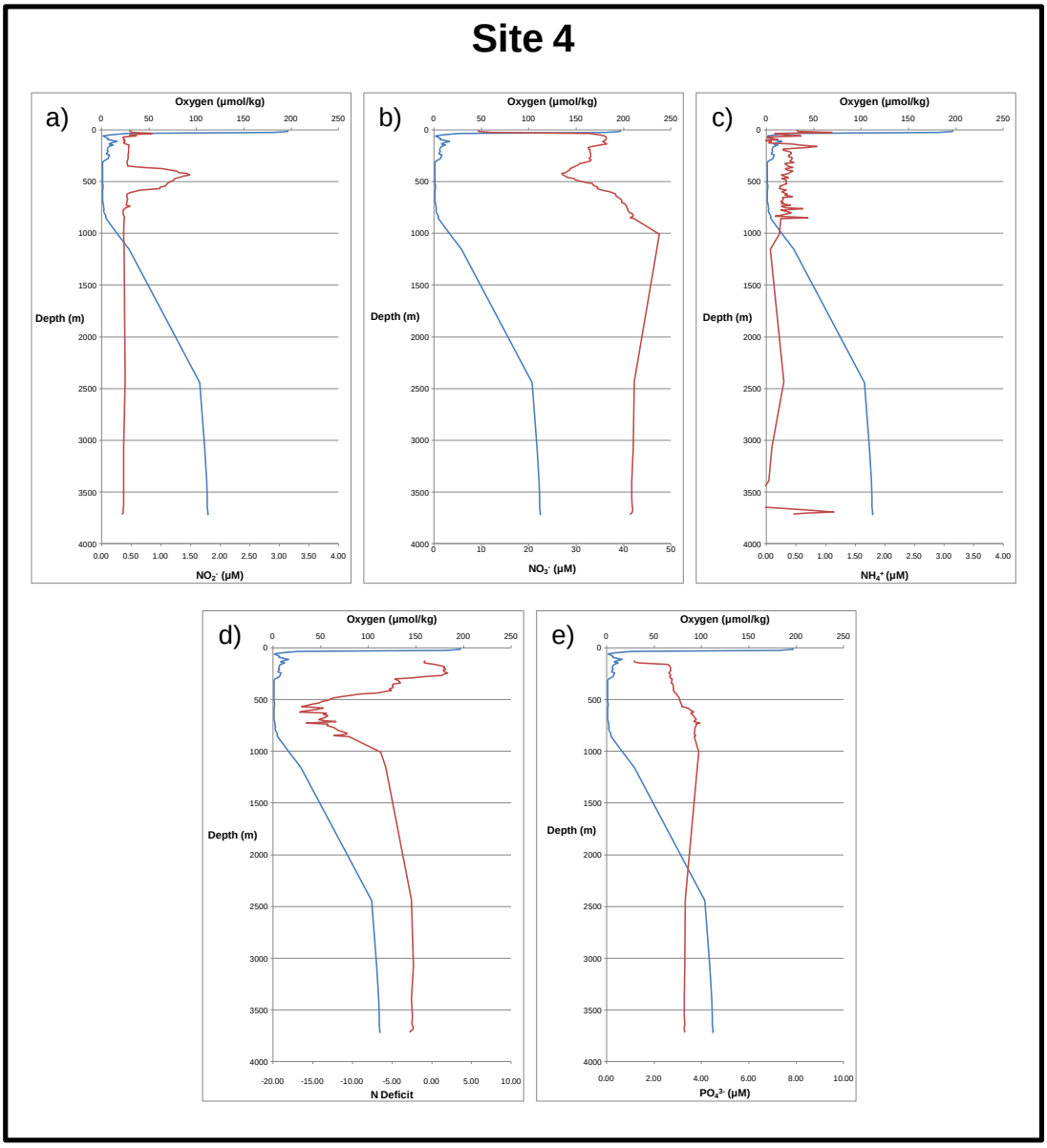


Figure 5.10: ETNP Profile data for Site 4. Profiles show data for a) NO2-, b) NO3-, c) NH4+, d) Nitrogen deficit and e) PO43-. Data for each graph is plotted in red and are plotted against O2 concentrations (blue).

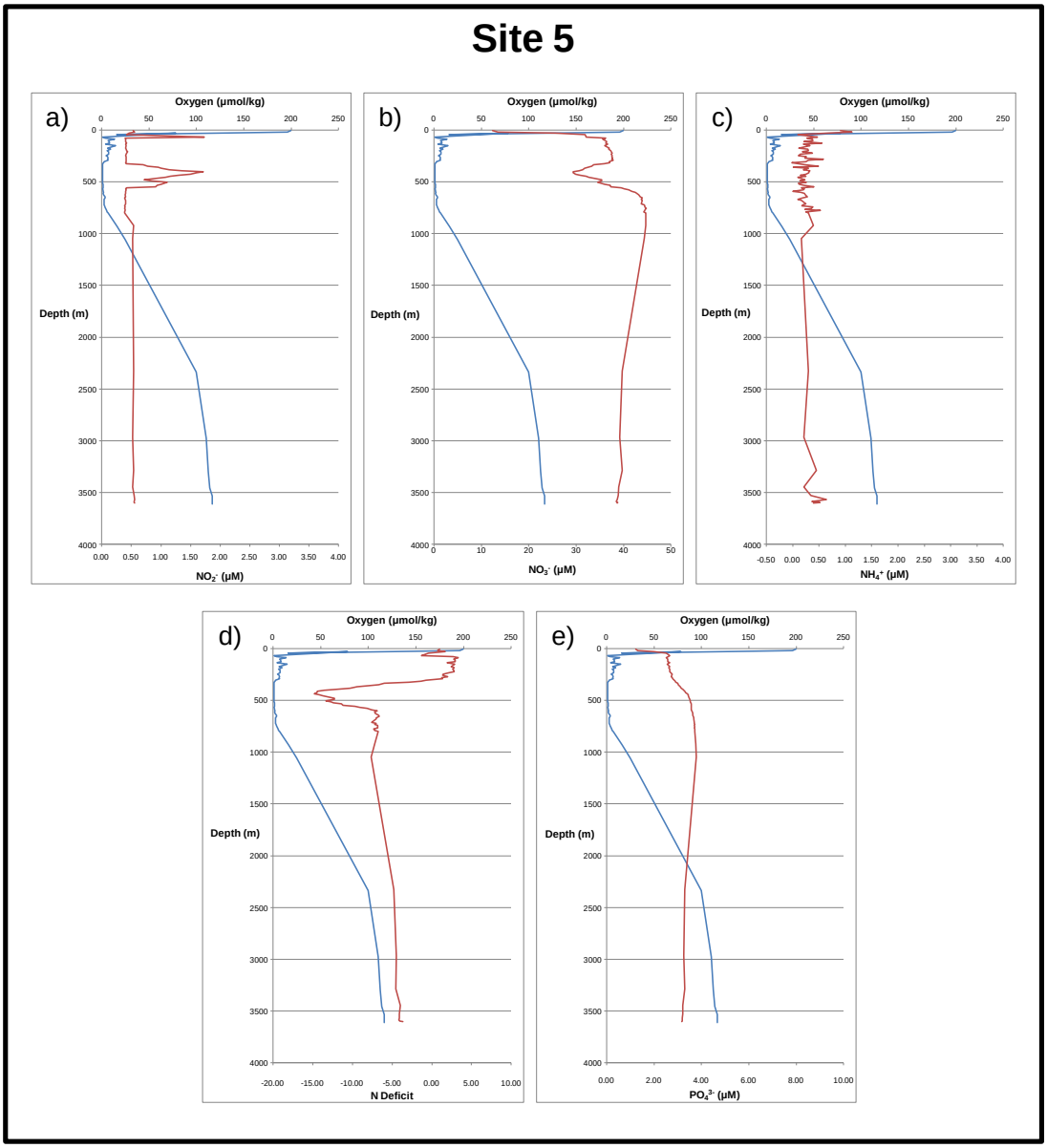


Figure 5.11: ETNP Profile data for Site 5. Profiles show data for a) NO2-, b) NO3-, c) NH4+, d) Nitrogen deficit and e) PO43-. Data for each graph is plotted in red and are plotted against O2 concentrations (blue).

Depth profiles for sites 2, 4 and 5 can be seen in Figure 5.9, Figure 5.10 and Figure 5.11. These data describe a reasonably constant nutrient profile throughout the OMZ. A peak in NO_2^- concentrations (up to ~1.5 µM) was observed at around 500 m. This correlates with a reduction in NO_3^- concentrations and a decrease in N_{def} (nitrogen deficit, indicating a loss of fixed N) calculations, consistent with a peak in denitrification at this depth (Devol, *et al.*, 2006; Chang, *et al.*, 2012; Nagel, *et al.*, 2013). PO_4^{3-} concentrations remained almost constant throughout the OMZ, ranging between 2.50 and 3.00 µM. NH_4^+ concentrations were low throughout the water column (~1.5–4 µM) but appeared to fluctuate greatly over fine horizons. However,

difficulties were reported to have been experienced on the cruise when making these measurements and hence these data may be prone to errors (personal communication).

5.4.3.2. Quality Control

454 pyrosequencing of anammox 16S rRNA PCR products obtained from triplicate* samples from sites 2, 4 and 5 produced a total of 264,050 reads. Of these reads, 225,602 contained both the forward and reverse primer sequences.

Length distribution analyses of these data indicated that the minimum read length was 46 nucleotides, the maximum was 557 nucleotides and the median was 477. 2,868 reads contained more than one ambiguous base (Ns). However the distribution of lengths (data not shown) was not entirely centred about the median (although this was the largest peak in length frequencies) and a secondary peak was observed centred around length 506 nucleotides (maximum peak height was 895 reads).

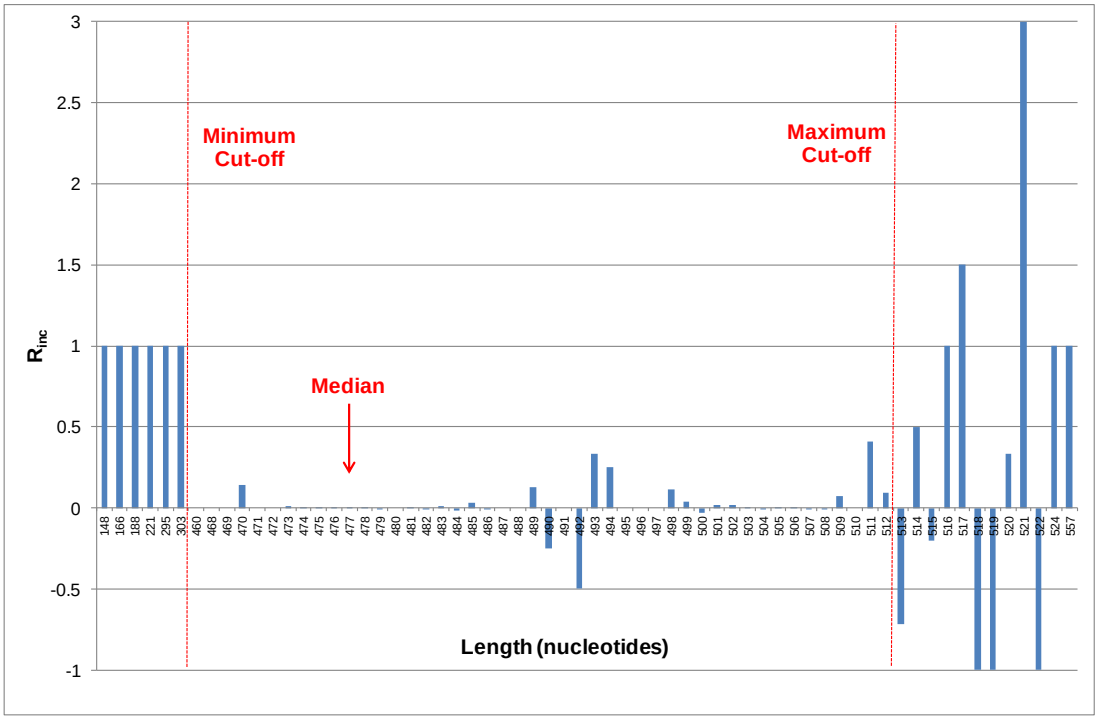


Figure 5.12: R_{inc} values for all 454 pyrosequencing data from ETNP sites 2, 4 and 5. Only read lengths which were represented in the data are shown. Red hashed lines indicate chosen minimum and maximum cut-off values from which reads between these two values were used for downstream analysis.

* N.B. Only two replicates were available for S2 100m as one triplicate was lost between sampling and DNA extraction

ReLIC and calculation of R_{inc} values (Figure 5.12) indicated a clear division between low and high R_{inc} values towards lower read lengths, suggesting that a minimum length cut-off of 460 nucleotides would omit poor quality reads whilst including good quality sequences and not removing true low abundance clusters. However, as previously found with this type of analysis, the upper boundary did not show such a clear definition and made the selection of a maximum length cut-off more difficult. R_{inc} values did however appear to increase significantly after a length of 512 and so this was chosen as the upper boundary. Trimming of read lengths lower than 460 bases and greater than 512 bases resulted in a total of 222,565 reads, with no ambiguous bases, for downstream processing.

5.4.3.3. Phylogenetic Analysis

Clustering of reads which passed quality control, using a similarity cut-off of 95% to define clusters, produced 44 potential OTUs, 8 of which were singleton clusters and 25 contained less than 0.1% of the total reads. Initially, these 33 clusters were included in phylogenetic analyses as, the low number of potential OTUs created (*i.e.* 44) meant that this was a manageable number of OTUs to process (*c.f.* 294 OTUs obtained from similar Medway analysis, see section 5.4.2.1)[*]. It was also decided to include such clusters in phylogenetic analysis to assess the validity of the assumption that low abundance clusters had a greater likelihood of representing erroneous and poor quality sequences.

[*] N.B. As stated previously, the logic behind omitting these low abundance clusters from analysis was to remove clusters which were highly likely to be as a result of poor quality sequencing, in the knowledge that true, rare taxa would likely be omitted. The omission of such 'good' OTUs was deemed unimportant as the aims of these investigations was to measure relative anammox diversity and therefore, providing such omissions were made consistently, this would have little, if any, effect on the overall conclusions due to the perceived low ecological significance of these clusters.

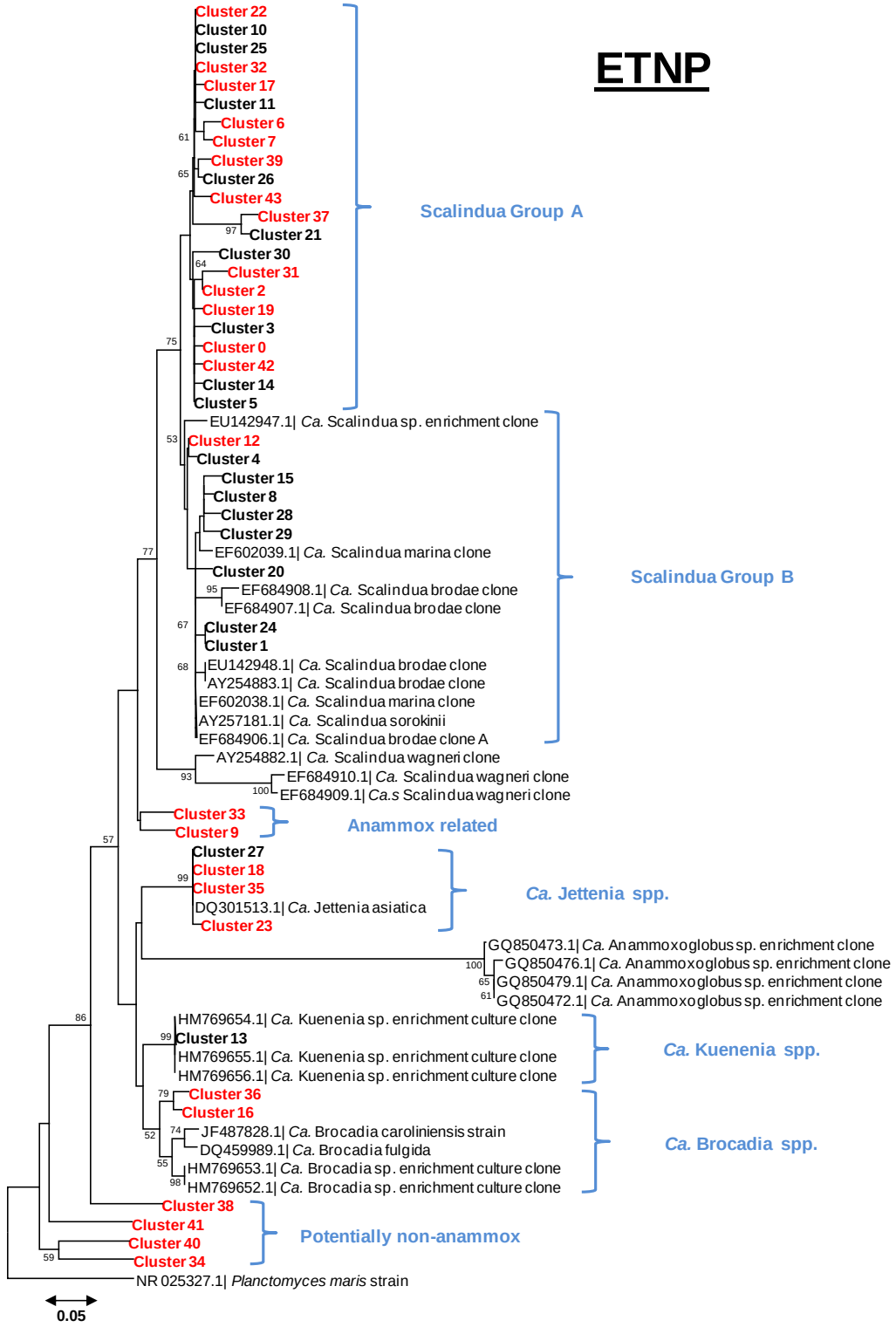


Figure 5.13: Phylogenetic tree of OTUs from ETNP. Blue text indicates taxonomic classifications designated to OTUs. Red text indicates OTUs which represent <0.1% of the total number of reads which passed quality control (222,565). OTUs were defined at a similarity cut-off of 95%. Reference sequences were obtained from NCBI and include accession numbers. Sequences were aligned using MUSCLE (Edgar, 2004) and phylogeny inferred using a neighbour-joining methodology with a bootstrap test involving 1000 replications using MEGA 5 (Tamura, *et al.*, 2011). Branch lengths with bootstrap values greater than 50% are labelled.

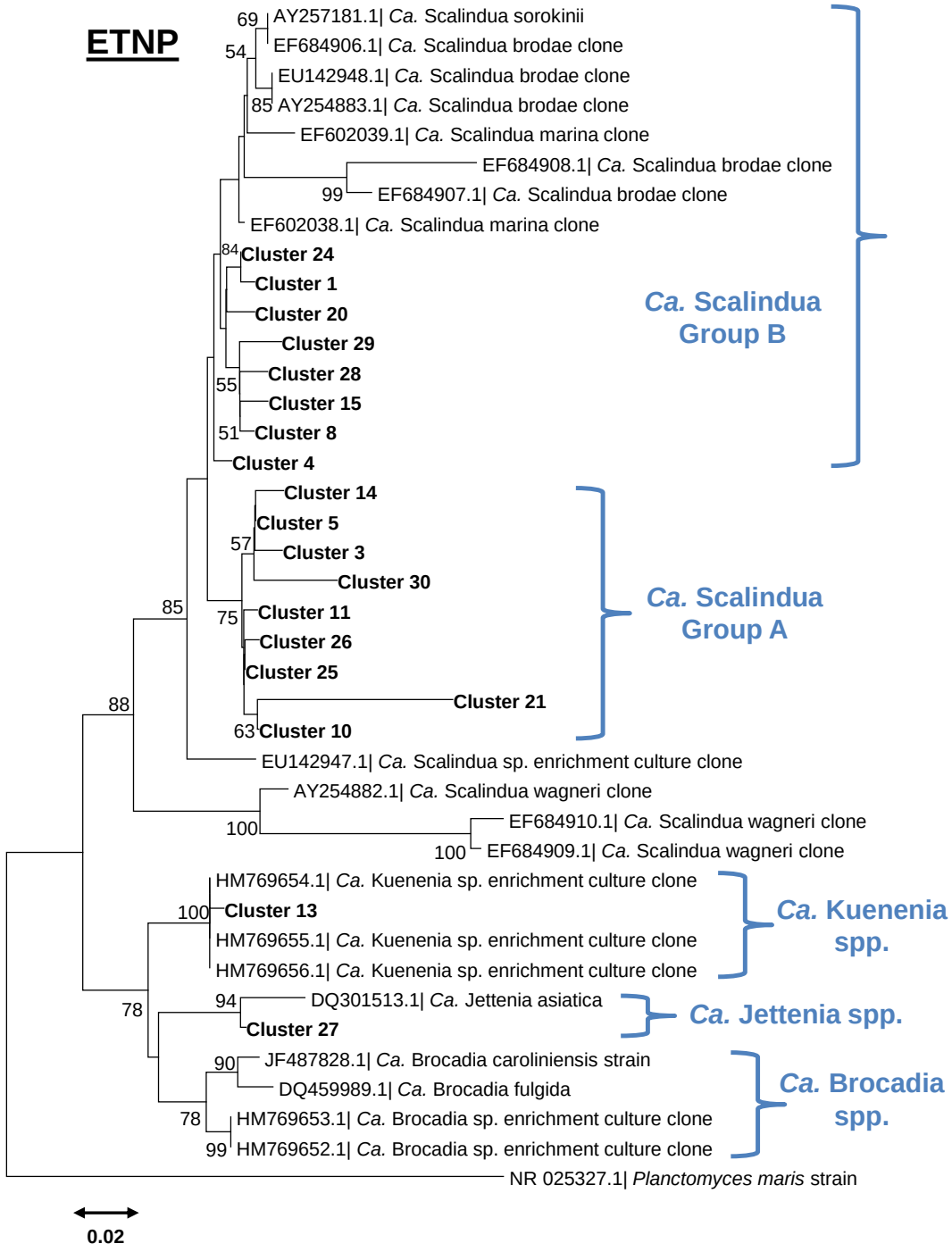


Figure 5.14: Phylogenetic tree of OTUs from ETNP omitting low abundance OTUs. Tree is identical to that shown in Figure **5.13** except that OTUs representing <0.1% of the total number of reads (highlighted in red in Figure **5.13**) had been omitted from phylogenetic analysis. The phylogeny expressed by this tree is congruent with that in Figure **5.13** and so the addition of these low abundance OTUs would not affect taxonomic classification and further analysis.

> 0.1 % of Reads	
Cluster	**Taxonomic group**
Cluster 1	Scalindua Group B
Cluster 3	Scalindua Group A
Cluster 4	Scalindua Group B
Cluster 5	Scalindua Group A
Cluster 8	Scalindua Group B
Cluster 10	Scalindua Group A
Cluster 11	Scalindua Group A
Cluster 13	*Ca.* Kuenenia sp.
Cluster 14	Scalindua Group A
Cluster 15	Scalindua Group B
Cluster 20	Scalindua Group B
Cluster 21	Scalindua Group A
Cluster 24	Scalindua Group B
Cluster 25	Scalindua Group A
Cluster 26	Scalindua Group A
Cluster 27	*Ca.* Jettenia sp.
Cluster 28	Scalindua Group B
Cluster 29	Scalindua Group B
Cluster 30	Scalindua Group A

<0.1 % of Reads	
Cluster	**Taxonomic group**
Cluster 0	Scalindua Group A
Cluster 2	Scalindua Group A
Cluster 6	Scalindua Group A
Cluster 7	Scalindua Group A
Cluster 9	Anammox related
Cluster 12	Scalindua Group B
Cluster 16	*Ca.* Brocadia sp.
Cluster 17	Scalindua Group A
Cluster 18	*Ca.* Jettenia sp.
Cluster 19	Scalindua Group A
Cluster 22	Scalindua Group A
Cluster 23	*Ca.* Jettenia sp.
Cluster 31	Scalindua Group A
Cluster 32	Scalindua Group A
Cluster 33	Anammox related
Cluster 34	Pot. non-anammox
Cluster 35	*Ca.* Jettenia sp.
Cluster 36	*Ca.* Brocadia sp.
Cluster 37	Scalindua Group A
Cluster 38	Pot. Non-anammox
Cluster 39	Scalindua Group A
Cluster 40	Pot. non-anammox
Cluster 41	Pot. non-anammox
Cluster 42	Scalindua Group A
Cluster 43	Scalindua Group A

Table 5.3: Table of assigned taxonomic groups to ETNP clusters. Clusters defined at 95% similarity. The left-hand table (blue) contains clusters which represent >0.1% of the total number of reads after quality control (222,565). The right-hand table (red) represent clusters which represent <0.1% of the total number of reads and therefore reads assigned to these clusters may have a number of errors.

A neighbour-joining phylogenetic tree of ETNP OTUs, including OTUs representing <0.1% of the total number of reads, from sites 2, 4 and 5 is presented in Figure 5.13. A phylogenetic tree was also conducted omitting these low abundance OTUs (Figure 5.14) to ensure that the inclusion of low abundance OTUs did not affect the overall phylogenetic consensus and taxonomic inference. Comparison of Figure 5.13 and Figure 5.14 clearly demonstrates that the same phylogeny was encountered with and without the inclusion of the 33 low abundance OTUs and that all of the 11 larger OTUs maintained the same taxonomies in both trees. Therefore the inclusion of low abundance OTUs did not affect the inference of taxonomic identities and could be included in the analysis without affecting the results.

OTUs were designated into taxonomic groups based on this phylogeny the results of which are depicted in Table 5.3. The majority of OTUs (n=30 OTUs) were associated with *Ca.* Scalindua spp. and clustered within two distinct groups designated 'Scalindua Group A' and 'Scalindua Group B'. 'Scalindua Group A' however did not contain any reference anammox 16S rRNA sequences and was comprised entirely of OTUs defined from these ETNP data. 'Scalindua Group B' was comprised of reference sequences from *Ca.* S. brodae, *Ca.* S. marina and *Ca.* S. sorokinii. No OTUs gleaned from ETNP sequence data were similar to *Ca.* S. wagneri 16S rRNA genes. The proportion of OTUs which represented less than 0.1% of the total number of reads was greater in 'Scalindua Group A' (13/22) than 'Scalindua Group B' (1/9).

7 OTUs were associated with non-Scalindua anammox organisms, *Ca.* Brocadia spp., *Ca.* Kuenenia spp. and *Ca.* Jettenia spp. No OTUs appeared to be related to the *Ca.* Anammoxoglobus genera. Of these non-Scalindua anammox OTUs, only 2 represented greater than 0.1% of the total number of reads in this dataset, Cluster 27 (*Ca.* Jettenia sp.) and Cluster 13 (*Ca.* Kuenenia sp.).

2 OTUs clustered outside of both the *Ca.* Scalindua and non-Scalindua anammox clades and were designated as 'Anammox related' OTUs, though both represented less than 0.1% of the total number of reads. 4 OTUs were seen to cluster outside of the known anammox diversity however all of these 'Potentially non-anammox' OTUs also represented less than 0.1% of the total number of reads, 3 of which were singleton clusters.

Rarefaction analysis on these data indicate that in the majority of samples, rarefaction curves (Figure 5.15) are plateauing suggesting that these data represent the entire diversity in these samples as described by these particular primers and hence and conclusions drawn from further measurements of relative anammox diversity are likely to be a fair representation of the true anammox diversity. However, some rarefaction curves are not seen to completely plateau, although they were beginning to do so suggesting that if taxa had been omitted from these samples, they are unlikely to affect measurements of relative anammox diversity greatly.

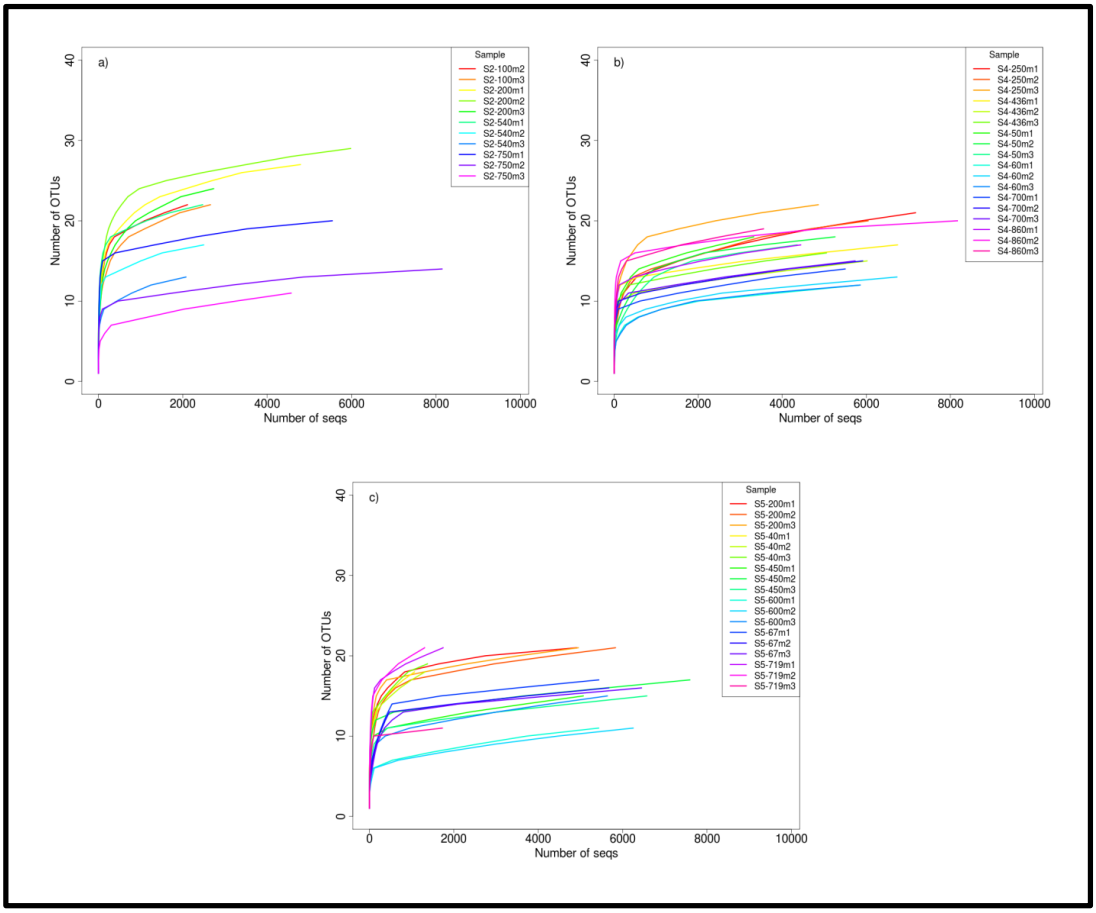


Figure 5.15: Rarefaction curves for all OTUs obtained from ETNP 454 pyrosequencing data. Rarefaction curves are shown for a) Site 2, b) Site 4 and c) Site 5. Rarefaction curves for most samples are shown to plateau suggesting that the entire anammox diversity at these sites is represented in these data (at least that is represented by the primers used). However, rarefaction analysis of some samples (e.g. S5-719m1) were not seen to completely plateau although they were beginning to do so. This might suggest that the sampling depth of these samples might not be sufficient to report the entire anammox diversity. However the low resolution of these rarefaction curves (i.e. < 20 OTUs) may prevent these curves to clearly plateau.

Taxa	Site 2				Site 4						Site 5					
	100 m	200 m	540 m	750 m	50 m	60 m	250 m	436 m	700 m	860 m	40 m	67 m	200 m	450 m	600 m	719 m
Ca. Brocadia sp.	0	0	0	0	0	0	0	0	0	0	2	1	0	1	0	0
Ca. Jettenia sp.	2	4	1	1	1	0	1	1	0	2	1	0	2	0	0	3
Ca. Kuenenia sp.	0	0	1	0	1	0	0	0	0	0	1	0	0	1	1	1
Scalindua Group A	13	16	12	14	10	6	15	12	13	14	13	8	15	14	12	11
Scalindua Group B	8	8	7	6	8	9	8	6	6	7	8	8	8	6	3	7
Anammox related	1	2	1	0	0	0	0	1	0	0	0	0	1	0	1	2
Pot. non-anammox	0	2	0	0	0	0	1	0	1	0	0	0	1	0	1	0
Total	**24**	**32**	**22**	**21**	**20**	**15**	**25**	**20**	**20**	**23**	**25**	**17**	**27**	**22**	**18**	**24**

Table 5.4: Table showing sample richness from sites 2, 4 and 5 per taxanomic group. Table shows the number of OTUs (95% similarity) in each sample (richness). Richness is calculated from the OTU abundances from triplicate samples and OTUs were counted even if they only appeared once across the three triplicates. At all sites, richness was greater at approximately 200-250 m.

5.4.3.4. Richness

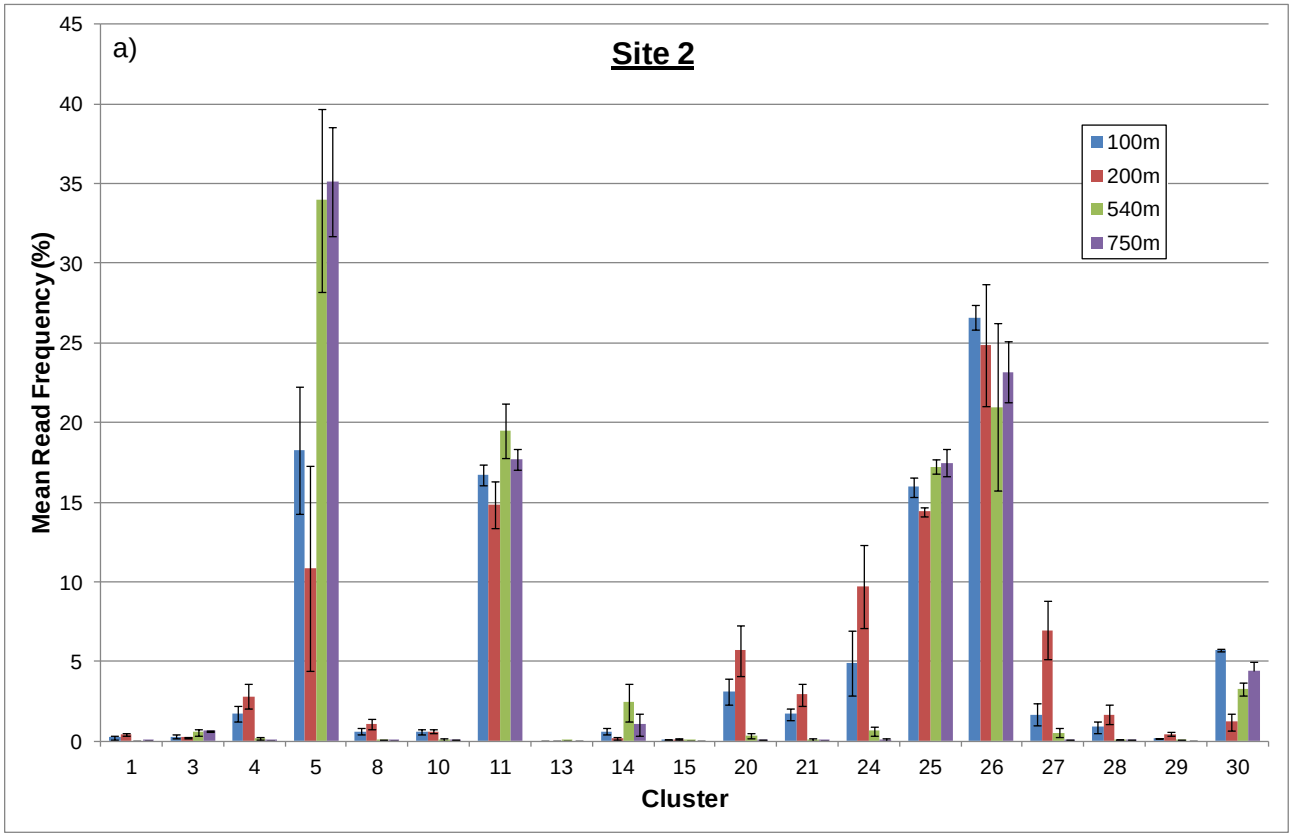
The richness of each sample, defined as the number of OTUs in each sample, is shown in Table 5.4. The majority of OTUs in every sample are associated with *Ca.* Scalindua organisms. In all 3 sites the greatest richness was observed at around 200 m corresponding to the upper core of the OMZ. The deeper core of the OMZ typically had a lower richness. Site 4 and 5 appear to show that the richness of these sites increases again towards the bottom of the OMZ and start of the bottom oxycline. This trend is not seen in samples from site 2 however this may be because the bottom oxycline was not sampled at this site (see Figure 5.8). At both sites 4 and 5 the lower half of the upper oxycline (S4-60 m and S5-67 m) appeared to show a marked drop in richness from the top half of the oxycline. Again, this portion of the oxygen profile had not been sampled at site 2.

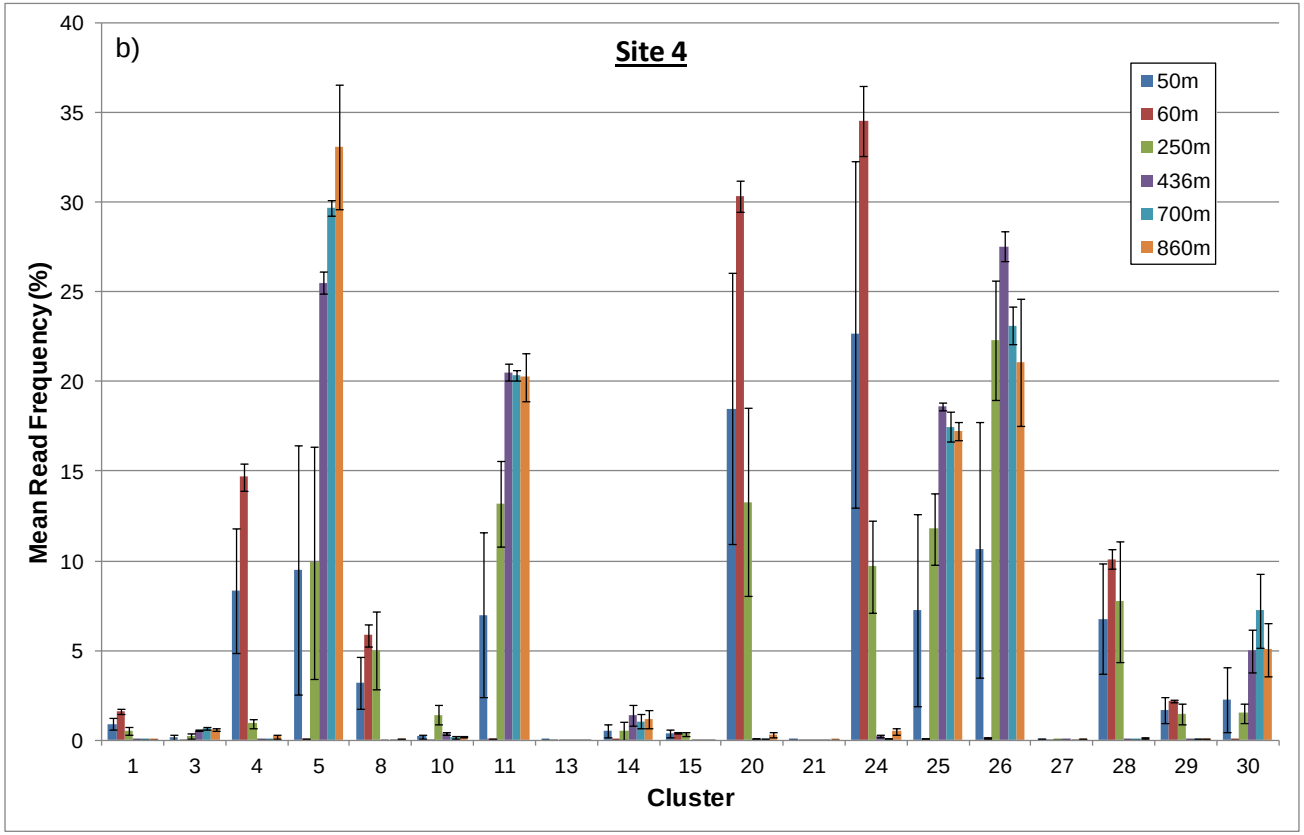
5.4.3.5. Diversity and Distribution

Frequency distributions for each OTU, which represented greater than 0.1% of the total number of reads, are depicted in Figure 5.16. Individual OTUs appear to group into three approximate classifications; OTUs which are ubiquitous and evenly distributed at all depths, OTUs which are more abundant towards the top of the OMZ and OTUs which have a greater abundance with depth. The majority of the 19 OTUs presented in Figure 5.16 are of a relatively low abundance, typically comprising of less than 1-2% of the total number of reads in each sample.

Clusters 4, 8, 20, 24 and 28 appeared to be more abundant in the upper oxycline and upper core of the OMZ at all 3 sampling sites. Other OTUs appear to show a similar trend but were not observed at all sites. Cluster 21 was significantly more abundant in the upper core of the OMZ than at depth but was only observed at site 2 (Figure 5.16a). Cluster 27 also demonstrated this trend but was only present in very low abundances at site 4 (Figure 5.16b) and site 5 (Figure 5.16c). Cluster 27 did however demonstrate a large frequency of reads at S5-719 m where it contributed to 34.8 ± 29.7% of the total reads. A number of OTUs appeared to be evenly distributed with depth but were absent in the lower half of the upper oxycline (*i.e.* at 60 m at site 4 and 67 m at site 5). This region of the depth profile was not sampled at site 2. Clusters 5, 11, 25, 26 and 30 all appeared to demonstrate this trend. Clusters 5, 11

and 25 appeared to demonstrate a greater abundance with depth (although this OTU was not absent in shallower depths) however this was only observed at sites 2 and 4. This trend was also observed in Cluster 11 at site 2 however the standard error bars (±1 SE) of the 100 m and 750 m depths are seen to overlap.


a)
Site 2
Mean Read Frequency (%)
45
40
35
30
25
20
15
10
5
0
100m
200m
540m
750m
1
3
4
5
8
10
11
13
14
15
20
21
24
25
26
27
28
29
30
Cluster


b)
Site 4
50m
60m
250m
436m
700m
860m
Mean Read Frequency (%)
0
5
10
15
20
25
30
35
40
1
3
4
5
8
10
11
13
14
15
20
21
24
25
26
27
28
29
30
Cluster

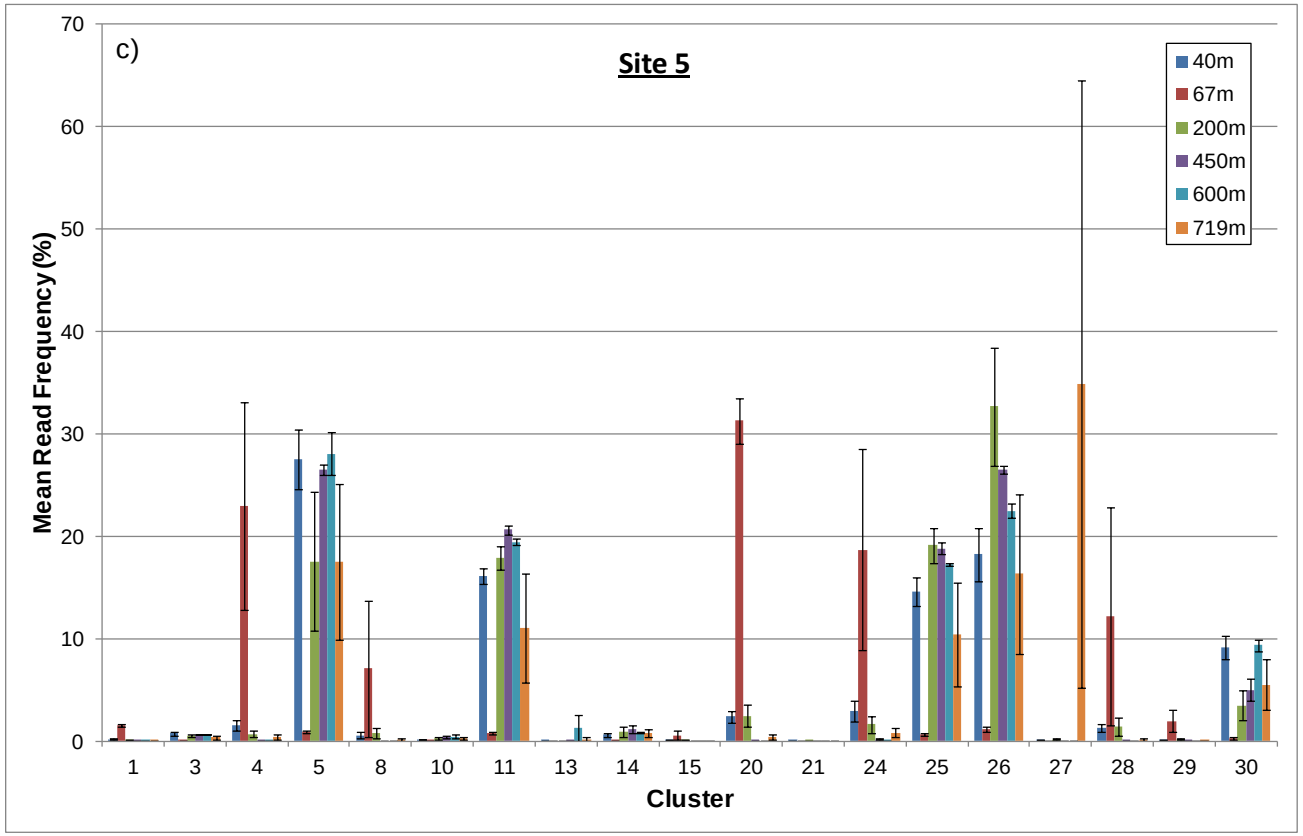


Figure 5.16: Mean frequency distributions of ETNP clusters. Frequencies are represented as the mean percentage of reads assigned to each cluster per sample for a) Site 2, b) Site 4 and c) Site 5. All samples were sequenced in triplicate. Error bars represent ± 1 SE Only clusters representing >0.01% of the total number of reads (222,565) are included in these figures.

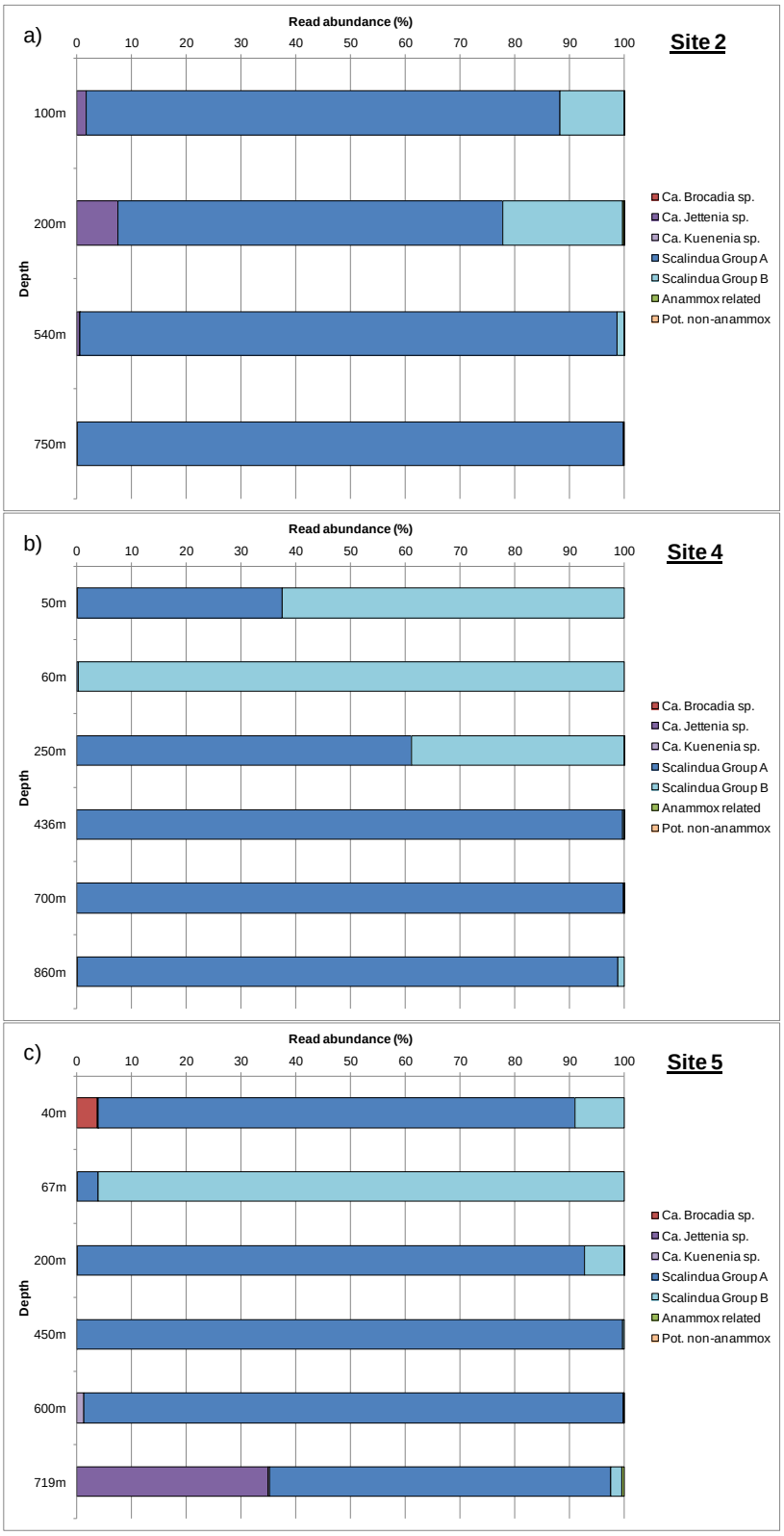


Figure 5.17: Mean ETNP read abundance per sample, expressed as a percentage. Graphs are shown for a) Site 2, b) Site 4 and c) Site 5. Mean values for each OTU were calculated from triplicate samples. The mean abundance of OTUs which had been assigned to specific taxonomic groups (see Figure **5.13**) were combined as mean values and are presented in these figures.

Mean read abundances for sites 2, 4 and 5, for each taxonomic group (as defined above in section 5.4.3.3) are presented in Figure 5.17. At all three sites, reads associated with 'Scalindua Group A' are found at all depths (although only 0.3% of the reads in S4-60 m are attributed to this group) and largely represent the majority of reads obtained from each sample. Reads associated with 'Scalindua Group B' were more frequently observed in the upper oxycline or upper core of the OMZ and was the dominant taxa in the lower half of the oxycline (*i.e.* S4 60 m and S5-67 m, Figure 5.17b and c respectively). At all three sites 'Scalindua Group B' was almost absent in the deeper core of the OMZ (<0.4% of reads except at S2-540 m where this group contributed to 1.3% of the total reads) but was present in the bottom oxycline at a slightly greater abundance (1.2% of reads in S4-860 m and 2.0% of reads in S5-719 m).

All sampling sites and depths were dominated by OTUs representing *Ca.* Scalindua spp. although non-anammox reads were also obtained from 454 pyrosequencing data. Of these only two OTUs (Cluster 13 representing *Ca.* Kuenenia spp. and Cluster 27 representing *Ca.* Jettenia spp.) contributed more than 0.1% of the total number of reads. Of these Cluster 27 was the most abundant. Both OTUs were almost absent in samples collected from site 4 (in these samples where these OTUs were present they contributed to less than 0.1% or the reads in each sample). The highest abundance of Cluster 13 was observed in S5-600 m (1.3 ± 1.3%) though this OTU was only observed in a single replicate sample at this depth. Cluster 27 was typically more abundant in these data, contributing 1.7 ± 0.7% of the reads in S2-100 m, 7.0 ± 1.9% in S2-200m and 34.8 ± 29.7% of reads in S5-719 m. In addition a further non-anammox cluster, Cluster 16 (representing *Ca.* Brocadia spp.) appeared to contribute to the diversity of S5-40 m (3.6 ± 3.6% of reads) though this was not significant and this OTU was only observed in one triplicate from this sampling depth and only appeared once within the data. The presence of non-Scalindua anammox OTUs did not appear to adhere to any observable trend in the data and appeared to be sporadically distributed across the dataset.

CCA and Mantel tests were calculated for these data and the environmental data associated with these sampling depths (see Figure 5.9, Figure 5.10 and Figure 5.11). These data failed to indicate a significant trend between anammox diversity across

these depth profiles related to NO_2^-, NO_3^-, NH_3, O_2 or PO_4^{3-}. Unconstrained CCA plots for anammox diversity data are presented in Figure 5.18. Figure 5.18a shows a CCA plot for all OTUs. In this plot the majority of the diversity in these samples is represented along the *x*-axis (CA1) though the data is skewed by a large diversity as represented by the *y*-axis (CA2) which separates sample S5-719 m and S2-200 m from the rest of the samples, presumably due to the near exclusive presence of Cluster 13 and Cluster 27 in these sites. Omitting non-Scalindua anammox OTUs (Figure 5.18b) aids the interpretation of this analysis and indicates that samples appeared to cluster into three groups of similar diversity; upper oxycline samples (blue), upper core OMZ samples (red) and deeper core and bottom oxycline (green). Samples fit into these three grouping perfectly (based on the depth which they were obtained) except for S5-40 m (upper half of the upper oxycline) which clusters with the deeper samples (green). These deeper samples (green) are more closely correlated, indicating a lower degree of diversity than the upper oxycline (blue) and upper core OMZ (red) samples which show greater diversity amongst these sampling depths.

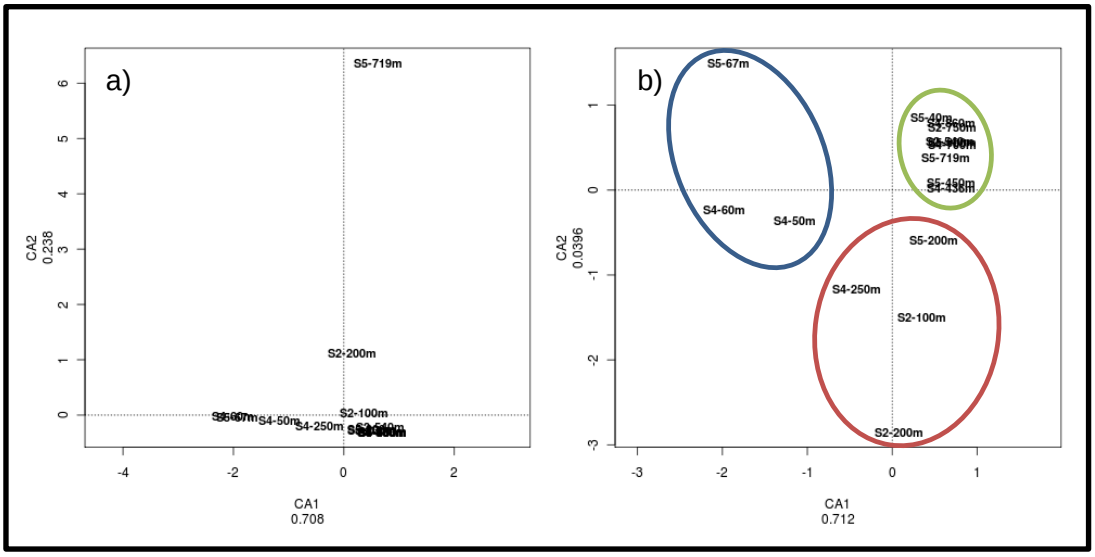


Figure 5.18: Unconstrained CCA plots created from ETNP 454 Pyrosequencing data for Sites 2, 4 and 5. CCA plots are shown for a) all observed anammox genera and b) only OTUs representing the Ca. Scalindua genus. Data for sampling sites and depths are calculated using mean values from biological triplicate samples. The presence or absence of non-Scalindua genera appeared to have the largest effect on sample similarity as these OTUs, though generally of low abundance, were only observed at a couple of locations. Omitting these OTUs show 3 clusters which appear to correspond to shallow depth within the oxycline (circled blue), upper core of the OMZ (red) and lower core to bottom of the OMZ (green) with the exception of S5-40m which clusters with the deeper OMZ sites.

5.4.4. *Ca.* Scalindua spp. Diversity

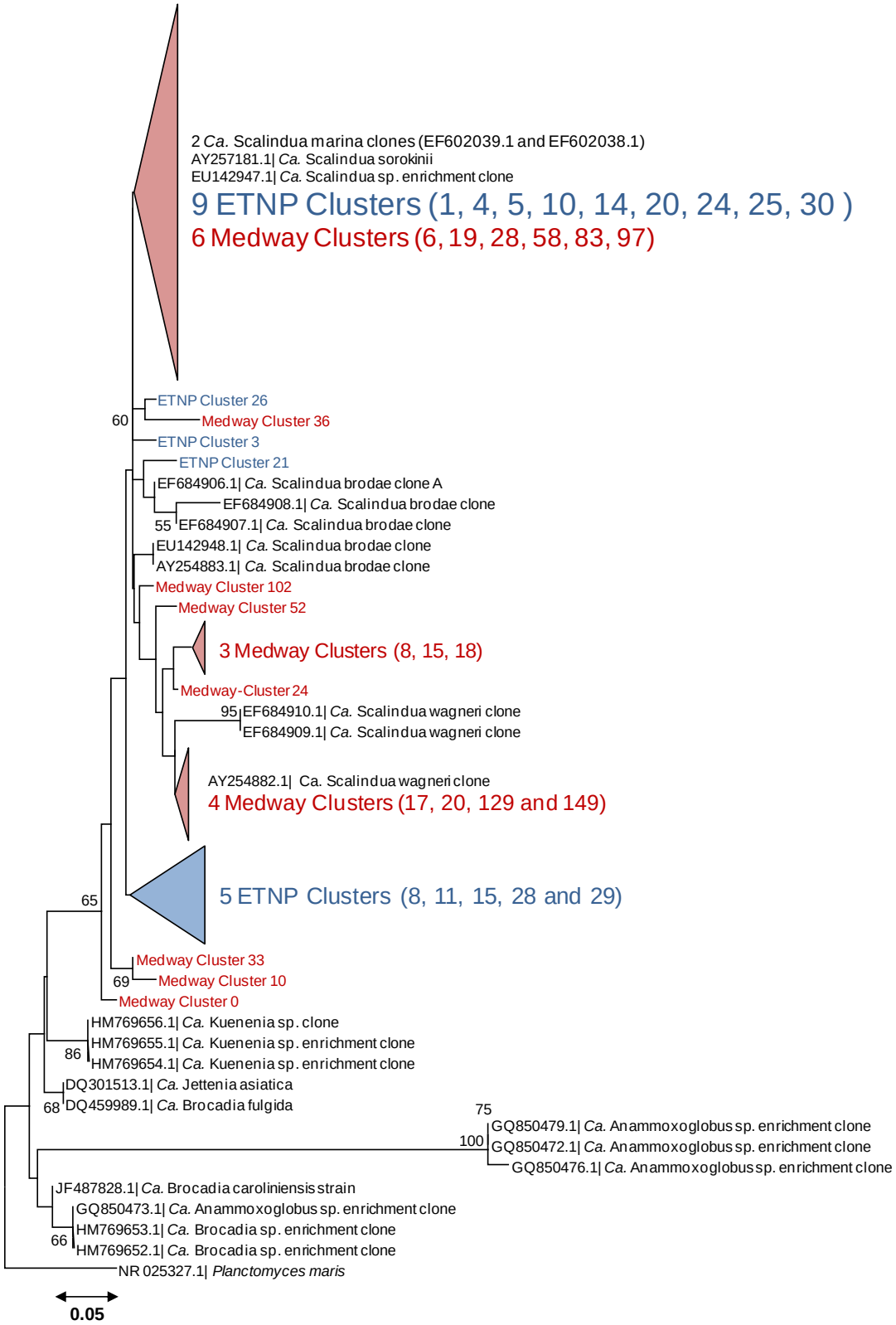


Figure 5.19: Phylogenetic tree showing relationship between Ca. Scalindua spp. OTUs from Medway and ETNP 454 pyrosequencing data. A bootstrap analysis involving 1000 repetitions was used as a test of phylogenetic analysis. Only branches showing bootstrap values greater than 50% are labelled. OTUs were defined at a similarity cut-off of 95% for both the Medway and ETNP datasets. Only OTUs representing greater than 0.1% of the total number of reads from each set and which had been positively identified as being related to the Scalindua genus were included in this phylogenetic analysis. Most OTUs from both sampling sites appeared to cluster together. However, two clear clusters of OTUs, related to Ca. Scalindua wagneri sequences, can be seen which contain only OTUs obtained from the Medway Estuary. A further Scalindua related cluster was also observed containing only OTUs from the ETNP OMZ.

OTUs obtained from Medway and ETNP 16S rRNA pyrosequencing data, which had been identified as being related to the Scalindua genus, were analysed in order to investigate the diversity of Scalindua organisms across these two environments. Only OTUs representing greater than 0.1% of the total number of reads from their respective dataset were included in this analysis. Phylogenetic relationships between these OTUs and reference sequences can be seen in Figure 5.19.

Phylogenetic analysis of *Ca.* Scalindua OTUs (Figure 5.19) indicates that a number of OTUs were present in both the Medway Estuary and ETNP OMZ. These OTUs primarily belonged to the '*Ca.* Scalindua brodae/marina/sorokinii Group' as defined from Medway data (Figure 5.4) and 'Scalindua Group A' and 'Scalindua Group B' from ETNP OMZ data (Figure 5.13). However, there was no clear distinction from this phylogenetic analysis between ETNP 'Scalindua Group A' and 'Scalindua Group B' as is seen when analysing only OTUs obtained from ETNP data (Figure 5.13). This may be due to an increased number of sequences improving the strength of phylogenetic relationships in this analysis compared with the analysis for only ETNP OTUs, perhaps suggesting that the designation of ETNP 'Scalindua Group A' and 'Scalindua Group B' is erroneous and due to a lack of data rather than a true phylogenetic phenomenon.

However, phylogenetic analysis of Scalindua-only OTUs also indicated the presence of distinct clusters which were only present in either the ETNP OMZ or Medway Estuary. Two OTU clusters, both associated with *Ca.* Scalindua wagneri 16S rRNA sequences, were observed containing solely OTUs obtained from Medway data. These OTUs represented ~8.9% of the total number of reads from the Medway dataset (see section 5.4.2.4).

In addition a further cluster was observed which contained only OTUs from ETNP pyrosequencing data. However, this cluster demonstrated limited similarity to any of the *Ca.* Scalindua spp. 16S rRNA sequences though it did cluster within the Scalindua clade. The OTUs present in this cluster were seen to cluster within 'Scalindua Group B' in Figure 5.13 (although branch lengths of these sequences were longer than their neighbouring sequences) and their unique clustering in the analysis presented in this section may be due to an improvement in the sequence database with the inclusion of both Medway and ETNP OTUs.

5.5. Discussion

5.5.1. Quality of 454 Pyrosequencing Data

A major and repeating issue in the use of 454 pyrosequencing technologies, and data obtained via this method, for microbial ecology is the quality of the data obtained and its reproducibility (Huse, *et al.*, 2008; Kunin, *et al.*, 2010). As such, it is important to stringently control and analyse the quality of 454 pyrosequencing data to provide validity to one's conclusions. This is especially true when utilising novel methods for analysing these types of data (as in this investigation). However, this does not negate the use of stringent controls when using more widely used methods *e.g.* QIIME (Caporaso, *et al.*, 2010) and Pyronoise (Quince, *et al.*, 2009).

Initial observations into the quality of the data presented in this thesis indicated that, in general, the quality of obtained reads was high. A large number of reads were produced for each dataset; 179.5K reads for the Medway Estuary (~8.5K per triplicate sample) and 264K reads for the ETNP (~5.5K per triplicate sample). These data would provide a sufficient depth of study for investigating the ecology of anammox bacteria using 16S rRNA primers Amx368F & Amx820R based on previous, preliminary investigations (~4K reads per sample to report the entire relative diversity based on rarefaction analyses based on the results from trial 454 pyrosequencing runs using these samples). Of these a suitable proportion of reads (47% and 85% for the Medway Estuary and ETNP OMZ datasets respectively) contained both the forward and reverse primer motifs. In the analysis of previously obtained data (not shown), a large number of reads which did not contain both of the primer regions was found to be indicative of a generally poor quality and error-prone

sequencing run. The large number of reads from the Medway dataset which did not contain the primer sequences would usually be of concern, however, even taking this into account, the number of 'good' reads obtained from these data would still be more than sufficient to assess the aims of this investigation (*i.e.* they would provide suitable sequencing coverage). The difference between the two sequencing runs analysed here is that the ETNP dataset contained a much larger number of samples than that of the Medway Estuary and so 454 pyrosequencing was performed using an entire half of a plate. Potentially, as the Medway data would have been obtained on a plate containing samples from other research projects using other primers, this may have caused interference during sequencing, resulting in a lower quality of data. Controlling this (by using the entire half plate) appeared to ameliorate this problem. In future research projects it is highly recommended to run samples on at least half a plate, even at greater financial outlay (though the steadily reducing cost of sequencing should offset this), in order to improve the quality of raw data.

During this project a novel approach (ReLIC) was used in aiding the quality control of these data, namely the investigation of minimum and maximum length cut-off values and R_{inc} ratios from read length data (see section 5.3 for full discussion). The data produced from this quality control step was of a good quality (see section 5.4). However, a small number of OTUs were encountered which clustered outside of the known anammox diversity and therefore the possibility of poor quality sequences affecting clustering results cannot be discounted. Therefore, this method may have been too liberal in terms of the reads which passed the quality control stages, however ReLIC does appear to be a reasonable starting position from which to research the use of this method further. Regardless, this method did produce data of a good standard (see subsequent sections) suggesting that the use of read lengths to determine good and poor quality reads may be an effective method. As can be seen in Figure 5.2 and Figure 5.12 a clear distinction between high and lower R_{inc} values at shorter read lengths could be determined for both datasets however the distinction at longer read lengths was not clear. This may be a potential source of poor quality sequences. As such using these exact cut-offs may not be the best method, but instead they could be used to determine significance levels about the median values (using mathematical modelling) to obtain an improved cut-off where the researcher is more confident that poor quality sequences are being excluded. Regardless of

future improvements which could be made to ReLIC, it appears to be a reasonable method, at least with these datasets of relatively low diversity, to remove the majority of poor quality reads from the data. Furthermore, this method removes arbitrary decision making from such processes to the improvement of the statistical robustness and reproducibility of the quality control of high throughput sequencing.

Pyrosequencing positive controls were conducted by sequencing a cloned anammox 16S rRNA gene of known identity (*Ca.* Brocadia sp.) in triplicate (section 5.4.1.1). These data produced only two OTUs at a clustering similarity cut-off of 95%, one of which comprised of only one read (total number of reads=12,306) and so is likely to be an erroneous OTU arising from a read with significant sequencing errors. Manual investigation of this read, in comparison with the other OTU, highlighted the presence of some potentially chimeric regions and the presence of indels associated with homopolymers within this singleton read, suggesting that this OTU had indeed arisen from a poor quality read within the dataset. The main OTU demonstrated a high similarity to *Ca.* Brocadia spp. (greater than other anammox sequences) suggesting that these data were accurate.

Technical replicates were also conducted in order to test the precision of these data. A summary of these data is shown in Figure 5.1 (see section 5.4.1.2). All three triplicates produced the same number of OTUs (n=8) at a cut-off of 95%. Furthermore, the standard errors associated with each OTU (error bars on Figure 5.1) were all small, showing that the variation between samples, in regards to the relative abundance of each OTU, was also small. As such, similarly precise data could be expected for the other samples in the dataset and any large errors are likely to be as a result of true biological variation amongst biological replicates, not poor replication from 454 pyrosequencing technologies.

5.5.2. Anammox Diversity Across Environmental Gradients

5.5.2.1. Medway Estuary

5.5.2.1.1. Data Quality

Medway Estuary data produced a large number of OTUs (294 OTUs) at a clustering cut-off of 95% similarity, however only 49 OTUs represented a sizeable proportion

of the total number of reads (>0.1%). As discussed previously, the aim of this investigation was to measure the relative diversity of anammox organisms and so low abundance OTUs were omitted in the knowledge that such OTUs may be ecologically significant but would not affect measurements of diversity providing such omissions were universally implemented. The ecological significance of such organisms is an important question but one that would require a more thorough knowledge of the diversity, distribution and environmental controls of anammox bacteria, the elucidation of which was the aim of this study. Rarefaction analyses for these data (Figure 5.3) indicate that, even with the omission of these low abundance OTUs, relative measurements of diversity are likely to be unaffected and that the entire anammox diversity was represented in these data (*i.e.* the sequencing coverage was sufficient to not omit any significant OTUs).

5.5.2.1.2. Phylogeny and Richness

The richness of the anammox community (Table 5.2) was greater in the middle of the sampled range (~30 OTUs) than at either end (~20 OTUs). This may represent a more specialist community at either end of the estuary (*i.e.* at the extremes of the salinity gradient which ranged from approximately 0 ppt at M1 to 25 ppt at M7) whereas the community in the middle of the estuary was more diverse, perhaps representing the changing state of these environments and periodic salinity intrusions. Twelve year averages for salinity data along the Medway Estuary indicate that the salinity at Aylesford (near site M1) remained low (0.54 ± 0.24 ppt, yearly average) throughout the year. Data for sites along the estuary however demonstrated a large salinity range of approximately 10 ppt over the year; Wouldham (near M4) 3.18 ± 1.83 ppt (December) to 13.25 ± 1.83 ppt (August) and M2 Bridge (near M6) 8.56 ± 2.61 (December) to 19.96 ± 1.06 ppt (September). However, data for Arethusa station (near M7) fluctuated between 18.79 ± 2.62 ppt (January) to 28.60 ± 0.67 (August), indicating that a similar annual salinity range (*i.e.* ~10 ppt) was observed at this end of the estuary as that observed over the middle of the estuary, despite M7 demonstrating a low degree of richness. This may suggest that the limiting salinity concentration affecting niche adaptation of anammox bacteria was approximately 20 ppt, *i.e.* above this salinity a specific anammox taxon or clade has adapted to specialise in such salinity ranges, explaining lower measurements of richness.

Alternatively it may suggest that anammox bacteria may not be affected by salinity, but instead are affected by a secondary variable associated with salinity and/or distance along the estuary (*e.g.* organic loadings) which in turn affects anammox specialisation. Indeed Kartal, *et al.* (2006) found that salinity did not affect the activity of the "freshwater" *Ca.* K. stuttgartiensis suggesting this may be the case.

Phylogenetic analysis of OTUs representing greater than 0.1% of the total number of reads (Figure 5.4) shows that OTUs were obtained from these data which represented all of the known anammox genera except *Ca.* Anammoxoglobus. The *Ca.* Anammoxoglobus sequences used in this analysis were obtained from enrichment cultures from a bioreactor (Viancelli, *et al.*, 2011) and so the failure to detect organisms related to such sequences in the natural environment might not be surprising. Indeed it is from such environments that *Ca.* Anammoxoglobus spp. were first identified (Kartal, *et al.*, 2007b) and no study in the literature has detected sequences in the natural environment which demonstrate a high similarity to this genus. More OTUs were related to *Ca.* Scalindua organisms than non-Scalindua bacteria perhaps suggesting that the diversity of *Ca.* Scalindua bacteria was greater along the estuary. Nine OTUs demonstrated a lower degree similarity to any known anammox bacteria ('Potentially non-anammox' group in Figure 5.4) and clustered outside of the anammox clade. These OTUs were typically associated with large standard errors (Figure 5.5d).

5.5.2.1.3. Diversity and Distribution

A clear change in anammox diversity and community structure was observed along the estuary in the data presented in Figure 5.6. Anammox bacteria have been found to have only one copy of the 16S rRNA gene (Strous, *et al.*, 2006) and so, although 454 pyrosequencing data cannot be used to determine absolute organism abundances, it can be used as a proxy for the relative abundance of anammox bacteria[*]. The relative abundance of non-Scalindua anammox bacteria is greatest at M1 (the most upstream site) where these organisms dominate the anammox community (98.64% of reads). Non-Scalindua anammox organisms also dominated M2 contributing to 90.32% of the total number of reads. The abundance of non-Scalindua anammox

[*] As shown in chapter 3, the primers used in this study did not show any bias towards different anammox genera.

organisms decreased downstream contributing 40.04% of the total reads at M3, 49.95% at M4 and 10.10% at M5 and were almost absent at sites M6 and M7 (0.43% and 0.08% respectively).

Of these, reads associated with *Ca.* Brocadia spp. were most abundant with 76.26% of reads associated with 'Brocadia A' and 21.94% of reads with 'Brocadia B'. This was true for all seven sites, where 'Brocadia A' and 'Brocadia B' represented the vast majority of the non-Scalindua anammox reads. These two groups contributed to >99% of the reads associated with non-Scalindua organisms in all sites except M5 where they contributed to 86.23% of reads (the rest being made up of reads associated with *Ca.* Kuenenia spp.). These findings agree with those of Hirsch, *et al.* (2011) where *Ca.* Brocadia spp. were found to be the most dominant organisms in freshwater river sediments. This observation is also consistent with the original hypotheses outlined in section 1.3.2.1.2.1. As shown in chapter 3, these primers (Amx368F & Amx820R) demonstrated the ability to target all the known anammox genera and did not appear to be biased towards particular organisms and so this dominance of *Ca.* Brocadia spp. in terms of the diversity of non-Scalindua anammox organisms is likely to be an accurate representation of the anammox community along the Medway Estuary.

Ca. Scalindua organisms demonstrated the opposite trend to non-Scalindua anammox organisms, decreasing in abundance upstream which is consistent with the hypotheses outlined in section 1.3.2.1.2.1. *Ca.* Scalindua organisms dominated the anammox community at site M5 (83.03% of reads) and decreased in abundance (to be replaced as the dominant community members by non-Scalindua anammox organisms) towards the freshwater end of the estuary, contributing to 47.69% of the total reads at M4, 53.31% at M3, 8.90% at M2 and 1.08% at M1. *Ca.* Scalindua organisms appeared to be less abundant at sites M6 and M7 (60.81% and 33.28% of reads respectively). However, the remainder of reads at these sites were not made up of non-Scalindua anammox bacteria, but reads which could only be attributed to 'anammox related' or 'potentially non-anammox' organisms (Figure 5.6) which demonstrate a low degree of similarity to known anammox organisms and hence may represent reads containing sequencing errors which could be associated with the large standard errors associated with these data (see Figure 5.5). However, the

presence of a large proportion of sequences belonging to OTUs classed as anammox 'related' or 'potentially non-anammox' may be indicative of a low copy number of anammox 16S rRNA genes at this site. Lower copy numbers of target genes could potentially produce proportionally fewer reads associated with anammox bacteria and an increase in the number of reads associated with chimeric or-non target amplified sequences gained from high-throughput sequencing technologies. However, if this claim were true, rarefaction curves for these samples may be expected to rise more sharply (see Figure 5.3) and no issues were encountered in the laboratory (*e.g.* lower PCR amplification efficiencies) which may suggest an unusually low copy number of anammox 16S rRNA genes at this site. Alternatively, these OTUs may represent a novel anammox phylogeny however, as previously stated, extreme caution must be used when attempting to assign novel taxa from such data, especially without the aid of other measurements of ecological diversity (*e.g.* a functional gene conforming to the phylogenetic consensus). Omission of such OTUs would reveal that M6 and M7 are dominated by *Ca.* Scalindua anammox organisms (>99% of reads) which would fit with the trend observed in the rest of the data.

The majority of OTUs (11 OTUs) associated with *Ca.* Scalindua organisms clustered with reference sequences from *Ca.* S. brodae, *Ca.* S. marina and *Ca.* S. sorokinii 16S rRNA genes (Figure 5.4). This OTU cluster also represented the most abundant group of *Ca.* Scalindua organisms at all but site M1 (Figure 5.6), where the 'Scalindua related' cluster represented the largest number of reads (53.20% of *Ca.* Scalindua reads). However, as shown previously, *Ca.* Scalindua organisms only represented 1.08% of the total number of reads at this site. Despite the '*Ca.* Scalindua brodae/marina/sorokinii' cluster representing the majority of *Ca.* Scalindua spp. reads at these sites, an interesting trend was nevertheless observed concerning the '*Ca.* S. wagneri' cluster. Although in general, *Ca.* Scalindua organisms appeared to demonstrate a decrease in abundance upstream, *Ca.* S. wagneri appeared to be most abundant at M5 (34.95% of *Ca.* Scalindua reads, 29.02% of total reads) and decreased in abundance both downstream and upstream of this site. The lowest abundance of *Ca.* S. wagneri related organisms was observed at M7 (0.14% of *Ca.* Scalindua reads and 0.05% of total reads at this site). The abundance of these organisms at M1 was also low representing 0.20% of the total number of reads (the contribution to the *Ca.* Scalindua community was larger,

18.54%, however the low number of reads associated with these organisms at M1 has likely produced misleading data in regards to this). The most abundant OTU representing *Ca.* S. wagneri organisms, Cluster 8, demonstrates this trend (Figure 5.5b) with M5 having a substantially higher frequency of reads assigned to *Ca.* S. wagneri than M4 and M6 which in turn contain more reads than M3 and M7. The error bars of all of these mean read frequencies from Cluster 8 (±1 SE) did not overlap. The other five *Ca.* S. wagneri OTUs, (Clusters 15, 20, 102, 129 and 149, Table 5.1) did not appear to demonstrate this trend (Figure 5.5b) though all these OTUs represented <5% of the total number of reads in data from these sites and therefore did not substantially affect the abundance of these organisms and their trend in distribution across the sites. As can be seen in Figure 5.6, *Ca.* S. wagneri organisms appear to be most abundant between salinities of 4-10 ppt. This may suggest that these organisms are adapted to mesohaline conditions (though the data present are not sufficient to define the exact range of these organisms) whereas other *Ca.* Scalindua organisms are abundant in brackish to saline environments. The existence of such an anammox community, adapted to mesohaline conditions was hypothesised previously (see section 1.3.2.1.2.1).

Unconstrained CCA analysis of these data (Figure 5.7) indicates that the greatest difference in anammox diversity (represented by CA1 along the *x*-axis) was observed along the estuary sequentially from M1 to M7. This trend was perfectly linear except for M3 and M4 (the orders of which is reversed, see Figure 5.7) however these sites are very similar (as demonstrated by the closeness of these points in Figure 5.7) and therefore do not detract from the validity of this trend. A trend in increasing anammox diversity was also observed towards the middle of the estuary (represented by CA2 along the *y*-axis, with M5 demonstrating the greatest diversity), with low diversity observed at either extreme of the sampled range (M1 and M7). These two axes (CA1 and CA2) represented 100% of the observed diversity (68% and 32% respectively). Dale, *et al.* (2009) reported lower diversity upstream of the Cape Fear River Estuary, correlating with the findings of this investigation, but reported a greater diversity with increasing salinity. CCA analysis conforms with the conclusions drawn from the rest of the data analysis (see above) and indicate that the greatest diversity was observed at M5, which corresponds with

the site where *Ca.* S. wagneri, a potentially mesohaline adapted anammox organism, was found to be most abundant.

5.5.2.1.4. Environmental Controls and Significance

Mantel tests were conducted on interpolated/extrapolated environmental data (Environment Agency, UK) and plotted on CCA plots to test for relationships between diversity measurements and environmental conditions. Figure 5.7 shows the effects of salinity, conductivity, NH_3, NO_2^-, NO_3^-, PO_4^{3-}, biological oxygen demand (BOD), dissolved O_2, temperature and pH on anammox diversity. The environmental factors showing the strongest correlation to diversity appear to be salinity and conductivity (which are related environmental factors) and demonstrate a positive correlation with the change in diversity observed from M1 to M7. A strong negative correlation is also observed with NO_3^- and PO_4^{3-} between M1 and M7, with increasing concentrations of these anions towards the freshwater end of the estuary. The concentration of NH_3 and NO_2^- also demonstrate a negative trend (increasing towards freshwater end of the estuary) with the trend in anammox diversity between M1 and M7 though this was not as strong as that demonstrated by NO_3^- and PO_4^{3-}. The increase in diversity towards the middle of the estuary (*y*-axis on Figure 5.7) was strongly correlated with a decrease in pH and less strongly correlated with decreasing temperature and dissolved O_2, with higher values for these environmental factors observed to either extreme of the sampling range. Hamersley, *et al.* (2009) also reported an increase in anammox diversity in brackish environments but stated that the reasons for this were unclear.

No direct measurements of anammox activity in Medway Estuary sediments were available for the locations and dates sampled. However, anammox rate measurement data was available from a previous study along the estuary by Nicholls and Trimmer (2009). These data suggest that the potential for anammox activity was highest around site M6 (~10.93% of N_2 produced) and decreased from this point both upstream and downstream, though measured rates were marginally higher upstream (~7.48% of N_2 around M4) than downstream (~4.49 % of N_2 around M7). Nicholls and Trimmer (2009) presented data for anammox rates further downstream, beyond the extent of this investigation, reporting rates of 3.09% and 5.39% (at the most

downstream site) of N_2 production. No anammox rate measurements were available for the most freshwater sites namely M1, M2 and M3.

However, these data must be used with caution as they were not taken at the same time as samples collected during this investigation nor where measurements taken at the exact same site locations as this study. Furthermore, Nicholls and Trimmer (2009) did not present error data associated with these measurements and the low number of data points makes interpolation (within the context of the data presented in this thesis) impossible and so no firm conclusions could be drawn from comparison with these data.

The data presented in this study clearly demonstrate a trend in anammox diversity and community structure along the Medway Estuary. The diversity of anammox organisms appeared to be strongly correlated with the increase of salinity along the estuary which corresponds with the findings of other investigations into anammox diversity along estuarine environments (Dale, *et al.*, 2009; Hirsch, *et al.*, 2011). If salinity is indeed the major controlling factor in anammox diversity and distribution, this may suggest a large degree of seasonal variation in the anammox community with annual salinity fluctuations (Dale, *et al.*, 2009). Freshwater site M1 was dominated by non-Scalindua anammox organisms. The abundance of non-Scalindua organisms decreased through M2, M3 and M4 (oligohaline) and M5 (mesohaline) and they were almost absent at M6 and M7 (very brackish to saline). The abundance of *Ca.* Scalindua spp. was greatest at the saline end of the estuary (M5 and M6) and lowest at the freshwater end (M1). Thus the data suggest that non-Scalindua organisms are adapted to freshwater/low salinity environments whereas *Ca.* Scalindua spp. are adapted to more high salinity environments. The diversity and richness of the anammox community was greatest towards the middle of the estuary which represents a community comprising of both *Ca.* Scalindua and non-Scalindua anammox bacteria. *Ca.* Scalindua wagneri organisms were most abundant toward the middle of the estuary, perhaps suggesting that these organisms are specifically adapted to mesohaline environments.

However, as stated previously, investigations on enrichments of anammox organisms have shown that these organisms can survive and are active at a range of salinities (Kartal, *et al.*, 2006). Therefore, the community change observed in the data

obtained from this study may not represent anammox organisms adapted to different salinities but rather to a secondary variable (such as the availability of DOM) also correlated with distance along the estuary as outlined in the original hypotheses (section 1.3.2.1.2). However these data are not sufficient to confirm this hypothesis and further research would be required to investigate this, including a more thorough investigation into the environmental conditions experienced along the Medway Estuary including measurements taken at the time of sampling. Nevertheless, such relationships can only be inferred from *in-situ* data and further experimentation would be required to confirm that variations of such environmental factors had a direct effect on the diversity of anammox bacteria. Caution must however be used when comparing the ecology of *in-situ* organisms with those grown *in-vitro*.

One hypothesis, developed during this study, as to the causes of anammox diversity and community separation in estuarine environments, might be the ability of anammox organisms to utilise organic substrates as a source of carbon and/or nitrogen. Other data obtained during this study (see chapter 6) suggest that non-Scalindua organisms may be able to metabolise organic substrates whereas *Ca.* Scalindua spp. cannot. Hence, these data might suggest a change in anammox metabolism along the estuary, from classical anammox in saline environments to 'Organammox' in freshwater environments. However this hypothesis cannot be asserted with the data available and further research, including the measurement of anammox activity along the estuary with the addition of organic substrates, would be required to test this hypothesis. Alternatively the concentration and bioavailability of DIN and other nutrients along the estuary may also affect anammox community change (as shown if Figure 5.7, DIN and PO_4^{3-} were seen to increase upstream and were correlated to anammox community structure) however there is no evidence in the literature that indicates that anammox organisms demonstrate a preference for low or high concentration of nutrients. There are however suggestions in the literature that anammox diversity may be strongly influenced by anthropogenic activity (Dale, *et al.*, 2009; Li, *et al.*, 2011; Wang and Gu, 2013). Sato, *et al.* (2012) further found that paddy field environments, which are typified by high nitrogen loadings from manure and fertilisers, were dominated by *Ca.* Brocadia spp., although paddy fields are also freshwater environments which may account for the presence of these organisms.

5.5.2.2. ETNP OMZ

5.5.2.2.1. Data Quality

A total of 44 OTUs were defined from ETNP OMZ data (n=222,565 reads after quality control). Of these, 25 OTUs did not represent greater than 0.1% of the total number of reads. However, in this instance, these low abundance clusters were included in order to investigate the validity of the assumption that such low abundance clusters are indicative of poor quality data and are unlikely to affect conclusions drawn from the data. Phylogenetic trees were constructed both including and omitting these low abundance OTUs (see section 5.4.3.3 and Figure 5.13 and Figure 5.14) and show that the phylogeny expressed by these two analyses was congruent and thus the inclusion of these low abundance OTUs did not affect the inferred phylogenies of the more abundant OTUs. Figure 5.15 demonstrates that in the majority of samples, rarefaction curves were plateauing, indicating that the sampling depth was sufficient to report the entire (relative) diversity of anammox bacteria in these samples. Those samples which were not observed to completely plateau were nevertheless all beginning to plateau, and so it can be assumed that no significant OTUs were missing from this dataset.

5.5.2.2.2. Phylogeny and Richness

Phylogenetic analysis of these OTUs (Figure 5.13) revealed seven clusters of OTUs (Figure 5.13). The majority of OTUs were shown to be related to *Ca.* Scalindua spp. and clustered into two clusters: 'Scalindua Group A' and 'Scalindua Group B'. 'Scalindua Group A' demonstrated no similarity to any known *Ca.* Scalindua 16S rRNA sequences (though this cluster definitely clustered within the *Ca.* Scalindua clade and outside the wider anammox phylogeny) though the majority of these OTUs (13/22 OTUs) represented <0.1% of the total number of reads. Conversely, only 1 OTU from 'Scalindua Group B' represented <0.1% of the total number of reads and showed a high similarity to sequences representing *Ca.* S. brodae, *Ca.* S. marina and *Ca.* S. sorokinii 16S rRNA genes. No OTUs were found which demonstrated any similarity to *Ca.* S. wagneri sequences which is consistent with previous studies (Woebken, *et al.*, 2008).

OTUs were also found to represent non-Scalindua anammox organisms namely *Ca.* Brocadia spp., *Ca.* Kuenenia spp. and *Ca.* Jettenia spp., however only 2 out of 7 OTUs (one showing a high sequence similarity to *Ca.* Jettenia spp. and the other to *Ca.* Kuenenia spp.) represented greater than 0.1% of the total number of reads from this dataset. Despite the general consensus within the literature that *Ca.* Scalindua spp. are solely marine anammox organisms and non-Scalindua anammox inhabit terrestrial and freshwater environments (Hamersley, *et al.*, 2009; Hirsch, *et al.*, 2011; Song and Tobias, 2011), non-Scalindua organisms, such as *Ca.* K. stuttgartiensis , have been detected in marine environments (Byrne, *et al.*, 2008; Ulloa, *et al.*, 2012). However Woebken, *et al.* (2008) found no evidence of non-Scalindua anammox organisms in the Peruvian OMZ. No OTUs from the ETNP demonstrated any similarity to *Ca.* Anammoxoglobus spp. These organisms were also not observed in the Medway Estuary (see section 5.5.2.1.3).

5.5.2.2.3. Distribution and Diversity

The vast majority of reads obtained from ETNP OMZ pyrosequencing data were associated with *Ca.* Scalindua spp. as hypothesised previously (see section 1.3.2.1.2.2). The *Ca.* Scalindua clade represented the majority of reads at all of the 16 sampling depths across the 3 sites (Figure 5.17). The lowest abundance of *Ca.* Scalindua reads was observed at S5-719 m (64.35%). In general *Ca.* Scalindua spp. reads contributed >99% of the total number of reads at each site except at S2-100 m (98.24%), S2-200 m (92.23%), S5-40 m (96.13%), S5-600 m (98.61%) and S5-719 m. In samples where *Ca.* Scalindua spp. did not contribute to 100% of the reads, the remainder were largely made up of non-Scalindua anammox organisms. 'Anammox related' and 'Potentially non-anammox' OTUs contributed very small numbers of reads to each site (<<0.1%) if they were present at all. The low abundance and large phylogenetic distances associated with these OTUs strongly indicates that they are a result of low quality sequences rather than true phylogenetic diversity and so were omitted from further analysis and discussion, except to highlight that the low number of erroneous reads and OTUs created from the dataset indicates that the methodology used in the analysis of this data (in terms of quality control) was good.

A trend was observed within the diversity of *Ca.* Scalindua spp. at all three sites (though the lower resolution of site 2 made this trend less pronounced[*]). 'Scalindua group B' organisms (*i.e.* those OTUs showing a high similarity to *Ca.* S. brodae, *Ca.* S. marina and *Ca.* S. sorokinii) were most abundant in the lower half of the upper oxycline (S4-60 m and S5-67 m). Organisms from the 'Scalindua Group B' clade were also observed in the upper half of the upper oxycline (S4-50 m and S5-40 m) and the upper core of the OMZ (S2-100 m, S2-200 m, S4-250 m and S5-200 m). These organisms were not detected in the lower core of the OMZ except at S2-540 m though this clade only represented 1.27% of the total number of reads and only two OTUs (Clusters 20 and 24) which were associated this clade showed a significant number of reads at this site. The deepest sampling depths at sites 4 and 5 (S4-860 m and S5-719 m), representing the lower oxycline below the core of the OMZ, also contained organisms from this clade though they only contributed 1.18% and 2.00% of the reads respectively. The deeper core of the OMZ at these three sites was dominated by organisms related to 'Scalindua Group A'.

Therefore, the data suggest a change in *Ca.* Scalindua diversity within the ETNP OMZ, with an increase in the abundance of 'Scalindua Group B' organisms (*Ca.* S. brodae, *Ca.* S. marina and *Ca.* S. sorokinii) in more oxygenated waters (*e.g.* the oxyxline and upper core of the OMZ, see Figure 5.8) whereas 'Scalindua Group A' organisms were dominant at the anoxic OMZ core. Dalsgaard, *et al.* (2012) reported a greater activity of anammox organisms in the upper core of the ETSP OMZ, compared with the lower core, which may be related to this community shift. Indeed, Woebken, *et al.* (2008) presented a case for different ecological niches for anammox bacteria in the OMZ but failed to speculate as to the causes of such niches. The data may also suggest that 'Scalindua Group B' begin to repopulate the anammox community in the lower oxycline however more sampling points at these depths would be required to confirm this observation as, as shown in Figure 5.8, Figure 5.9, Figure 5.10 and Figure 5.11, the sampling sites chosen during this investigation only just sampled the lower oxycline and failed to do so at site 2. These data therefore support the hypothesis outlined in section 1.3.2.1.2.2. These observations, considering they are observed at all three sites and that the change in abundance of most of these OTUs is significant (Figure 5.16), would suggest that 'Scalindua

[*] N.B. Only four sites were sampled.

Group A' does indeed represent a true clade within the *Ca.* Scalindua genus and is not as of a result of low quality sequencing reads (as suggested previously).

Although previous studies have found evidence of non-Scalindua anammox organisms being present in marine settings (Byrne, *et al.*, 2008; Ulloa, *et al.*, 2012), this study presents the first data to suggest significant numbers of these organisms (*i.e.* not just single clones from clone libraries) using an established method for reporting phylogenetic diversity (*i.e.* not the *hzo* gene, see chapter 4). This would confirm hypotheses that the diversity and distribution of anammox organisms is not solely restricted to the salinity of the environment (see section 1.3.2.1). Reads were found which represented *Ca.* Brocadia spp., *Ca.* Jettenia spp. and *Ca.* Kuenenia spp. Of these, *Ca.* Jettenia spp. was the most abundant and *Ca.* Kuenenia spp. the least abundant. However, unlike *Ca.* Scalindua spp., non-Scalindua anammox organisms did not appear to be ubiquitously dispersed throughout the OMZ but instead occupied one or two distinct sampling sites. In only five of the sixteen sites did non-Scalindua anammox organisms contribute to greater than 1% of the total reads from that site (Figure 5.17). Furthermore, there did not appear to be any obvious trend as to the presence or absence of non-Scalindua anammox organisms, though the highest abundances of these organisms were not found in the core of the OMZ (although they were present at these locations, *e.g.* S5-600 m).

CCA analyses (Figure 5.18) of these data indicated that the majority of anammox diversity was correlated to depth (CA1 representing approximately 70% of the observed diversity). However, this diversity was skewed by the presence of non-Scalindua anammox organisms (Figure 5.18a) as one would expect due to their large abundances but scarce distribution. Omission of these OTUs (Figure 5.18b) revealed three distinct clusters related to depth: the upper oxycline, the upper core OMZ and the lower core OMZ and edge of the lower oxycline. The greatest difference in diversity appeared to exist between the upper oxycline and deeper sites, presumably due to the effect of 'Scalindua Group B' which was most abundant in the upper oxycline.

5.5.2.2.4. Environmental Controls and Significance

Oxygen profiles and nutrient profiles can be seen in Figure 5.8, Figure 5.9, Figure 5.10 and Figure 5.11. The ETNP OMZ between 8°N and 13°N and appeared to be fairly constant throughout, exhibiting a standard OMZ depth profile (though the upper oxycline was very steep with the periphery of the OMZ beginning at a depth of approximately 50 m). The core of the OMZ appeared to be deeper towards the Central American coast (*i.e.* towards the North), presumably due to increased upwelling at the continental shelf boundary. Profile data for the ETNP OMZ (Figure 5.9, Figure 5.10 and Figure 5.11) show a typical OMZ depth profile (Paulmier and Ruiz-Pino, 2009) with low NO_2^- concentrations throughout except for a spike in the upper core of the OMZ. This NO_2^- peak is typically associated with high denitrification, corresponding to a decrease in NO_3^- concentrations and a lower value of N_{def} (as seen in Figure 5.9, Figure 5.10 and Figure 5.11). A pronounced N_{def} is indicative of a significant loss of fixed nitrogen (either from anammox or denitrification) from these regions (Devol, *et al.*, 2006; Chang, *et al.*, 2012; Nagel, *et al.*, 2013). However, Rush, *et al.* (2012) found this NO_2^- peak to correlate with an increase in anammox 16S rRNA copy numbers and ladderane lipids, implying either an increase in anammox activity and/or increase in bacterial abundance, perhaps indicative of an increase in both anammox and denitrifier activity at this depth. Although this study contains no quantitative measurements of anammox bacteria (except those inferred by 454 pyrosequencing read abundance), these observations by Rush, *et al.* (2012) would correspond to the change in *Ca.* Scalindua spp. diversity between upper oxycline and upper OMZ sites and core OMZ sites as discussed in section 5.5.2.2.4. NH_4^+ concentrations were low (< 2 μM) however technical difficulties were experienced in making these measurements and so these data may overestimate true concentrations. However these data were not too dissimilar to those reported by Rush, *et al.* (2012) and Chang, *et al.* (2012) from the ETNP OMZ.

However CCA analyses combined with mantel tests, conducted using these data, failed to demonstrate any significant trend between anammox diversity and these profile data and as such the hypothesis outlined in section 1.3.2.1.2 concerning the environmental factors controlling the diversity of anammox organisms cannot be

asserted with these data. The diversity of *Ca.* Scalindua spp. appears to be approximately correlated with O_2 (or at least between regions of oxygen deficient and anoxic regions) and perhaps higher resolution data (*e.g.* every 50 m) would be required to demonstrate a significant trend with these data. The reasons for the distribution of non-Scalindua anammox organisms are also unclear as no observable trend is evident (although the actual abundances of these reads are significant within the data and so are likely to represent real anammox diversity). Further and more extensive research would be required in order to investigate the ecological significance and environmental factors of the observed anammox diversity.

5.5.3. *Ca.* Scalindua spp. Diversity

454 pyrosequencing data was compared between Medway Estuary data and ETNP OMZ data in order to investigate the diversity of the genus *Ca.* Scalindua in different marine settings. As previously stated, *Ca.* Scalindua spp. have previously been shown to solely inhabit marine environments (Hamersley, *et al.*, 2009; Song and Tobias, 2011). However, to date no single study has investigated the diversity of these organisms in such different marine settings as estuaries and OMZs, nor has any study investigated these environments with the high resolution presented in the data of this study.

Anammox 16S rRNA gene sequences and OTUs obtained from these samples indicate that a number of organisms are shared between these two settings. OTUs obtained from both the Medway Estuary and ETNP OMZ data were shown to cluster together (Figure 5.19) and demonstrated a high degree of similarity to known 16S rRNA genes from *Ca.* S. brodae (Schmid, *et al.*, 2003; van de Vossenberg, *et al.*, 2008), *Ca.* S. marina (Brandsma, *et al.*, 2011) and *Ca.* S. sorokinii (Kuypers, *et al.*, 2003) which have been previously found in marine environments. *Ca.* Scalindua OTUs obtained from ETNP data appeared to produce a distinct cluster 'Scalindua Group A' (Figure 5.13), perhaps similar to the "novel subcluster" from the Eastern Tropical South Pacific (ETSP) OMZ as suggested by (Galan, *et al.*, 2009). However, the alignment of these OTU sequences with data from the Medway Estuary (Figure 5.19) effectively removed this independent cluster, clustering these OTUs, and those from the Medway Estuary, with *Ca.* S. marina and *Ca.* S. sorokinii 16S rRNA sequences. This may suggest that there may be a degree of phylogenetic diversity

between *Ca.* Scalindua spp. from the ETNP (and potentially other OMZs) and other *Ca.* Scalindua spp. (*i.e.* those from which the reference sequences have been obtained) however *Ca.* Scalindua spp. from estuarine environments may lie somewhere between the two.

A number of OTUs were however found which were unique to either dataset. Five OTUs obtained from the ETNP dataset were found to cluster independently of all of the other *Ca.* Scalindua spp. (including all reference sequences) and represented 19.57% of the total number of reads. It is unlikely that this cluster represents poor quality data as these 5 OTUs represent such a large proportion of reads and so must be treated as a true independent cluster. These OTUs were also found in most of the sampling sites and depths investigated during this study. However, whether this cluster represents a true novel taxa would require further investigation, not least as to the reproducibility and seasonality of this cluster (*i.e.* would it be encountered during further investigations of this region).

Eight OTUs obtained from Medway Estuary data appeared to cluster with reference sequences representing *Ca.* S. wagneri organisms whereas no ETNP OTUs demonstrated a high degree of similarity with these sequences. As such these data appear to suggest that *Ca.* S. wagneri is not found in the ETNP OMZ. *Ca.* S. wagneri was first observed in and enriched from a wastewater treatment plant (Schmid, *et al.*, 2003) but has since been detected in estuarine environments (Amano, *et al.*, 2007; Dale, *et al.*, 2009) and a stratified lake (Hamersley, *et al.*, 2009). However, no such study on anammox diversity in OMZ environments has reported to find this species. As such, *Ca.* S. wagneri appears to be associated with marine environments, but those found closer to terrestrial settings and not in the open ocean. Analysis of the distribution of this clade along the Medway Estuary (Figure 5.6) indicates that *Ca.* S. wagneri related OTUs were most abundant at site M5 (29.02% of reads), was present at low abundances at M1 and M2 (0.20% and 0.63% of reads respectively) and was almost absent (<0.05% of reads) at M7, the most saline of the seven sites. This might suggest that *Ca.* S. wagneri inhabits a mesohaline environmental niche ranging from salinities of ~4-10 ppt. However the elucidation of the role of this organism within this niche requires further investigation.

Nevertheless, these data indicate that the current view that *Ca.* Scalindua spp. are found in marine organisms and non-anammox organisms are not (Hirsch, *et al.*, 2011) may be overly simplistic and that the ecology of anammox organisms and diversity and distribution of the *Ca.* Scalindua clade is more complex than this. Although this has been suggested in previous studies (Hamersley, *et al.*, 2009), this study presents the first data of anammox diversity and distribution which samples two different environments using exactly the same methodology and thus provides the best comparison between these environments.

5.5.4. Comparison of Anammox Diversity in Estuarine and OMZ Environments

The data presented above demonstrate a depth of study previously unseen in investigations of the microbial ecology of anammox bacteria and provides new insights into the *in situ* diversity of these organisms. Furthermore, it is the first study to use such an approach to compare two different ecosystems: a temperate estuary and a tropical OMZ region.

The observed diversity was greater in the Medway Estuary than the ETNP OMZ, with more OTUs obtained from Medway Estuary. Rarefaction curves for each dataset were shown to be plateauing and therefore it is assumed that no significant members of the anammox community have been missed in these data. Furthermore, the 16S rRNA primers used for this study have been shown to target the entire known diversity of anammox bacteria[*]. The ETNP OMZ was dominated by *Ca.* Scalindua spp. whereas the Medway Estuary appeared to show equal abundances (over the entire sampled range) of *Ca.* Scalindua spp. and non-Scalindua anammox organisms (dominated by *Ca.* Brocadia spp.). Evidence for the presence of non-Scalindua anammox organisms were also observed in the ETNP, dominated by reads demonstrating the closest similarity to *Ca.* Jettenia spp. These observations contradict those of Hirsch, *et al.* (2011), who stated that marine diversity was greater than that in freshwater environments, though they used *hzo* gene sequencing (in conjunction with 16S rRNA methods) which have been shown to be unreliable markers of phylogenetic diversity (chapter 4). Hirsch, *et al.* (2011) did however

[*] Though obviously, the potentially for previously undiscovered anammox community members cannot be accounted for.

observe similar trends as to the distribution of anammox organisms along estuarine gradients, comparable with further observations in the literature (Dale, *et al.*, 2009). A similar trend in the anammox community, changing from a *Ca.* Brocadia spp. dominated to a *Ca.* Scalindua spp. dominated community with increasing salinity was also observed in a stratified lake environment (Hamersley, *et al.*, 2009).

Diversity was also observed within the *Ca.* Scalindua clade between these two environments. OTUs associated with *Ca.* S. wagneri were only observed in the Medway Estuary data. Furthermore these organisms were only observed along the middle stretch of the estuary (from ~4-10 ppt salinity) suggesting that these organisms may be adapted to mesohaline/brackish environments. A cluster of *Ca.* Scalindua spp. ('Scalindua Group B') were also observed in ETNP OMZ data which were not found in the Medway Estuary. This cluster appeared to cluster separately from known *Ca.* Scalindua spp. and was only present in samples from the oxycline and upper core of the OMZ. A similar cluster was observed by Galan, *et al.* (2009) in the upper core of the ETSP OMZ (off the coast of Chile), which was also dissimilar to known *Ca.* Scalindua spp., and so this cluster may represent a truly novel *Ca.* Scalindua sp., potentially adapted to low O_2 concentrations rather than complete anoxia (as encountered in the core of the OMZ).

These data indicate a level of anammox community structure, distribution and diversity which has previously been unreported. The large number of reads gleaned from high-throughput sequencing technologies and robust data quality control reinforce previous findings within the literature as well as reporting previously unseen observations. The presence of non-Scalindua organisms in the ETNP OMZ and diversity within the *Ca.* Scalindua clade suggest that the current opinion that anammox organisms are either freshwater or saline adapted may be overly simplistic (although salinity may be a significant environmental factor). However, the environmental data obtained during this project are insufficient to ascertain the driving forces of anammox diversity and distribution and further research is required to fully understand the ecology of anammox organisms, preferably combining both *in-situ* and *in-vitro* studies.

5.6. Summary

- Anammox organisms are environmentally important however we understand little about their diversity, distribution and environmental controls in natural environments.
- This study utilised high-throughput sequencing analysis of established and validated marker genes for anammox, which accurately represent the phylogeny of these organisms, to obtain a comprehensive understanding of the diversity and distribution of these organisms in the ETNP OMZ and Medway Estuary.
- Data from the Medway Estuary demonstrated a pronounced change in anammox diversity along the estuary, from *Ca.* Brocadia spp. dominated freshwater environments to *Ca.* Scalindua spp. dominated saline environments.
- *Ca.* S. wagneri appeared to be adapted to a salinity range from ~4 ppt-10 ppt.
- The diversity of anammox bacteria along the Medway Estuary (from upstream to downstream) was positively correlated with salinity and negatively correlated with nutrient concentrations.
- The ETNP OMZ data demonstrated a change in anammox diversity between the upper oxycline and upper core of the OMZ, the core of the OMZ and the deep core and bottom oxycline of the OMZ.
- Two distinct clusters within the *Ca.* Scalindua clade were observed in ETNP OMZ data, one representing known *Ca.* Scalindua spp. and one representing a potentially unidentified group of organisms which were not observed in the anoxic core of the OMZ.

- Evidence was also found for the presence of non-Scalindua anammox organisms in the ETNP though these were typically of low abundance and not ubiquitously dispersed in the OMZ.
- These findings elucidate the distribution and diversity of anammox organisms in these environments but further research is required to fully understand the controlling factors and ecology of these organisms.

6. Investigation into the Potential for Organic Routes to Anammox

6.1. Introduction

6.1.1. Metabolic Diversity of Anammox

The potential for an anammox-like reaction was first predicted thermodynamically by (Broda, 1977) following observations of NH_4^+ anomalies in anoxic basins (Richards, 1965; Richards, *et al.*, 1971). However, it was not until much later that the reaction was discovered in a wastewater treatment plant in The Netherlands (Mulder, *et al.*, 1995). Soon after, the microorganisms responsible for this process (see chapter 1) were characterised (van de Graaf, *et al.*, 1995) and were discovered to be autotrophic bacteria, solely using CO_2 for biomass and energy (Van de Graaf, *et al.*, 1996). To date, anammox organisms are still described as chemolithoautotrophs, gaining energy solely from the catabolism of inorganic compounds (namely NH_4^+ and NO_2^-) to fix inorganic carbon via the acetyl coenzyme A (acetyl-CoA) pathway (Russ, *et al.*, 2012; van Niftrik and Jetten, 2012). This pathway to anammox has been proven both experimentally (van de Graaf, *et al.*, 1997; Kartal, *et al.*, 2011b) and through genomic studies (Strous, *et al.*, 2006; Gori, *et al.*, 2011; Hira, *et al.*, 2012; Hu, *et al.*, 2012; van de Vossenberg, *et al.*, 2012).

However, this view appears to be overly simplistic and anammox bacteria appear to have far more diverse metabolic capabilities than previously thought (Kartal, *et al.*, 2008). Anammox organisms have been shown to be able to use ferrous iron (Fe^{2+}) or organic acids such as propionate, acetate and formate as electron donors (Güven, *et al.*, 2005; Kartal, *et al.*, 2007a; Kartal, *et al.*, 2007b; Gori, *et al.*, 2011; van Niftrik and Jetten, 2012) and can use ferric iron (Fe_3^+) and manganese oxide (MnO) as electron acceptors (Strous, *et al.*, 2006; Kartal, *et al.*, 2008). Furthermore Kartal, *et al.* (2008) found *Ca.* Brocadia fulgida to be capable of oxidising methylamines, releasing NH_4^+ in the process. However, it is maintained that anammox bacteria retain their autotrophic lifestyle, only fixing CO_2 into biomass and not organic carbon sources (Kartal, *et al.*, 2008). Güven, *et al.* (2005) stated that propionate was oxidised to CO_2 which was then fixed by anammox bacteria, however they also claimed that the anammox organisms may have also directly incorporated some of

the propionate into biomass. *Ca.* K. stuttgartiensis has also been found to be capable of DNRA (Kartal, *et al.*, 2007a), further highlighting the potential metabolic diversity of these organisms.

To date, such investigations into the potential metabolic diversity of anammox organisms have only been conducted on enrichment cultures from wastewater treatment plants. There does not appear to be any reports in the literature of similar studies being conducted on *in situ* organisms. As such the ability to oxidise organic compounds has only been observed in anammox organisms belong to the genera *Ca.* Anammoxoglobus, *Ca.* Brocadia and *Ca.* Kuenenia (Kartal, *et al.*, 2008). Kartal, *et al.* (2008) suggested that anammox bacteria have a high degree of evolutionary divergence and hence could occupy distinct ecological niches, insinuating a difference between *Ca.* Scalindua spp., incapable of oxidising organic substrates and other anammox organisms which possess this ability. However it is yet to be shown that *Ca.* Scalindua spp. are obligate autotrophs.

6.1.2. Scientific Rationale

Clearly, the current characterisation of anammox bacteria as being solely chemolithoautotrophic does not represent the metabolic diversity of these organisms which has been previously described. Recently, Karlsson, *et al.* (2009) suggested that anammox bacteria should not be thought of as specialists, but rather as generalists, utilising different metabolic pathways to survive in a wide range of ecological niches and under different environmental stresses. Indeed, anammox bacteria have been found to be active in a diverse range of environments, extending from low to high salinities, from sub-zero to temperatures in excess of 60°C and in a wide range of substrates including terrestrial soils, sediments, pelagic waters (both free floating and as part of particulate aggregates), groundwater, permafrost, estuarine and lacustrine environments (see chapter 1). It would be unreasonable to presume that such dissimilar environments would be inhabited by a narrow suite of organisms with restricted metabolic capabilities. Furthermore, a completely inorganic based metabolism may restrict anammox bacteria in certain environments (*e.g.* sediments) where the high availability of organic substrates could result in these organisms being out-competed by denitrifiers, especially as a solely chemolithoautotrophic metabolism requires 15 anammox reactions in order to fix

one mole of CO_2 (van Niftrik and Jetten, 2012). Kuypers, *et al.* (2006) further stated that organic compounds cannot be excluded as either a potential carbon or energy source for anammox.

Karlsson, *et al.* (2009) further discussed that anammox bacteria showed a much higher degree of genetic diversity than other bacteria with the *Ca.* K. stuttgartiensis genome containing approximately 200 genes involved in anammox catabolism and respiration. Furthermore it is evident from the literature that anammox organisms have many genes with currently unknown functions or regions of their genome which are yet to be sequenced (*e.g.* see Strous *et al.*, 2006). Horizontal gene transfer (HGT) has been shown to be an increasingly important in bacterial genetic diversity (Ochman, *et al.*, 2000; McDaniel, *et al.*, 2010). Considering that anammox bacteria coexist in environments with a diverse consortia of other bacteria (Byrne, *et al.*, 2008; Li, *et al.*, 2010a) and that they may be have developed early in The Earth's history (Klotz, *et al.*, 2008; Klotz and Stein, 2008) the potential for HGT is high. Hence anammox may have acquired a greater metabolic diversity through this method. Indeed, Russ, *et al.* (2012) recently discovered genes from anammox bacteria more closely associated with methanogenic bacteria, potential suggesting the sharing of genetic material between such organisms. Clearly there is a high potential for a greater metabolic diversity of these unique organisms than previously described.

However, to date, no evidence has been presented that anammox can directly utilise organic substrates in natural environments. Recently, Trimmer and Purdy (2012) discovered that organic nitrogen from allylthiourea (ATU) was converted with NO_2^- into N_2 gas by a biological process in the Arabian Sea oxygen minimum zone (OMZ). This process appeared to have a 1:1 stoichiometry as in the anammox reaction (Trimmer and Purdy, 2012). Furthermore, anammox organisms have previously been shown to be present and active in the Arabian Sea OMZ (Nicholls, *et al.*, 2007; Woebken, *et al.*, 2008). ATU is a hydrocarbon containing two amine groups at positions 1 and 3 along the carbon chain (Figure 6.1). Hence, Trimmer and Purdy (2012) concluded that anammox, or an anammox-related bacteria could be responsible for this coupling of an amine group from ATU and NO_2^-. This discovery was made as ATU was added to incubations as an inhibitor of nitrification (Trimmer

and Purdy, 2012). However, due to the serendipitous nature of this discovery, further observations either into this reaction or the microbial consortia responsible for it could not be made. Therefore, further research was needed in order to confirm the existence to such a pathway to the anammox reaction. This study aimed to investigate the microbial potential for "organammox" or organic pathways to the anammox reaction.

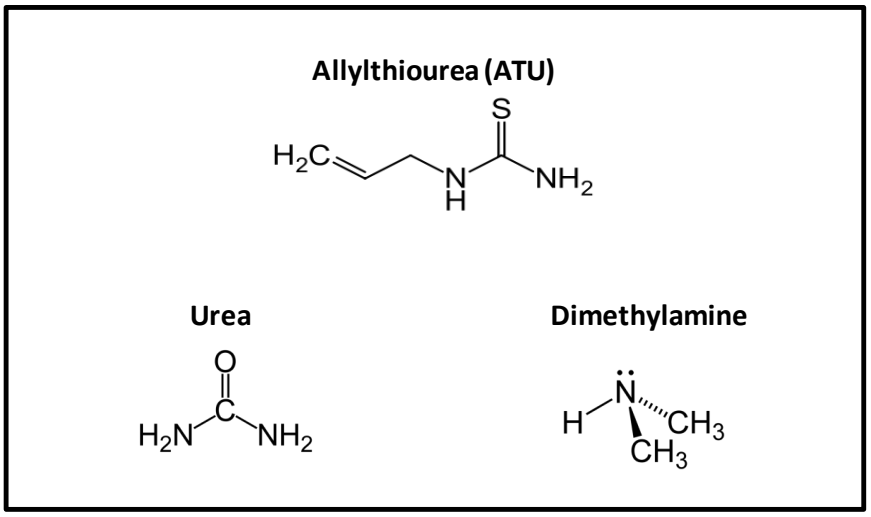


Figure 6.1: Molecular structure of allylthiourea (ATU), urea and dimethylamine (DMA). Urea and dimethylamine were chosen for this experiment as both are common organic substrates in the environment and shared similar functional groups with ATU.

6.1.3. Hypotheses

ATU is not a compound commonly found in the natural environment. However it was hypothesised that the potential for an anammox-like reaction to utilise this compound would suggest that the organisms responsible for this reaction may potentially utilise other analogous compounds. For the purpose of this investigation two compounds were chosen; urea and methyl-amine (Figure 6.1). Both of these compounds are abundant in the natural environment (Fitzsimons, *et al.*, 2001; Sliekers, *et al.*, 2004) and contain available amine groups.

It was hypothesised (see section 1.3.2.2.2), that if anammox could utilise organic substrates, this would occur in one of two ways:

a) A syntrophic relationship with other micro-organisms

b) Direct utilisation of organic nitrogen or carbon by anammox organisms

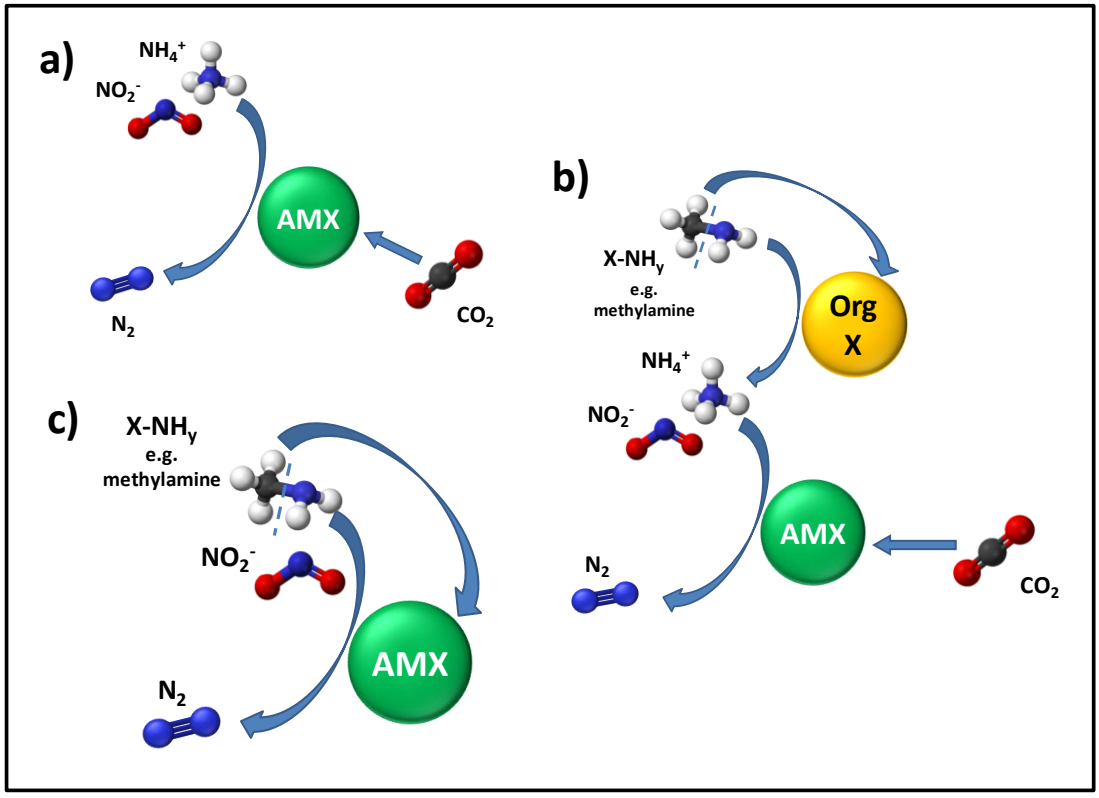


Figure 6.2: Hypothetical models showing potential organic routes to anammox. a) Classic chemolithoautotrophic anammox. b) Syntrophic pathway to anammox. c) Direct pathway to anammox. See text for full discussion.

As discussed above, classic anammox is defined as a solely chemolithoautotrophic process mediated by a unique suite of strictly anaerobic bacteria (van Niftrik and Jetten, 2012). One molecule of NH_4^+ is oxidised to one molecule of N_2 with one molecule of NO_2^-, providing the energy required to fix CO_2 (Figure 6.2)[*]. The classic anammox reaction requires the acquisition of freely available dissolved inorganic nitrogen (DIN) from the environment.

Most DIN is originally derived from organic matter which has been liberated by heterotrophs in the environment. The environment is inhabited by numerous microbial consortia operating within different ecological niches. The interactions of such microorganisms, both between themselves and the environment, are often overlooked when investigating microbial ecology. Therefore anammox bacteria may

[*] See chapter 1 for a full discussion of the anammox reaction and biochemistry.

oxidise liberated amine groups (released as NH_4^+) from organic compounds as a waste product of other microorganisms (Figure 6.2), *e.g.* methanogens release NH_4^+ as a waste product from breaking down methyl-amines. Anammox could therefore operate within a syntrophic relationship with such organisms, feeding off waste N-compounds while potentially aiding their syntrophic partner by reducing the toxic effect of high NH_4^+ concentrations. If this were true, a close spatial relationship might be expected between anammox and potential syntrophic partners. This would potentially explain the findings of Trimmer and Purdy (2012), as anammox, if forming aggregates with other microorganisms, would preferentially utilise liberated NH_4^+ from organic substrates over DIN as this source of NH_4^+ would exhibit a higher degree of bioavailability. This would explain the low measurements of $^{29}N_2$ from incubation with $^{15}NO_2^-$ and $^{14}NH_4^+$ and far greater values when amended with $^{15}NO_2^-$ and $^{14}NH_4^+$ and ATU (Trimmer and Purdy, 2012). There are also further reports in the literature suggesting that anammox may be found within flocs of microorganisms and particulate matter within open waters and exhibit close relationships with other microorganisms (Woebken, *et al.*, 2007; Quan, *et al.*, 2008; Hu, *et al.*, 2012).

Alternatively, anammox bacteria may have the ability to catabolise organic nitrogen directly (Figure 6.2). Genomic and transcriptomic studies on anammox organisms have been severely limited by the inability to grow these organisms in pure culture (Jetten, *et al.*, 2009). Although the metabolic pathway of the anammox reaction has been empirically proven (Kartal, *et al.*, 2011b), there remains much discrepancy in the literature as to the genomic information available for these novel bacteria and many sequenced genes have yet to be assigned a function. Furthermore, anammox organisms have a far greater catabolic gene diversity than other bacteria (Karlsson, *et al.*, 2009) potentially suggesting a far greater metabolic diversity than hitherto thought. Therefore, anammox may be essentially mixotrophic, capable both of the classical chemolithoautotrophic anammox reaction and of utilising organic substrates. The presence of mixotrophic anammox organisms may lead to the discovery on new anammox organisms, potentially outside of the anammox diversity previously reported. However, the potential for the direct utilisation of organic carbon by anammox bacteria is, as of yet, unproven and so how such a process may occur is entirely speculative.

6.2. Methodology

6.2.1. Overview

SIP experiments were carried out in order to investigate the potential for organic pathways to the anammox reaction. Microcosms were set up at the locations and sites discussed in section 2.4 (see also Table 2.6).

Initially, a pilot study was conducted on sediment collected from the Medway Estuary site M6 collected on 15/02/2011 in order to investigate the potential for organammox in these sediments and to highlight potential issues which may arise from the chosen methodology and experimental plan. Sediment from this site was chosen as it had previously demonstrated the highest anammox rates along the Medway Estuary (Nicholls and Trimmer, 2009). The pilot study did not involve a complete suite of controls (*e.g.* addition of unlabelled substrates) in order to facilitate sample processing and analysis. As such only no addition controls were used in conjunction with experiments labelled with ^{13}C DMA and urea.

Following the pilot study, two subsequent experiments were carried out investigating the potential for the existence of this process (see Table 2.2). Experiment 1, set up with sediment collected on 25/07/2011, incorporated incubations amended with ^{12}C and ^{13}C DMA, TMA and urea with no addition controls. In addition, experiments were set up with double the concentration of urea and DMA in order to investigate whether the use of greater substrate concentrations aided downstream analysis and amplification of ^{13}C DNA. Experiment 2, set up with sediment collected on 15/03/2012, investigated the potential for the incorporation of organic carbon into anammox biomass at sites M1 and M6. Incubations were set up amended with ^{12}C and ^{13}C DMA and urea with no addition controls (see Table 2.6).

The varying conditions and concentrations of substrates used in these experiments and the date of sample collection can be seen in Table 2.6. Experiments were carried out as outlined in the following sections and in chapter 2.

6.2.2. Microcosm Set-up – Sediments

6.2.2.1. Set-up

5 cm^3 sediment cores were taken in triplicate (within a 10 cm radius) from estuarine sediment at the low-water mark and transported back to the laboratory in cool-bags. Samples for environmental population analysis were also taken at this time from which pore water was extracted. 5 ml of this pore water was added to each 5 cm^3 of sediment in a glass serum bottle.

Slurries were amended with either ^{12}C or ^{13}C labelled di-methylamine (DMA), tri-methylamine (TMA)[*] or Urea (Table 2.6). All labelled isotopes were 99.9% labelled and were supplied by Sigma-Aldrich, UK. Microcosms were spiked with 100 µl of 1M solutions (final concentration 0.01 M[†]) which had been filter sterilised. Microcosms were sealed with a butyl stopper and crimped with an aluminium seal. Headspace was flushed with oxygen-free-nitrogen (OFN) via a hypodermic needle, with a second needle as an outflow, for approximately 20 minutes. The outflow was removed before the inflow to create a slight positive pressure and to prevent potential back-flow of oxygenated air. Microcosms were incubated at room temperature in the dark for approximately three weeks (see section 6.3.1.2) at which point they were frozen at -20°C until analysis.

6.2.2.2. Gas Chromatography

CH_4 concentrations in headspace gases from microcosms amended with methylamines (either DMA or TMA) were measured periodically during incubation in order to determine when the substrates had been completely utilised and therefore when would be a suitable time to terminate the experiment. CH_4 measurements were taken using an Agilent 6890N gas chromatograph (GC) with a flame ionisation detector (FID) and a Porapak-Q (2 m x 3 mm) column (Waters, USA). A nitrogen (N_2) carrier gas was used (BOC, UK). The injector temperature was set to 180°C, the detector to 250°C and the column temperature to 90°C. 0.1 ml of headspace gas was

[*] It was hypothesised that if the anammox reaction could use product from methyl-amines the mechanism for utilising different methyl-amines would be similar and therefore different compounds (either DMA or TMA) were used interchangeably. However, in one experiment solely one or the other compound was used, variation only existed between experiments conducted on samples collected at different times.

[†] Some experiments were amended with 0.02 M of substrate (see Table 2.6).

injected directly into the injector using a syringe, needle and a Supelco Mininert valve (Sigma-Aldrich, UK). Headspace gas measurements for each microcosm were taken in triplicate.

6.2.3. Microcosm Set-up – Marine

Water was captured at specific depths (Table 2.6) using a deck-controlled CTD with a Niskin rosette. On deck, water was extracted from Niskins in triplicate directly into 1 L serum bottles using PVC tubing. Tubing was inserted into the bottom of the serum bottle and allowed to fill up 3 times. The tubing was removed allowing that the serum bottle to be overfilled and a butyl stopper carefully inserted so as to prevent oxygenated air entering the serum bottle. Serum bottles were crimped with aluminium seals. A 4 ml headspace of He was introduced into each serum bottle. Microcosms were amended with TMA or urea (final concentration 100μM). Microcosms were also set-up with both TMA/urea and NO_2^- (final concentration 25 μM). Microcosms were incubated at 4°C (the *in-situ* temperature) for approximately 2 weeks at which point they were filtered through 0.2 μm poly-carbonate filter (Whatman, USA) using a vacuum pump and flash frozen in liquid nitrogen prior to DNA extractions.

6.2.4. DNA Extraction

DNA was extracted from both sediment and marine microcosms using the protocol outlined in Purdy (2005). DNA was extracted from approximately 5 g of wet sediment or the entire polycarbonate filter[*] (depending on microcosm substrate). The protocol was followed exactly as in Purdy (2005) except that, after the removal of proteins, hydroxyapatite columns were washed with 150 mM K_2HPO_4 in order to remove RNAs. DNA was then eluted from hydroxyapatite columns using 300 mM K_2HPO_4 and purified as per the protocol. DNA was not purified via PEG precipitation, as this tended to result in a loss of DNA and thus may have been detrimental to the SIP protocol.

[*] There was not a sufficient amount of biomass on half a filter after SIP incubations to obtain a suitable yield and therefore extractions were conducted on the whole filter (*c.f.* chapter 2).

6.2.5. SIP Fractionation

Samples were fractionated in order to separate heavy (^{13}C-labelled) and light (^{12}C-labelled) DNA as outlined by (Neufeld, *et al.*, 2007) to identify organisms which had assimilated carbon from labelled organic sources. The densities of fractions were measured using a AR200 digital refractometer (Reichert, USA) to ensure that density gradients had formed correctly during centrifugation. Separation of ^{12}C and ^{13}C fractions was expected to be observed at densities ≈1.7250 g ml^{-1} (Neufeld, *et al.*, 2007). Fractions were then PEG precipitated (see section 2.4.5). For a full description of the SIP protocol used in this study see chapter 2.

6.2.6. PCR

6.2.6.1. 16S rRNA primers

PCR was performed on purified SIP fractions as outlined in chapter 2. To amplify product from these samples it was necessary to perform a nested PCR, using Pla46F & 1390R as a first round PCR followed by Amx368F & Amx820R, as DNA concentrations were frequently low.

6.2.6.2. *hzo* Primers

DNA from selected SIP fractions (representing suspected ^{12}C and ^{13}C containing fractions from Urea and DMA incubations) was also amplified using primers targeting *hzo* genes in order to investigate whether the diversity of this gene represented metabolic, rather than phylogenetic diversity. DNA was amplified using a nested PCR approach as it proved difficult to amplify from these SIP samples directly with a single round approach, potentially due to low yields of target DNA after fractionation (see section 6.3). As such a first round PCR using primers HZO4F & HZO1R was conducted before amplification with "*hzo* cluster 1" specific primers hzocl1F1l & hzocl1R2. This methodology is similar to that outlined in Hirsch, *et al.* (2011) except that primer hzocl1F1l was used instead of hzocl1F1 based on the results presented in chapter 4. PCR was performed as outlined in chapter 2.

6.2.7. DGGE

PCR products were analysed using DGGE to investigate community shifts in the anammox population between heavy and light DNA fractions (Schafer and Muyzer, 2001). Samples to be analysed by this method were amplified as per the standard PCR protocol except that primer Amx368F was substituted with the same primer with a GC clamp attached to the 5' end (Amx368F-GC see Table 2.1). Polyacrylamide gels were stained for 30 minutes using ethidium bromide and then destained for 30 minutes in pure water, both with gentle shaking. Polyacrylamide gels were viewed using a UV gel imager.

6.2.8. Sequencing

In order to identify notable bands from DGGE gels, bands were excised from polyacrylamide gels for sequencing. Poor results were experienced using the commercial gel extraction kit (Qiagen, UK). As such the gel extraction protocol described by Schafer and Muyzer (2001) was used. 1 µl of supernatant was amplified via PCR using primers Amx368F & Amx820R (see section 2.3.2.1). PCR product was cloned into the vector pGEMT-Easy (Promega, UK) and transformed into JM109 competent *E. coli* cells (Promega, UK) as per the established protocol (see section 2.3.3 for full details). Clones were grown overnight at 30°C on LAXI plates and white colonies selected. Colonies were boiled for 10 minutes and then amplified via PCR with primers M13F and M13R (see chapter 2.3.3) in order to check that the colony contained an insert of the correct size. M13 PCR product was purified and submitted to GATC, UK for Sanger sequencing from both ends of the read. Reads were checked manually for errors and contigs assembled using SeqMan II (DNASTAR, USA). Alignments and phylogenetics were conducted using Mega 4.0-5.1 (Tamura, *et al.*, 2007; Tamura, *et al.*, 2011).

6.2.9. 454 Pyrosequencing

Some samples were also submitted for 454 pyrosequencing in order to determine the diversity in heavy and light DNA fractions. Samples were submitted as outlined in section 2.4.9. Clusters were defined at a similarity cut-off value of 95%. Clusters were identified via inference after phylogenetic analysis of reference sequences from each cluster with published sequences from NCBI (see section 2.3.5.2.2), *i.e.* if the

reference sequence for a particular cluster was shown to have the greatest similarity with a group of known anammox sequences (*e.g. Ca.* Kuenenia spp.), it was inferred that this cluster represented these organisms.

6.3. Results

6.3.1. Pilot Study

6.3.1.1. Overview

Initially a pilot study was conducted on sediment collected from Medway Bridge Marina site M6 (Figure 2.2) on 15/02/2011 in order to test the above experiment procedure and identify potential areas which required optimisation prior to a more comprehensive experiment. Microcosms were incubated with either ^{13}C-labelled DMA or ^{13}C-labelled urea in triplicate. No addition controls were also set-up in triplicate. Killed controls were not performed during this experiment as, due to the speculative nature of the hypotheses being tested, it was not imperative at this stage to investigate whether organic substrates were being degraded biotically or abiotically, only whether these organic substrates were made available for anammox organisms (or their syntrophic partners) to utilise.

6.3.1.2. Loss of Substrates

The production of CH_4 in pilot experiment microcosms amended with DMA was measured during incubation in order to aid the assessment of when to terminate the experiment[*].

Figure 6.3 shows the production of CH_4 in pilot microcosms over the period of incubation. After an initial lag phase of 5-7 days, CH_4 production increased rapidly in DMA amended microcosms. After 13-14 days the production of CH_4 appeared to plateau with a maximum concentration of 578.47 ± 44.36 ppb ml^{-1} of headspace gas. In comparison, no CH_4 was seen to accumulate in "no addition" microcosms over this period (Figure 6.3) indicating that the increase in CH_4 concentrations in DMA amended microcosms was likely to be as a result of the degradation of DMA.

[*] At the time of experimentation no method was available within the laboratory for measuring the loss of urea in these substrates and therefore it was assumed (if a similar consortia of organisms utilised both urea and methylamines in similar ways) that urea would be utilised at a similar rate to DMA.

Therefore, on the basis of these data, it was decided to terminate these experiments at approximately 3 weeks, a week after it appeared as if all the ^{13}C labelled substrate had been utilised.

Attempts were made to determine what percentage of amended DMA had been converted to CH_4 (and therefore what proportion had been incorporated into biomass) though these were unsuccessful, presumably due to the low accuracy of this instrument and difficulties associated with accurately determining the volume of gas in the microcosm.

CH_4 measurements were also taken during the incubation of Medway Estuary SIP experiments 1 and 2 (see Table 2.6). These demonstrated similar results (*i.e.* plateauing of CH_4 production after ~14 days) to those of the pilot study and hence are not presented in this thesis.

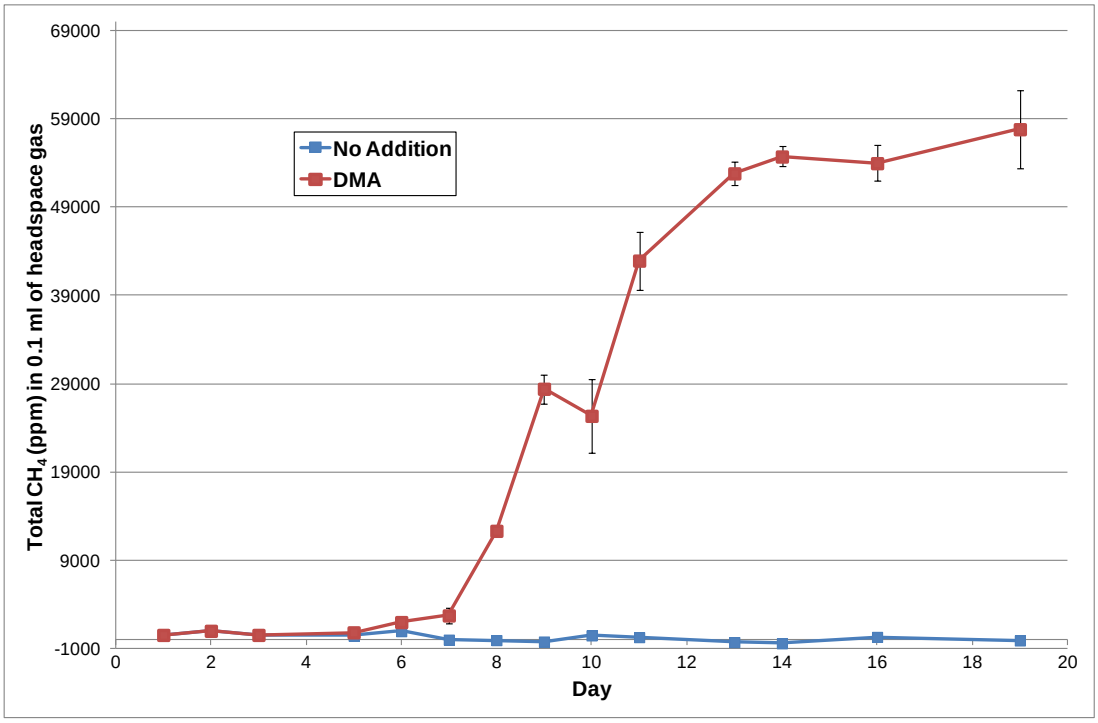


Figure 6.3: Concentration of CH4 (ppm) in headsace gas frompilot SIP microcosms. Figure shows the production of CH_4 in 0.1 ml headspace gas from pilot SIP microcosms amended with DMA. Error bars represent standard errors from triplicate measurements of each microcosm (3 microcosms per incubation in total). CH_4 production was measured as an indicator of when all DMA in the microcosm had been utilised and hence when was a suitable time to terminate the experiment. After an initial lag phase of 5-7 days, CH_4 production in DMA amended microcosms (red) increased rapidly until day 13 at which point it began to plateau. In comparison, CH_4 was not seen to accumulate in any of the no addition (blue) microcosms demonstrating that CH_4 was not produced in Medway Estuary M6 sediments under anaerobic conditions. Therefore the increase in CH_4 concentrations observed in DMA amended microcosms is a result of the conversion of DMA. Therefore it was decided to terminate this experiment at approximately 3 weeks.

6.3.1.3. Fractionation

Following DNA extraction and centrifugation samples were fractionated to separate heavy and light DNA fractions. Fraction densities, calculated after fractionation, are represented in Table 6.1. Frequently the density gradient was truncated, as the mixing of water with CsCl during the end of fractionation produced lower density readings. These erroneous data were omitted when calculating gradients (Table 6.1). In most samples the density gradient is constant between all fractions indicating that all fractions have remained intact and that there has been no (or at least little) mixing of DNA from different fractions. However, in ^{13}C-DMA 2, issues encountered during fractionation resulted in lower observed densities though these still formed a reasonable gradient and so this sample was included in downstream analysis.

	Density (g/ml)								
Fraction	**No Addition 1**	**No Addition 2**	**No Addition 3**	**^{13}C DMA 1**	**^{13}C DMA 2**	**^{13}C DMA 3**	**^{13}C Urea 1**	**^{13}C Urea 2**	**^{13}C Urea 3**
1	1.7461	1.7438	1.7449	1.7472	1.7427	1.7449	1.7449	1.7472	1.7438
2	1.7438	1.7427	1.7438	1.7461	1.7393	1.7438	1.7427	1.7438	1.7427
3	1.7416	1.7404	1.7404	1.7427	1.7359	1.7404	1.7393	1.7416	1.7404
4	1.7382	1.7370	1.7370	1.7393	1.7325	1.7370	1.7359	1.7370	1.7370
5	1.7337	1.7325	1.7348	1.7370	1.7269	1.7337	1.7303	1.7337	1.7337
6	1.7303	1.7292	1.7292	1.7325	1.7224	1.7292	1.7269	1.7292	1.7303
7	1.7247	1.7258	1.7258	1.7280	1.7190	1.7258	1.7224	1.7258	1.7258
8	1.7213	1.7213	1.7213	1.7247	1.6650	1.7213	1.7179	1.7202	1.7224
9	1.7179	1.7168	1.7179	1.7213	1.7134	1.7179	1.7134	1.7179	1.7190
10	1.7145	1.7145	1.7156	1.7179	1.6965	1.7134	1.7100	1.7145	1.7156
11	1.7111	1.7123	1.7111	1.7145	1.4363	1.7111	1.6773	1.7089	1.7100
12	1.6683	1.6683	1.6548	1.6796	N/A	1.6650	1.2561	1.5805	1.5647
Mean	**1.7294**	**1.7288**	**1.7293**	**1.7319**	**1.7290**	**1.7290**	**1.7237**	**1.7291**	**1.7292**
Gradient	**-0.0037**	**-0.0035**	**-0.0035**	**-0.0034**	**-0.0038**	**-0.0036**	**-0.0054**	**-0.0038**	**-0.0034**
Intercept	**1.75**	**1.75**	**1.75**	**1.75**	**1.75**	**1.75**	**1.76**	**1.75**	**1.75**
R^2	**0.9937**	**0.9903**	**0.9922**	**0.9938**	**0.9917**	**0.9940**	**0.8480**	**0.9958**	**0.9888**

Table 6.1: Table showing densities of fractions from SIP samples. Fractions show that density gradients remained intact during fractionation and therefore there was minimal mixing of DNA from these fractions. Due to issues experienced during fractionation (see main text), density measurements from fraction 12 in samples No Addition 1-3, ^{13}C DMA 1 & 3 and ^{13}C Urea 1-3 were deemed anomalous and excluded from gradient calculations. Due to further complication associated with the fractionation of samples 13C DMA 2, fractions 8, 10 and 11 were excluded from gradient calculations.

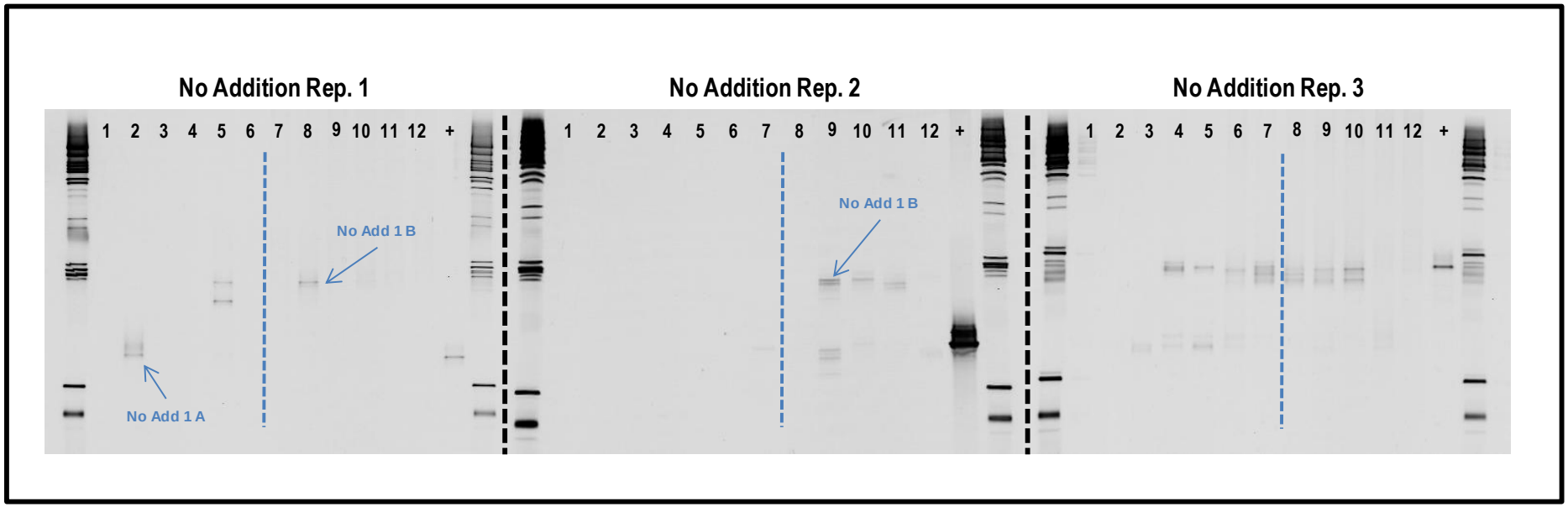


Figure 6.4: DGGE images for No Addition replicates from pilot SIP experiment. Fractions are labelled 1 to 12, corresponding with table PILOT_1 (*i.e.* Fraction 1 is the highest density and fraction 12 the lowest). Blue hashed lines represent the expected separation between ^{12}C and ^{13}C DNA bearing fractions. Highlighted bands are those from which bands were sequenced and band names and colours are the same as those in Figure **6.7**. No clear enrichment between ^{12}C and ^{13}C containing fractions is evident.

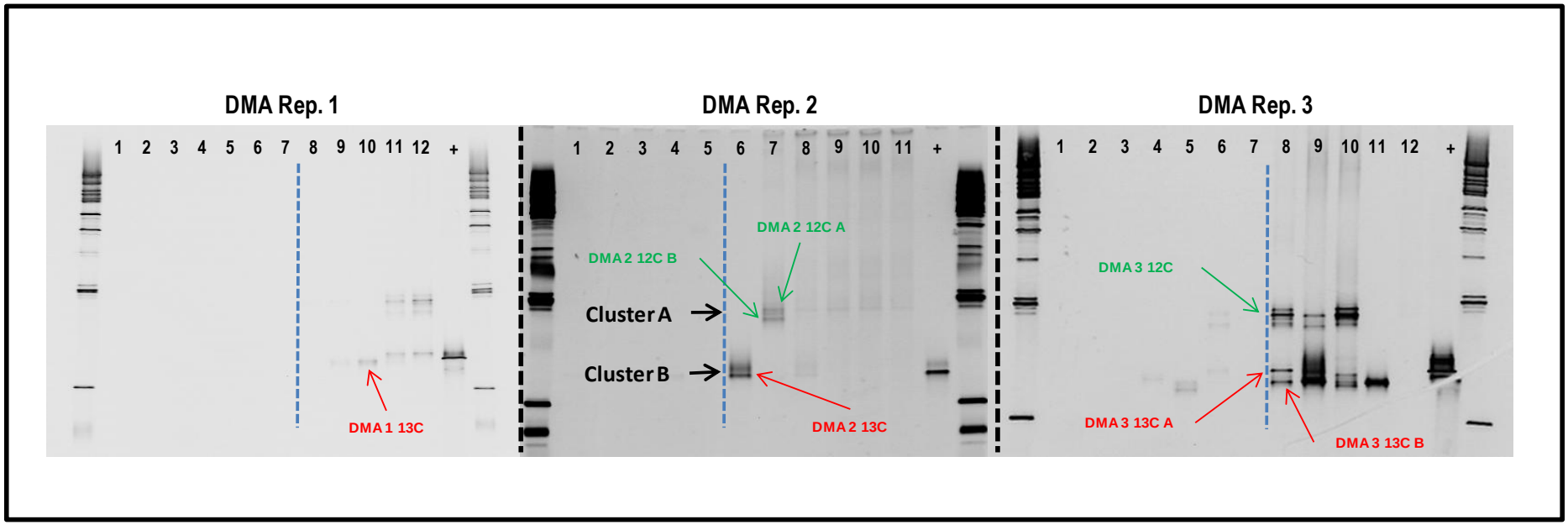


Figure 6.5: DGGE images for DMA replicates from pilot SIP experiment. Fractions are labelled 1 to 12, corresponding with table PILOT_1 (i.e. Fraction 1 is the highest density and fraction 12 the lowest). Blue hashed lines represent the expected separation between ^{12}C and ^{13}C DNA bearing fractions. Highlighted bands are those from which bands were sequenced and band names and colours are the same as those in Figure **6.7**. A community shift is evident between ^{13}C and ^{12}C fractions (fractions 7 and 8) in replicate 2 though this was not observed in the other two replicates.

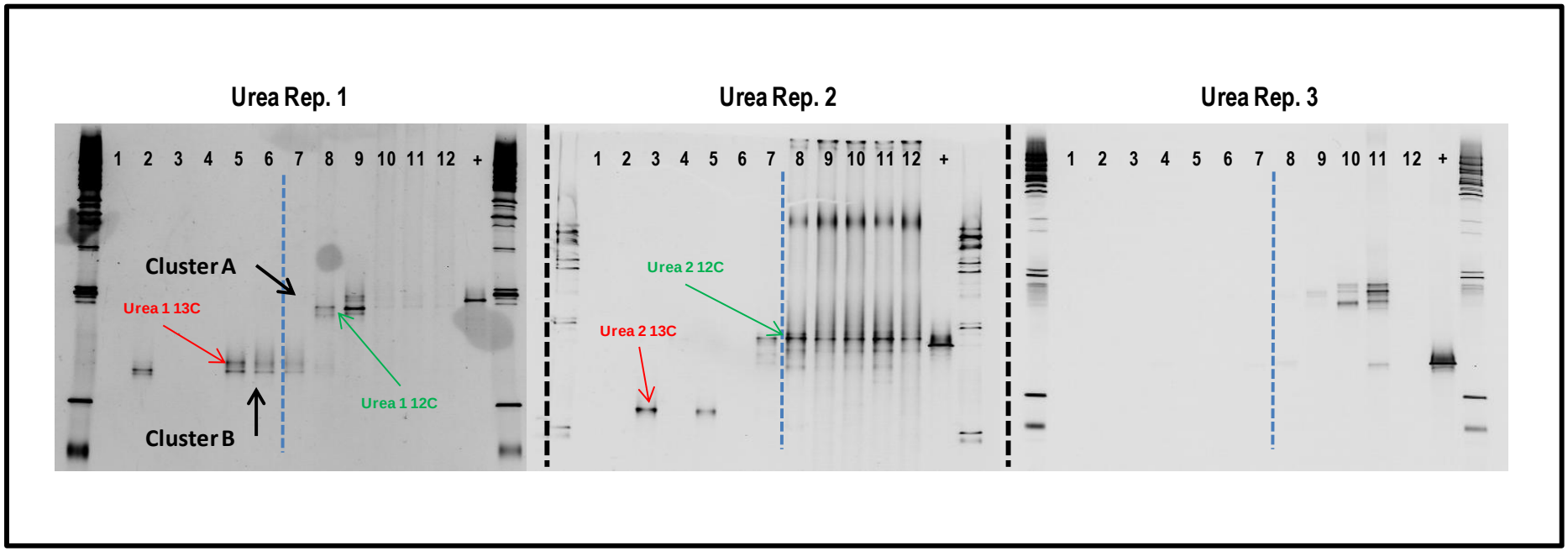


Figure 6.6: DGGE images for Urea replicates from pilot SIP experiment. Fractions are labelled 1 to 12, corresponding with table PILOT_1 (*i.e.* Fraction 1 is the highest density and fraction 12 the lowest). Blue hashed lines represent the expected separation between ^{12}C and ^{13}C DNA bearing fractions. Highlighted bands are those from which bands were sequenced and band names and colours are the same as those in Figure **6.7**. A community shift is evident between ^{13}C and ^{12}C containing fractions (fractions 7 and 8) in replicate 1. A similar community shift may also be observed in replicate 2 (fractions 5 and 7) though this is less apparent. No change in the anammox community was observed in replicate 3.

6.3.1.4. PCR and DGGE of 16S rRNA Genes

DNA purified from SIP fractions was amplified via PCR. Due to the low yields of DNA obtained after fractionation, it was necessary to perform a nested PCR, as outlined in section 6.2.6, in order to amplify anammox 16S rRNA genes from these samples. It was not possible to amplify DNA from all fractions and, in general, those which were amplified were from less-dense fractions (*i.e.* towards fraction 12). This was as expected as DNA should have collected towards the top half of the ultracentrifuge tube during centrifugation (Neufeld, *et al.*, 2007). Amplification of DNA from lower fractions (*i.e.* fractions 1-3) was probably due to a smearing effect of DNA throughout the ultracentrifuge tube (see section 6.4.2.2 for full discussion).

All samples which were amplified via the standard anammox PCR were also able to be amplified using anammox specific DGGE-primers. Images of DGGE gels from this pilot study can be seen in Figure 6.4, Figure 6.5 and Figure 6.6.

All three "no addition" replicates (No Addition 1-3) did not demonstrate any clear community change between fractions that would have been expected to contain ^{13}C-labelled and unlabelled DNA. Amplification efficiencies of these samples appeared to be poor with low yields of DNA, even after two rounds of PCR. Thus, these data show that there was no enrichment of DNA in ^{13}C fractions with no addition of labelled substrate. Thus, any enrichment observed in ^{13}C amended samples is likely to be as a result of incorporation of ^{13}C into microbial DNA.

It is important to note that the expected density of ^{13}C-labelled DNA (*i.e.* 1.725 g ml^{-1}) is only an approximation and will vary slightly based on the mass of DNA of the particular organisms involved. Hence, such experiments conducted on a diverse, *in situ* community are unlikely to produce clear and definitive trends. However, the separation between ^{12}C and ^{13}C bearing fractions was commonly insinuated (Neufeld, *et al.*, 2007) to be between fraction 6 and 7 or 7 and 8.

Fractions from DMA replicate 1 (DMA-1) also demonstrated poor amplification efficiencies. Regardless, two clear bands (or groups of bands) can be observed in fractions from this sample. DMA replicates 2 and 3 (DMA-2 and DMA-3) exhibited much greater yields after PCR amplification. DMA-2 appears to show a community shift between fractions 6 and 7, approximately where heavy and light DNA would be

expected to diverge. However, DMA-3 did not demonstrate a clear community change between these fractions, though the lower of the two bands did seem to be enriched compared with other samples, perhaps indicating greater numbers of these anammox organisms in this microcosm (however caution must be used when comparing bands from different DGGE gels).

Urea amended samples showed similar results. Again these samples appeared to show a greater amplification efficiency compared with no addition controls, potentially suggesting a greater amount of target DNA in these samples (however this can only be inferred with the present data). Replicate 1 (Urea 1) shows a clear community change between fractions 7 and 8, similar to that observed in DMA 2. Urea-2 may also show such a trend though this is less clear due to the absence of DNA in fractions 4 and 6. However, Urea-3 exhibits no discernible trend in the data though it was only possible to amplify DNA from fractions 9, 10 and 11 from this sample which represent ^{12}C bearing fractions.

6.3.1.5. 16S rRNA Clone Library of Notable Bands from DGGE

Notable bands from the DGGEs described above (Figure 6.4, Figure 6.5 and Figure 6.6) were excised from DGGE gels and sequenced as outlined in section 6.2.8 in order that potential anammox community changes, observed within the data, might begin to be investigated. Initially 20 bands were selected representing bands from fractions suspected of containing ^{12}C and ^{13}C DNA as well as no addition controls. Single colonies from cloning were sequenced in both the forward and reverse direction in order to obtain sequence data for the entire 16S rRNA region amplified by primers Amx368F & Amx820R. 6 bands/clones produced poor sequencing data and so were removed from the analysis. The remaining 14 sequences were processed as outlined in section 6.2.8.

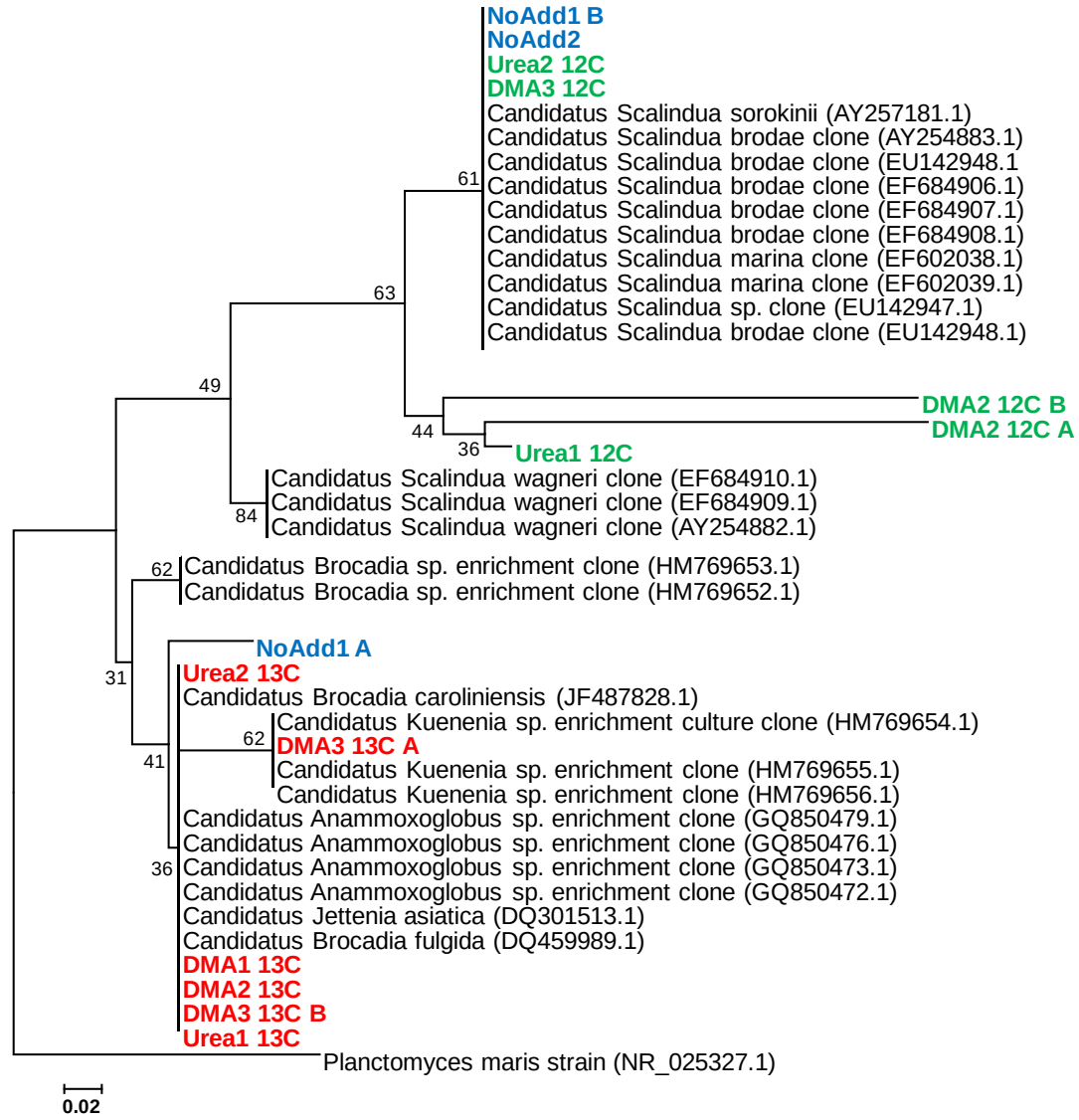


Figure 6.7: 16S rRNA phylogenetic tree of bands excised from DGGE gels from pilot SIP investigation. Bands excised from ^{12}C fractions were found to cluster within the genus Ca. Scalindua whereas ^{13}C fractions appeared to cluster with the remaining anammox genera, suggesting that ^{13}C-labelled organic substrates had been assimilated into biomass by non-Scalindua organisms. Phylogenetic tree was calculated and built in MEGA 5 (Tamura, *et al.*, 2011). Phylogenetic tree was calculated using a neighbour-joining methodology using a p-distance model. 1000 bootstrap replications were calculated as a test of phylogeny. Branch labels represent the percentage of bootstrap trees which demonstrated this phylogenetic relationship. Scale bar represents a 2% sequence divergence.

Figure 6.7 shows the phylogenetic relationship of these sequences with other known anammox sequences. All sequences fall within the anammox clade, as would be expected with these PCR primers (see chapter 3). Sequences obtained from ^{12}C bearing fractions (labelled green in Figure 6.7), which correspond with 'DGGE Cluster A', all clustered with 16S rRNA genes related to *Ca.* Scalindua spp.

Sequences obtained from ^{13}C DNA fractions (red in Figure 6.7), 'DGGE Cluster B', however, showed a high degree of similarity to non-Scalindua anammox organisms, specifically *Ca.* Brocadia spp. and *Ca.* Kuenenia spp.

Sequences obtained from no addition controls clustered with both *Ca.* Scalindua spp. and non-Scalindua anammox 16S rRNA sequences, suggesting that non-Scalindua organisms are present in detectable quantities in these samples prior to the addition of these substrates.

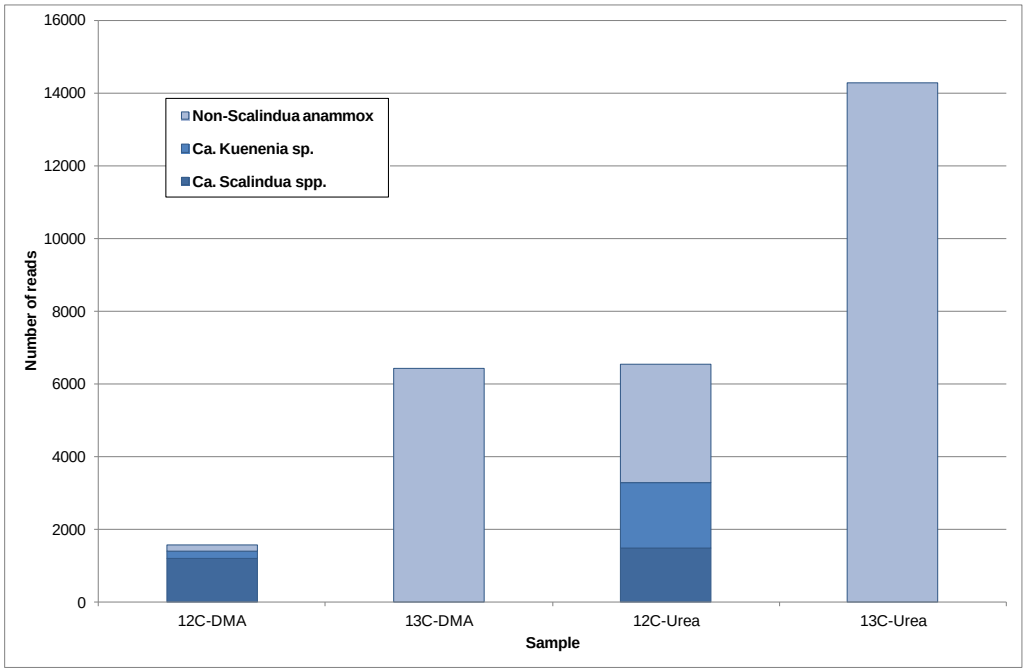
6.3.1.6. 454 Pyrosequencing of ^{12}C and ^{13}C Enriched Fractions

Selected fractions from the pilot SIP experiment were amplified via 16S rRNA PCR, specific for anammox bacteria, and submitted for 454 pyrosequencing in order to investigate whether the relationships observed in the sequencing of bands from DGGE remained true with the application of more stringent measurements of ecological diversity.

4 fractions from 2 samples were chosen for analysis: DMA-2 fractions 6 and 7 and Urea-1 fractions 5 and 8. A total of 64,923 raw pyrosequencing reads were obtained of which 28,785 passed quality control. Reads were clustered into OTUs at a cut-off of 95% sequence similarity. OTUs representing less than 0.1% of the total number of reads (5 OTUs, representing a total of 23 reads) were omitted from the analysis leaving 10 OTUs for downstream analysis.

As can be seen from Figure 3.1, heavy, ^{13}C-labelled DNA fractions were dominated by reads belonging to non-Scalindua anammox organisms. Non-Scalindua anammox organisms were identified as such as, after phylogenetic analysis of these clusters with reference sequences from the NCBI database, these clusters appeared to cluster outside the *Ca.* Scalindua spp. 16S rRNA clade but neither did they cluster closely with known, non-Scalindua anammox. These clusters may represent a previously un-reported anammox clade, potentially capable of heterotrophic anammox, however it would be impossible to assert this and characterise these sequences based on inferred relationships with a relatively short sequence of 16S rRNA. ^{12}C-labelled DNA fractions however showed a much greater diversity of anammox bacteria. Sequences obtained from these fractions were identified as *Ca.* Scalindua spp., *Ca.* Kuenenia spp. and non-Scalindua anammox. The proportions of different anammox organisms

in these two samples are not consistent however, without the presence of replication, it is impossible to draw any conclusions from this.



Cluster #	^{12}C-DMA	^{13}C-DMA	^{12}C-Urea	^{13}C-Urea	Identity
Cluster 0	157	6411	2924	14267	Non-Scalindua anammox
Cluster 1	149	0	73	0	*Ca.* Scalindua brodae/marina
Cluster 2	909	0	929	0	*Ca.* Scalindua wagneri
Cluster 3	78	0	169	0	*Ca.* Scalindua wagneri
Cluster 4	19	0	67	0	*Ca.* Scalindua brodae/marina
Cluster 5	11	0	144	1	Non-Scalindua anammox
Cluster 6	15	0	198	0	*Ca.* Scalindua wagneri
Cluster 7	0	0	188	0	Non-Scalindua anammox
Cluster 8	209	0	1794	0	*Ca.* Kuenenia sp.
Cluster 9	6	0	44	0	*Ca.* Scalindua wagneri

Figure 6.8: Number of 16S rRNA 454 reads associated with different OTUs from sequenced from SIP DNA fractions. Table shows cluster identities and abundance of anammox 16S rRNA 454 pyrosequencing data from pilot SIP DNA fractions. OTUs were defined at a cut-off of 95% sequence similarity. OTUs representing < 0.1% of the total number of reads have been omitted (95 OTUs). Both ^{13}C-labelled DNA fractions are dominated by sequences belonging to non-Scalindua anammox organisms whereas ^{12}C-labelled DNA fractions appear to show a greater diversity of anammox organisms including Ca. Scalindua spp. and non-Scalindua organisms.

6.3.1.7. Analysis of *hzo* Gene Sequences Obtained from ^{12}C and ^{13}C Enriched Fractions

As it had been concluded that *hzo* genes did not represent the phylogeny of anammox organisms (see chapter 4) it was hypothesised, in light of the above

findings, that the diversity of *hzo* may represent metabolic diversity within anammox organisms (*i.e.* representing anammox capable of mixotrophic versus autotrophic growth).

Figure 4.3 in chapter 4 shows the similarity of *hzo* gene sequences obtained from these samples in relation to other *hzo* sequences obtained from this study and those downloaded from the NCBI GenBank database. *hzo* clones obtained from SIP fractions are highlighted in blue. Clones labelled DMA-2-6 and Urea-1-5 were obtained from fraction believed to contain ^{13}C-labelled DNA whereas those labelled DMA-2-7 and Urea-1-8 were from ^{12}C-labelled DNA fractions[*].

Sequences obtained from SIP fractions showed a greater diversity of *hzo* than those obtained from Medway Estuary M6 samples (Figure 4.3). This was expected as 16S rRNA diversity analysis of these fractions had produced sequences identified as *Ca.*Scalindua spp., *Ca.* Brocadia spp. and *Ca.* Kuenenia spp. (see sections 6.3.1.5 and 6.3.1.6). However, clones of *hzo* sequences obtained from these samples did not clearly cluster into ^{12}C and ^{13}C containing fractions (as shown above) but instead were distributed amongst the whole of *hzo* clusters 'A' and 'B'. Only one clone was observed in the suggested "non-Scalindua" clusters of 'C', 'D' and 'E' (Li, *et al.*, 2010b). The apparent increase in diversity in these samples compared with the previous investigation into *hzo* diversity (see chapter 4) may be due to the nested PCR (used on SIP samples) being more specific at amplifying target sequences rather than an increase in *hzo* diversity. Thus, before further investigations into the use of *hzo* as a molecular marker of anammox it would be imperative that Medway Bridge Marina DNA samples are amplified using the same approach as that used on SIP DNA. However, due to the reasons stated in chapter 4, *hzo* was deemed not to be a suitable functional marker for anammox based on the available data and as such, it was decided not pursue this avenue of research any further.

Clones from ^{13}C-labelled DNA fractions did not cluster together and neither did those from ^{12}C-labelled DNA fractions. This suggests that the diversity of the *hzo* gene (at least represented by the primers used during this study) does not represent a metabolic diversity within anammox bacteria between organisms capable of

[*] In these clones, the first number represents the replicate number and the second number represents the fraction from which the clone was obtained (see Figure 6.4, Figure 6.5 and Figure 6.6).

mixotrophic and autotrophic lifestyles. However all DMA related sequences (n=10) did cluster within 'Anammox cluster A' and all except 2 (*i.e.* Urea-1-8 clone 8 and Urea-1-8 clone 9) urea sequences (n=10) clustered within 'Anammox cluster B'. This may indicate that the diversity of *hzo* may be dependent on metabolic diversity within anammox organisms capable of utilising different organic substrates. However the data presented here are not sufficient to verify such a claim and further experimentation would be required to investigate this.

6.3.2. Medway Estuary SIP

Based on the results of the pilot study, further experiments were conducted on sediment from the Medway Estuary in order to investigate the potential for organic pathways to anammox. This primarily involved two experiments. The first involved a repeat of the pilot study, with the addition of unlabelled urea and DMA controls, on sediment collected from Medway Bridge Marina on 25/07/2011. This experiment also included microcosms amended with twice the amount of unlabelled or labelled substrate (*i.e.* 0.02 M) as, with experiments on *in situ* sediment, the amount of labelled organic substrate may not be sufficiently enriched compared with ambient concentrations to produce a significant signal. All microcosms were set-up in triplicate. The second experiment was set up on sediment collected from Medway Bridge Marina and Allington Lock on 15/03/2012. Both unlabelled and labelled controls were set-up with urea and DMA amended samples (final concentration was 0.01 M) as outlined previously. Samples were collected in the same month as the original, pilot study. Sediment was also collected from the Allington Lock site in order to investigate the ability of less saline anammox communities to utilise organic substrates. For full details of sampling locations refer to Figure 2.2.

	Density g/ml								
Fraction	**No add 1**	**No add 2**	**No add 3**	**^{12}C DMA 1**	**^{12}C DMA 2**	**^{12}C DMA 3**	**^{13}C DMA 1**	**^{13}C DMA 2**	**^{13}C DMA 3**
1	1.7325	1.7337	1.7314	1.7359	1.7258	1.7370	1.7427	1.7427	1.7416
2	1.7303	1.7314	1.7292	1.7337	1.7235	1.7348	1.7416	1.7404	1.7404
3	1.7269	1.7280	1.7258	1.7303	1.7202	1.7314	1.7382	1.7382	1.7382
4	1.7224	1.7247	1.7224	1.7269	1.7156	1.7280	1.7337	1.7348	1.7348
5	1.7179	1.7202	1.7179	1.7235	1.7123	1.7247	1.7303	1.7303	1.7303
6	1.7145	1.7168	1.7134	1.7190	1.7078	1.7202	1.7269	1.7269	1.7269
7	1.7100	1.7123	1.7089	1.7145	1.7044	1.7156	1.7224	1.7224	1.7224
8	1.7055	1.7066	1.6987	1.7111	1.7010	1.7123	1.7179	1.7179	1.7190
9	1.7010	1.7033	1.6841	1.7066	1.6965	1.7078	1.7145	1.7145	1.7145
10	1.6976	1.6987	1.6672	1.7021	1.6920	1.7044	1.7111	1.7111	1.7111
11	1.6942	1.6954	1.6334	1.6987	1.6897	1.7010	1.7055	1.7078	1.7078
12	1.6255	1.6605	1.5782	1.6785	1.6548	1.6920	1.5805	1.6672	1.6807
Mean	**1.7065**	**1.7110**	**1.6926**	**1.7151**	**1.7036**	**1.7174**	**1.7138**	**1.7212**	**1.7223**
Gradient	**-0.0040**	**-0.0040**	**-0.0084**	**-0.0039**	**-0.0038**	**-0.0038**	**-0.0038**	**-0.0037**	**-0.0036**
Intercept	**1.74**	**1.74**	**1.75**	**1.74**	**1.73**	**1.74**	**1.75**	**1.75**	**1.75**
R^2	**0.9973**	**0.9949**	**0.8303**	**0.9957**	**0.9975**	**0.9967**	**0.9939**	**0.9945**	**0.9914**

Table 6.2: Density of fractions obtained after fractionation of SIP (10 mM DMA) experiments set-up using Medway Bridge Marina sediment collected in 03/2012. All samples show a steady gradient between fractions indicating that density gradients formed satisfactorily during centrifugation and remained intact during fractionation. N.B. Fraction 12 was omitted from gradient calculations due to mixing of water and CsCl solution producing erroneous results (see discussion in main text). The density gradient of 'No Add 3' does not show a steady gradient but this is likely due to excessive mixing of water with the CsCl density gradient during fractionation. However the gradient between fractions 1 and 6 did show an even gradient and so it can only be presumed that the rest of the fractions would also, had water contamination not affected density measurements.

Samples were processed as outlined in section 6.2 although no attempt to sequence these samples was made (see below). Density measurements of individual fractions after fractionation indicated that density gradients had formed satisfactorily during centrifugation and these had remained intact during fractionation. Density gradients for ^{12}C and ^{13}C DMA amended microcosms set up in 03/2012, along with no addition controls, can be seen in Table 6.2 as an example. These data are representative of the typical data obtained from these experiments.

In general difficulties were encountered in the amplification of DNA via PCR from these samples. Even after two rounds of amplification, fractions from 07/2011 microcosms showed minimal amplification. Therefore DGGE gels from these samples exhibited poor resolution. However, regardless of poor band intensities, none of the DGGE gels clearly demonstrated any observable community change.

Experiments conducted on sediment collected in 03/2012 presented similarly inconclusive results. Again difficulties were encountered in the amplification of DNA via PCR from fractions obtained from these samples. There appeared to be insufficient quantities of target DNA in each fraction to obtain a suitable amplification product however, a nested PCR approach tended to result in amplification of all fractions. DNA was not expected in the densest fractions (approximately fractions 1-4) however these were still amplified after 2 rounds of PCR. Such ubiquitous amplification probably represented the amplification of minimal amounts of DNA, which perhaps adhered to the inside of the ultracentrifugation tube. This smearing effect made it difficult to separate ^{12}C and ^{13}C DNA, exacerbated by over-efficient PCR.

Considering that the density gradients had been shown to have formed satisfactorily during the SIP protocol, it can be assumed that the PCR product amplified from these fractions represents smearing inside the ultracentrifuge tube. Hence the amplification efficiency of the nested PCR may be too great, obscuring any real trends in the data by over-amplifying small quantities of DNA to similar levels as that of other fractions, limited by the availability of DNA polymerase.

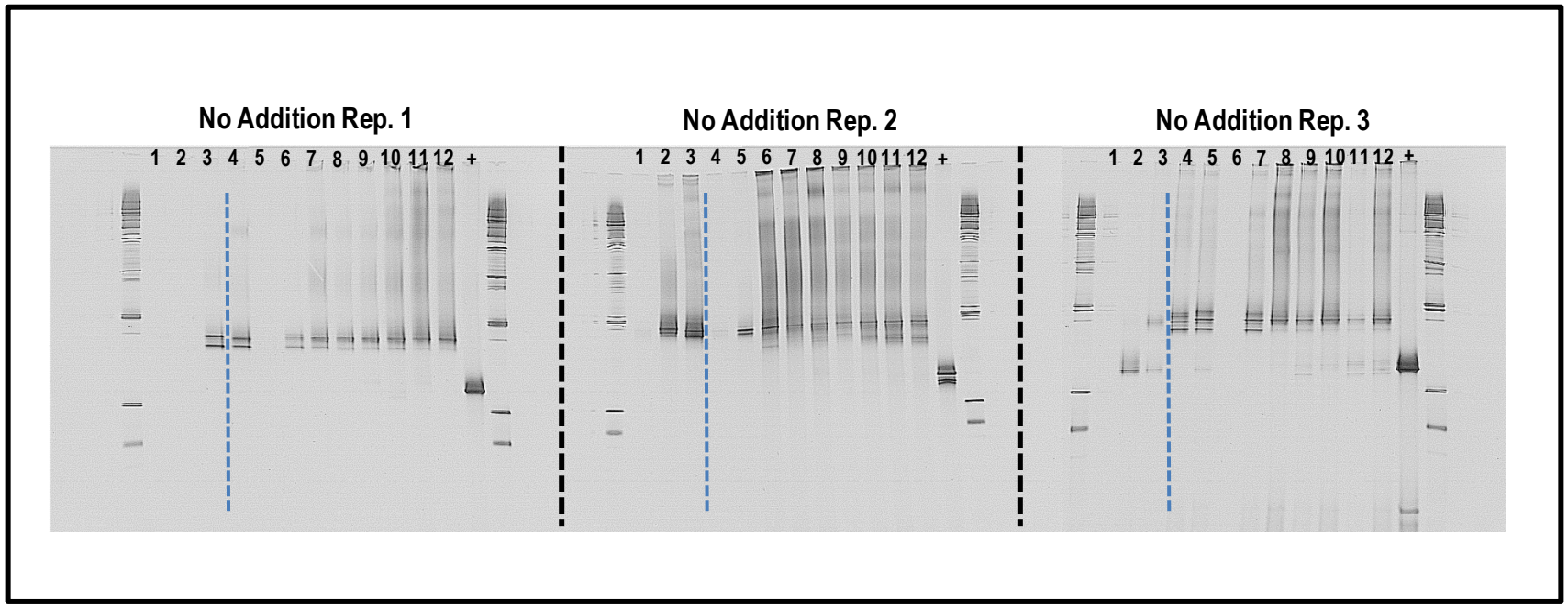


Figure 6.9: Anammox specific DGGE gel images for 03/2012 SIP experiment no addition controls. Fractions are labelled 1 to 12, corresponding with table MEDWAY_1 (i.e. Fraction 1 is the highest density and fraction 12 the lowest). Blue hashed lines represent the expected separation between ^{12}C and ^{13}C DNA bearing fractions. Amplifiable product was obtained from the majority of fractions. No obvious community change can be observed in any of the samples. Replicate 3 appears to show a greater diversity than replicates 1 and 2. The banding pattern seen in this replicate appears to be comparable to that seen in the pilot study (Figure **6.4**, Figure **6.5** and Figure **6.6**) with two distinct band clusters apparently representing Ca. Scalindua spp. and non-Scalindua anammox organisms.

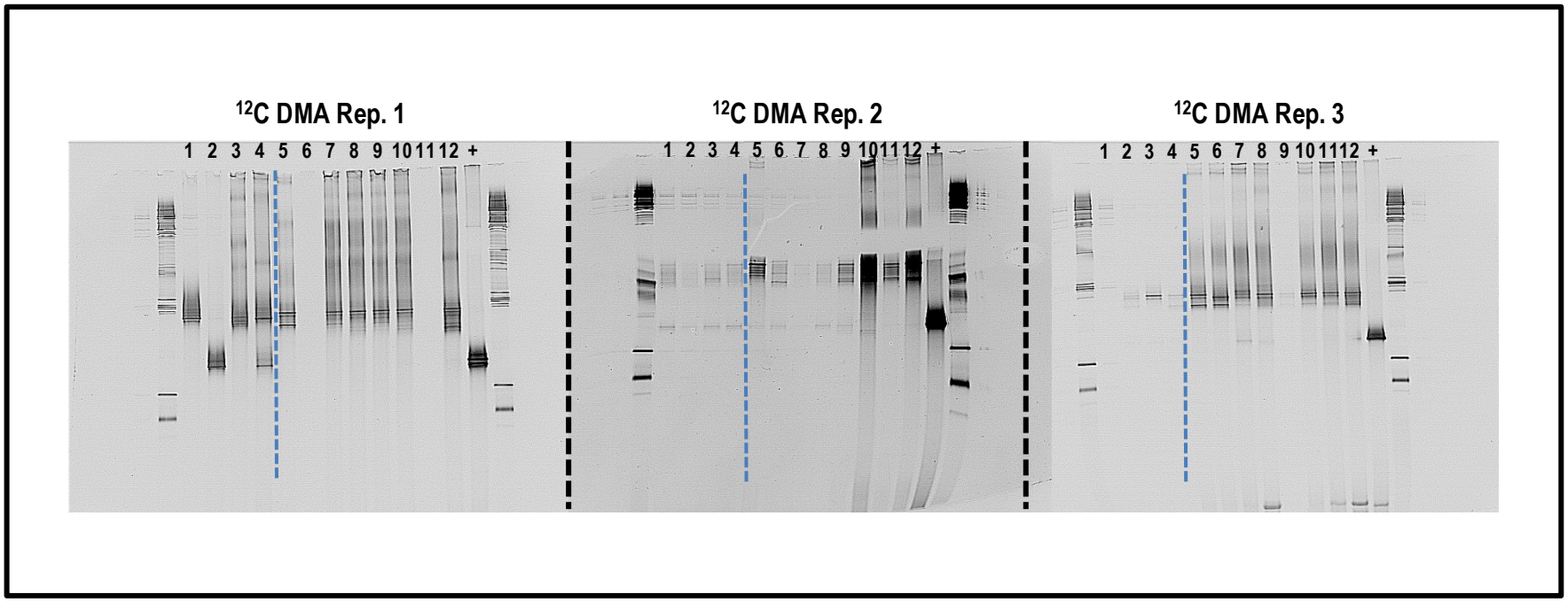


Figure 6.10: Anammox specific DGGE gel images for 03/2012 SIP experiment ^{12}C DMA (10 mM) controls. Fractions are labelled 1 to 12, corresponding with table MEDWAY_1 (i.e. Fraction 1 is the highest density and fraction 12 the lowest). Blue hashed lines represent the expected separation between ^{12}C and ^{13}C DNA bearing fractions. Most fractions were amplified after 2 rounds of PCR. ^{12}C amended samples appear to show a greater diversity than no addition controls, perhaps indicating an enrichment of anammox bacteria, though such conclusions are highly speculative.

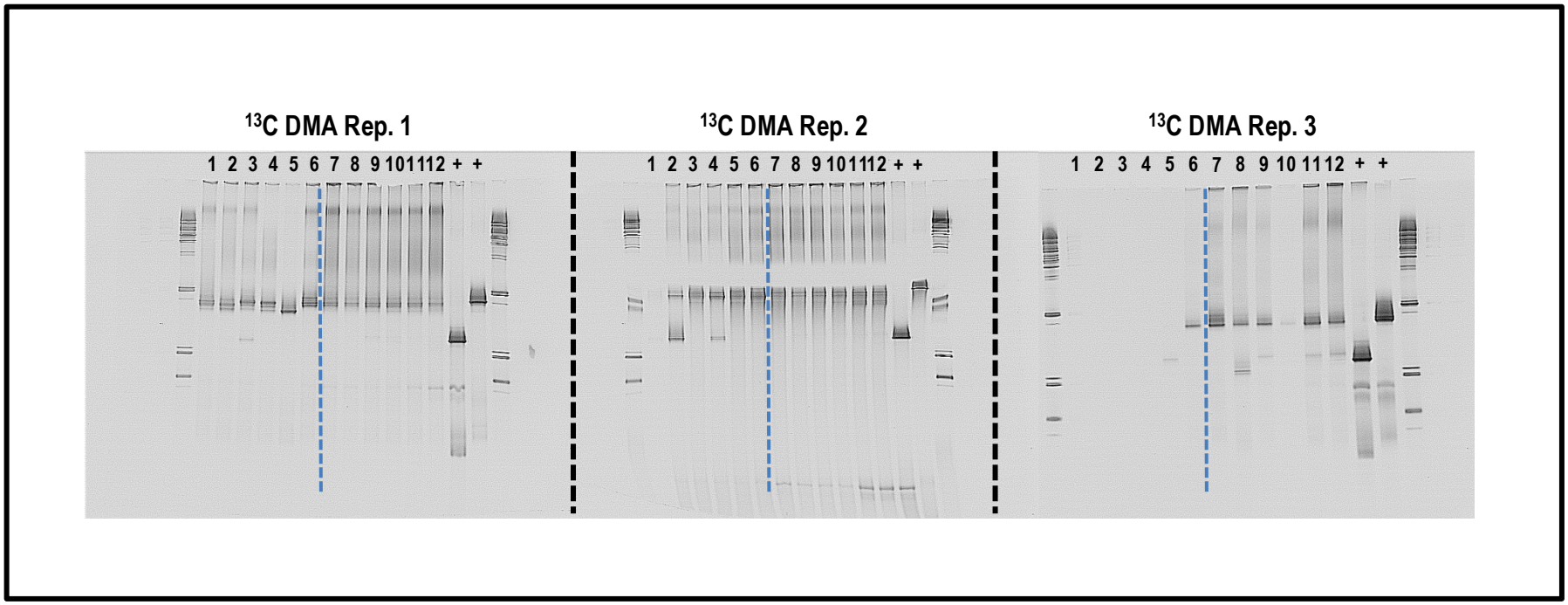


Figure 6.11: Anammox specific DGGE gel images for 03/2012 SIP experiment ^{13}C DMA (10 mM) amended microcosms. Fractions are labelled 1 to 12, corresponding with table MEDWAY_1 (i.e. Fraction 1 is the highest density and fraction 12 the lowest). Blue hashed lines represent the expected separation between ^{12}C and ^{13}C DNA bearing fractions. All but 1 fraction in replicates 1 and 2 were amplified by PCR. Thus no enrichment of ^{13}C labelled DNA was observable. Replicate 3 did not show ubiquitous amplification however, though it may demonstrate a definitive enrichment of ^{13}C DNA, it did not show an observable community change (as seen in the pilot study) and, considering the other two replicates, does not support a claim for the utilisation of organic substrates by anammox bacteria.

Nevertheless, none of the samples, when analysed via DGGE, demonstrated any obvious trend in the data in terms of enrichment of ^{13}C DNA or a change in the anammox population. Figure 6.9, Figure 6.10 and Figure 6.11 show typical data from samples amended with ^{12}C and ^{13}C DMA as well as no addition controls. Although microcosms inoculated with DMA appeared to show stronger amplification of anammox 16S rRNA genes (perhaps alluding to greater quantities of target DNA in these samples) no clear difference was observed between ^{13}C DMA amended samples and either ^{12}C DMA or no addition controls. Replicates neither show a clear enrichment of ^{13}C DNA nor any definitive change in the anammox community. These data are representative of the typical data obtained from this experiment.

6.3.3. ETNP SIP

Data obtained from similar experiments conducted on water from the ETNP produced similar results. In general very low quantities of DNA were gleaned from microcosms which proved challenging when attempting SIP. Nevertheless, amplifiable product was obtained from fractions after the use of a nested PCR approach. Furthermore, density measurements from these fractions demonstrated that an even gradient had formed during centrifugation and had remained intact during fractionation. However, analysis of the microbial community via DGGE neither showed enrichment of ^{13}C-bearing fractions compared with ^{12}C controls nor any observable community change over these fractions.

6.4. Discussion

6.4.1. Potential for Organic Pathways to Anammox

Data from the pilot study appeared to suggest that anammox bacteria may be able to assimilate organic carbon directly, similar to hypothesis b) described in sections 1.3.2.2.2 and 6.1.3. To date such a discovery of heterotrophic anammox has not been reported and would be a completely novel metabolism of anammox organisms. Both urea and DMA amended microcosms appeared to show an enrichment of heavy labelled DNA in two out of three replicates. Furthermore, a community change within the anammox population was evident between ^{12}C and ^{13}C labelled fractions, potentially indicating a separation between anammox bacteria capable of

assimilating organic carbon and those which could not. Further analysis of these communities, via the sequencing of bands obtained from DGGE, indicated that light DNA fractions were comprised of organisms representing *Ca.* Scalindua spp. whereas heavy fractions contained organisms representing *Ca.* Brocadia spp. and *Ca.* Kuenenia spp. These findings were collaborated with 454 pyrosequencing data. These data potentially suggest that *Ca.* Brocadia spp. and *Ca.* Kuenenia spp. have a more diverse metabolism than *Ca.* Scalindua spp., being capable of heterotrophic growth. This is in direct contradiction to findings previously reported in the literature (van Niftrik and Jetten, 2012). This may indicate a change in metabolic capabilities between marine anammox organisms (*i.e. Ca.* Scalindua spp.) and terrestrial/freshwater anammox organisms which have gained more versatile metabolic capabilities in order to survive and compete in more diverse environments.

It could be argued that anammox may simply be assimilating CO_2 which has been produced by the breakdown of organic compounds by other organisms as has been suggested by Güven, *et al.* (2005) and Russ, *et al.* (2012), hence retaining their autotrophic lifestyle. Thus anammox DNA may come to contain ^{13}C via a similar feed-back loop during such an experiment. However these data do not support this claim as, if this were true, although there would be an enrichment of ^{13}C DNA, there would be no observed change in the anammox community as different anammox organisms would not preferentially assimilate heavy and light CO_2. Thus, if these data represent a true phenomenon, then they must represent anammox organisms capable of heterotrophy.

However, the data cannot support a definitive conclusion as to the potential for organic pathways to anammox. The trends discussed above were only observed clearly in one replicate of the DMA amended sample and two replicates of those with urea. Although the absence of such trends do not disprove the hypothesis that anammox are capable of a mixotrophic metabolism, the lack of reproducibility does shed doubt on the accuracy of such findings. Furthermore, as this experiment was only a pilot study (in order to discover potential problems with the methodology) a comprehensive suite of controls was not conducted and so observations cannot be corroborated in this way. Sequence data, from the sequencing of DGGE bands and via 454 data, appear to suggest that a community change between heavy and light

DNA fractions was observable however, the low number of clones (n=14) and lack of replication in these respective analyses mean that any conclusions drawn from these data must be used with caution. As such, a more comprehensive experiment was needed in order to confirm the findings of this pilot study.

The first experiment, conducted on sediment collected in 07/2011, did not show any observable trends in the data, similar to those seen in the data of the pilot study. No enrichment or discernible community changes were seen across any samples. However, it was speculated that this may be due to low or non-existent anammox activity during this period as anammox rates have been shown to be low along the estuary during the summer (Trimmer*, et al.*, 2005).

Therefore a similar experiment was conducted on sediment collected from the same month as the pilot experiment (*i.e.* March). However, this experiment produced similar data as the first experiment, with no discernible trends, either in enrichment or a change in anammox diversity being observed across these samples. Sediment from a less saline site along the Medway estuary (*i.e.* Allington Lock) was also investigated during this study as findings from the pilot study suggested than non-Scalindua anammox organisms may be more capable of heterotrophy and these organisms were more numerous upstream along the estuary (see chapter 5). However, microcosms of sediment from this site also failed to show evidence that anammox organisms had assimilated carbon from organic compounds.

A final experiment was conducted on water from the ETNP OMZ, an environment similar to that in which the findings of Trimmer and Purdy (2012) were made, which was the foundation of the hypotheses of this study. However data from this experiment also failed to demonstrate any evidence that anammox bacteria had utilised urea or TMA.

The above data do not support the hypothesis that anammox bacteria could potentially utilise organic substrates such as urea or methylamines. However the observations made during the pilot study do suggest that some anammox organisms may have the potential to utilise DMA and urea in Medway Estuary sediments, though these findings could not be supported statistically. Furthermore, the experiments conducted after the pilot experiment failed to demonstrate any evidence

for the utilisation of organic carbon by anammox bacteria. Nevertheless, in light of such observations and the potential significance that organammox may have on our understanding of global N cycling, the hypotheses outlined in section 1.3.2.2.2 warrant further investigation. To this end it is imperative that the shortcomings of the investigation presented in this chapter are analysed and subsequent improvements to the methodology are made.

6.4.2. Methodological Shortcomings

6.4.2.1. Potential Reasons for the Lack of Positive Data

It is entirely feasible that the lack of solid positive evidence obtained during this investigation, into the presence of an organammox like reaction in estuarine and OMZ environments, is because such a process does not exist and that the hypotheses (section 1.3.2.2.2 and 6.1.3) are false. As such the positive observations observed during the pilot study may be false positives. However, following the analysis of these experiments there are a significant number of issues which have arisen regarding the methodology employed which could be improved. Therefore it would be prudent to repeat the experiment with such improvements before rejecting the hypotheses, especially considering the potential importance to global geochemical cycling of such a process (Trimmer and Purdy, 2012).

Anammox bacteria are typically of low abundance in the environment (Kuypers, *et al.*, 2003; Hamersley, *et al.*, 2007) and thus the yield of anammox DNA obtained from these experiments, especially those from the ETNP OMZ, would be low. This may have hindered analysis of SIP experiments as typically high yields of DNA are required for SIP experiments as large quantities of DNA can be lost during fractionation and precipitation (Neufeld, *et al.*, 2007). Lower yields of DNA would have weakened the anammox signal and potentially impaired downstream analysis (*e.g.* DGGE).

The slow growth rates of anammox bacteria (Jetten, *et al.*, 2009) also may have contributed to difficulties in analysing the results of these experiments as the rate of which ^{13}C was incorporated into anammox cells would have been slow. Therefore there would be lower concentrations of ^{13}C labelled DNA which would make fractionation and identification of this fraction difficult. However, without a

comprehensive knowledge of the potential metabolic rates of these organisms this is an issue which would be difficult to resolve.

In addition there are a number of issues associated with using SIP for explorative microbiological experiments which may have hindered this investigation (see below).

6.4.2.2. Using SIP to Investigate Anammox in Complex Communities

SIP can be a powerful technique to investigate microbial ecology allowing for the fate of carbon and nitrogen from the environment to be traced into microbial biomass (Manefield, *et al.*, 2002; Chen and Murrell, 2010). However the technique is not without its caveats, especially when investigating complex *in situ* microbial communities.

SIP experiments conducted on pure or enrichment cultures have often led to a successful conclusion. Such organisms have frequently been well characterised and as such the conditions favourable for growth have been extensively investigated. Furthermore, in pure culture, organisms are not presented with the problems of environmental stress or competition. Characterisation of such organisms also provides information as to their metabolic rates, especially in terms of the assimilation of carbon and nitrogen into biomass and mass balance calculations. With knowledge of the organisms involved in the system of choice the exact densities of light and heavy DNA (taking into account assimilation rates and GC content) which can be expected after centrifugation can be calculated, allowing for more efficient targeting of ^{12}C and ^{13}C bearing fractions. Comprehensive data of metabolic rates can also aid decisions as to the incubation time of experiments, allowing for the maximum amount of carbon or nitrogen incorporation with the minimum amount of cross-feeding into other organisms. This said, with the knowledge gained from such experiments, experiments conducted on the micro-organisms of interest *in situ* are likely to be more successful.

However, if such investigations into the metabolic capabilities of the organism in question cannot be completed (as is the case with anammox bacteria) then SIP methodologies lose some of their power. Without knowledge of the assimilation rates and toxicity of compounds to be investigated, it is difficult to ensure that a suitably sufficient concentration of labelled substrate has been amended to the

samples which is imperative for a successful experiment (Manefield, *et al.*, 2002). Furthermore without such data, it is impossible to determine the exact location of heavy and light DNA after fractionation except by subjective inference. This is further complicated as, with complex microbial communities, the varying molecular weights of DNA from different organisms is unlikely to produce a clear separation between labelled and unlabelled DNA. As such, SIP is not a particularly efficient tool for explorative microbial ecology as the scientist ideally needs to know something of the organisms which they wish to investigate.

Further complications of the SIP methodology also have the potential to limit its effectiveness. Despite due care and attention and good laboratory practice, it is difficult to ensure that DNA does not smear along the sides of the ultracentrifuge tube during centrifugation. DNA can adhere to the surface of the ultracentrifuge tube and therefore, may not appear in the correct fraction during fractionation. This can produce a "smearing" effect across fractions during molecular fingerprinting techniques, making analysis difficult. This problem can be exacerbated when dealing with low yields of DNA (as in this study) and over-efficient PCR. In addition, investigation of mixed, microbial communities using SIP can further exacerbate the problem of determining a clear labelled and unlabelled fraction due to varying AT/GC concentrations (and hence molecular masses), cross-feeding and dilution (Cadisch, *et al.*, 2005).

As such, it is impossible to determine whether a "negative" result (*i.e.* no clear enrichment of heavy DNA or noticeable community change across fractions) is an accurate representation of the microcosm or due to difficulties arising from the SIP methodology. SIP can also resist the effective use of statistics to explore observations within the data as, without the ability to ensure that fractionation is consistent throughout replicates (*i.e.* to ensure that the exact same density ranges exist in the same fraction between different replicates), it is difficult to compare samples and fractions. Thus subjective observations must be made on trends observed across samples and fractions which are not conducive to robust statistical methods.

Therefore, it must be questioned whether the above methodology is the best approach to investigate the potential for mixotrophy/heterotrophy in anammox

bacteria. SIP may not be the best tool for the exploration of novel, unknown pathways as a certain degree of prior knowledge is required in order to plan a successful SIP experiment. As the existence of such a pathway in anammox bacteria is yet to be investigated (other than the findings presented here), the design of such an experiment is based on completely hypothetical assumptions, from observations made in other organisms, which may or may not be valid. That is not to say that SIP would not be useful technique for future investigations, but rather other experiments should be conducted initially.

6.4.3. Further Research

Considering the limitations of the methodology used in this study (discussed above) a number of improvements should be made to further investigations of the hypotheses in order to be able to draw more definitive conclusions.

Future studies should focus initially on determining the loss of organic substrates in such environments. This would allow for a greater understanding of the concentrations of substrates to use during these experiments and a better assessment of incubation times. The concentration of substrates such as urea and methylamines should be monitored over the incubation period in order to assess the microbial population's ability to metabolise these compounds. A number of different calorimetric assays are available for the measurement of urea (*e.g.* Quantichrom™, BioAssay Systems, USA) though further investigation would be required to assess whether such assays are compatible with the substrates being investigated. Concentrations of methylamines in the environment have previously been measured using GC (King, 1984) and IC (Gibb, *et al.*, 1995). This data could be collaborated with measurements of CH_4 and CO_2 from headspace gases in order to assess the fate of carbon in these microcosms. Killed controls should also be used to determine how much of the conversion of these amended substrates are biologically mediated.

Anammox bacteria and activity should also be quantified, *e.g.* by qPCR, FISH and IPT (using ^{15}N labelled urea and DMA/TMA), in order to determine whether the incorporation of organic compounds can stimulate the growth and activity of these organisms. Such an initial investigation would provide a better understanding of the

fate of such organic compounds in these environments and the rate at which anammox bacteria can potentially utilise them.

With a greater knowledge of the fate and utilisation rates of urea and methylamines in anaerobic estuarine sediments and OMZs, subsequent SIP experiments are likely to be more successful. Due to the problems associated with replication of the results presented in section 6.3, it may be prudent to increase the number of replicates of future SIP experiments investigating the potential for anammox and also to include technical replicates for these microcosms (*i.e.* repeated fractionation of DNA from the same microcosm). Although increasing the number of replicates would increase the cost (both financially and in terms of time) a greater number of replicates would improve the statistical validity to any conclusions drawn from future experiments.

In addition to DNA-SIP, RNA-SIP should also be attempted. It has been suggested that RNA may be a more responsive biomarker for SIP experiments due to the higher turn-over rates and greater copy numbers of RNA (Manefield, *et al.*, 2002). Therefore RNA-SIP may improve the strength of the recovered ^{13}C signal after fractionation and aid downstream analysis.

A major limitation of SIP experiments is the difficulty in accurately determining which fractions represent the heavy and light labelled DNA/RNA. This problem may be exacerbated when investigating complex microbial communities (Cadisch, *et al.*, 2005). Therefore, the fingerprinting (DGGE) and theoretical (based on the molecular weight of DNA) techniques used in this study to pinpoint ^{13}C DNA may not be an accurate method for determining the location of ^{13}C DNA. Alternatively ethidium bromide staining (Neufeld, *et al.*, 2007; Chen and Murrell, 2010) could be used to more accurately locate ^{13}C DNA. Previous studies have also suggested the use of IRMS (Manefield, *et al.*, 2002; Chen and Murrell, 2010) or HPLC/IRMS (Cadisch, *et al.*, 2005) to investigate whether labelled DNA/RNA had been enriched above natural abundance and identify ^{13}C bearing DNA fractions. However the range of molecular weights of DNA associated with complex communities may also hinder the isolation of ^{13}C labelled DNA using this methods. It is possible that flow cytometry (Amann, *et al.*, 1990) or associated methodologies such as FLOW-FISH (Baerlocher, *et al.*, 2006) could be used to isolate anammox organisms (or other organisms of interest) from the ambient microbial community which would aid the

isolation of ^{13}C labelled DNA. Methodologies such as Raman-FISH (Huang, *et al.*, 2007) could also be utilised in conjunction with SIP in order to investigate bacterial cells which had incorporated labelled carbon from organic sources and would allow for the quantification of and enumeration of these bacteria.

The above improvements to the methodology used in this study would allow for a more targeted approach to investigate whether anammox bacteria could utilise such compounds directly or via syntrophic relationships and lead to a more rigorous testing of the hypotheses.

6.5. Summary

- Anammox appear to have a greater metabolic diversity than previously thought.
- Observations within the literature appear to suggest that anammox bacteria can utilise organic substrates either catabolically or anabolically.
- Anammox organisms may either use such organic substrates directly or via syntrophic reactions with other organisms.
- The data presented in this study fail to conclusively show anammox organisms to be able to utilise methylamines or urea, though some evidence eluded to the possibility of such a pathway to the anammox reaction.
- Further, in-depth studies are required in order to elucidate the possibility of such a reaction.
- Such further studies should focus initially on how much of such substrates are lost to the environment and whether anammox activity or cell numbers increase during incubation.

7. Concluding Remarks

Anammox bacteria are clearly organisms of great environmental significance and the anammox process has been shown to be an important sink of fixed nitrogen in anaerobic environments, potentially rivalling denitrification in its contribution to N_2 production. Anammox organisms belong to a deep branching, monophyletic clade comprising of only five genera, perhaps surprising concerning that their environmental rivals, denitrifiers, demonstrate a far greater diversity spanning different domains of life. Anammox organisms appear to be ubiquitous in anaerobic environments, where O_2 concentrations are low and DIN (and potentially DOC/DON) concentrations are high. As such an interesting parallel can be drawn between the specialist denitrifiers (with many different organisms occupying specific niches) and anammox bacteria which appear to be generalists. However, it has been suggested that anammox organisms also have a specific niche adaptation, the details of which are unknown. In fact, despite a relatively low richness (compared with other environmentally significant organisms), ubiquitous nature and monophyletic characteristics, very little is known about the diversity and distribution of these unique organisms and their controlling environmental factors. This study attempted to elucidate some of these discrepancies and provide a greater understanding of anammox bacteria in the natural environment.

The data presented in this study provide an unprecedented level of detail to intra-anammox diversity in estuarine and OMZ environments due to the use of high-throughput sequencing technologies with robust and validated PCR primers, which specifically target the entirety of the known anammox 16S rRNA diversity.

A clear shift in the anammox community was observed along the extent of the Medway River Estuary, from the freshwater end (Allington Lock) to the saline mouth of the estuary (Hoo Marina). Freshwater sites were dominated by *Ca.* Brocadia spp. whereas saline sites were dominated by *Ca.* Scalindua spp. Brackish sites, along the middle of the estuary, demonstrated a greater degree of diversity and the anammox community was evenly split between *Ca.* Scalindua and non-Scalindua organisms (though the relative abundance of non-Scalindua and *Ca.* Scalindua organisms gradually increased upstream and downstream respectively). Evidence for the presence of *Ca.* Jettenia spp. and *Ca.* Kuenenia spp. was also observed in the

data, although organisms from these genera were typically of low relative abundance. *Ca.* Scalindua wagneri organisms appeared to inhabit brackish environments with salinities ranging from approximately 3 to 15 ppt and were absent in both saline and freshwater environments. As such, the data describe a pronounced change in anammox diversity along the estuary which appeared to be correlated with salinity. However, the data also suggested that the diversity of anammox organisms was also correlated to increasing concentrations of DIN and nutrients upstream. As such, changes in salinity are unlikely to account entirely for the observed anammox diversity but instead contribute to the consortia of environmental factors which affect anammox distribution and community structure. However, further research is required to fully comprehend the niche adaptations of anammox bacteria in such environments.

Sequencing data obtained from the ETNP OMZ indicate that the anammox community in this environment was dominated by *Ca.* Scalindua spp. as suggested in the literature. However, non-Scalindua organisms were also detected in the OMZ. These organisms, though significant, were generally of low relative abundance and appeared to inhabit only a few depths within the OMZ. However no observable trend was observed within these data to explain this distribution. While no substantial change occurred in the anammox data along the latitudinal transect, a change in anammox diversity and community structure appeared to exist within the *Ca.* Scalindua clade at different depths. Two main clusters of *Ca.* Scalindua spp. were observed from the data. One of these clusters, showing less similarity to known *Ca.* Scalindua sequences from the literature than the other cluster, dominated the community in the upper oxycline and upper core of the OMZ, where O_2 concentrations were found to be greater and potentially more transient. The core of the OMZ was dominated by organisms showing a close similarity to *Ca.* S. brodae, *Ca.* S. marina and *Ca.* S. sorokinii. Data from the lower oxycline (*i.e.* approaching ambient oceanic O_2 concentrations below the OMZ) appeared to demonstrate that the first cluster of *Ca.* Scalindua spp., observed in the upper oxycline, had begun to re-populate the anammox community; however a greater resolution of the data, sampling deeper sites, would be required to confirm this observation. No sequences were obtained from the data which demonstrated a high degree of similarity to *Ca.* S. wagneri organisms, perhaps suggesting that these organisms are indeed adapted to

brackish or mesohaline environments. There appeared to be no significant changes in the anammox community over a longitudinal scale, only with depth.

Despite such clear trends in anammox diversity in the OMZ, no significant correlation could be observed with any environmental data. However, the community changes observed in the sequencing data did appear to suggest that two phylogenetically distinct groups of *Ca.* Scalindua spp. were observed; one of which appeared to be adapted to anoxic conditions within the core of the OMZ while the other was potentially adapted to higher, transient O_2 concentrations which flank the core of the OMZ. A similar, previously unidentified group of *Ca.* Scalindua spp. had been observed in OMZ environments (Woebken, *et al.*, 2008; Galan, *et al.*, 2009), however the data presented in this study confirm these previous observations (made from scant data) and are the first to demonstrate that such a cluster may be environmentally significant and to speculate as to the niche adaptation of this cluster. Future research should be directed in observing the change in the anammox community below the OMZ and to sample the transition between the oxycline and core of the OMZ in greater detail; as it appears to be here that anammox communities are likely to be the most dynamic and hence would provide the greatest source for increasing our understanding of the diversity, distribution and niche adaptations of these organisms. Investigation of the temporal and global changes in anammox diversity in such environments would also be of great benefit to understanding anammox in OMZ environments.

Investigation of the potential for anammox to utilise organic substrates in the natural environment was also attempted using labelled stable isotopes (^{13}C). Some initial evidence was obtained which suggested that anammox organisms in Medway Estuary sediments (Medway Bridge Marina) were capable of assimilating carbon from organic compounds such as urea and DMA into biomass. This was only observed in non-Scalindua anammox organisms and *Ca.* Scalindua spp. were not shown to assimilate organic carbon. Such a metabolism would be novel in anammox bacteria, which are previously defined as strict chemolithoautotrophs. It could be suggested that these data did not represent direct assimilation of organic carbon into anammox biomass, but rather the conversion of DOC to CO_2 which in turn was fixed via the conventional anammox pathway. However this is unlikely, as if this were true,

a distinction between ^{13}C-labelled non-Scalindua and ^{12}C-labelled *Ca.* Scalindua organisms would not be observed as there would be no preferential uptake of $^{12}CO_2$ and $^{13}CO_2$.

However, despite initial success in the investigation of this hypothesis, subsequent experiments failed to show any evidence for such a process. These data make the results from this investigation inconclusive and unable to prove or disprove the hypotheses. As such further investigations into the potential for organammox (such as investigating the fate of DON or potential syntrophic partners) were not possible.

However, the data appear to suggest, if anammox were capable of such a process, that the potential for organammox would be greater towards the freshwater end of the estuary, where non-Scalindua anammox organisms were more abundant. Experiments conducted on water from the ETNP OMZ demonstrated no evidence that such a process occurred, which corroborates with the above conclusion as the OMZ was dominated by *Ca.* Scalindua spp. However, this is in direct contradiction of the observations of Trimmer and Purdy (2012), which led to this investigation, who demonstrated the potential for this process in the Arabian Sea OMZ.

Nevertheless, the fact that some data, however sparse, were obtained which may suggest that organammox may occur, suggests that the potential for organic pathways to the anammox reaction warrants further investigation, especially considering the reported genomic diversity and environmental significance of these organisms. The lack of cultured isolates presents significant difficulties to such future research and this investigation suggests that a SIP-based methodology may not be the best approach to investigating these hypotheses (as SIP does not appear to be a strong enough tool for explorative microbiology). However the difficulty of the task and abundance of the problems facing further investigations should not dissuade such attempts and the risk of failure should not affect future research. Future investigations may first begin by attempting to clarify the seasonal and spatial variability in anammox bacteria and their activity so as to be able to ensure that further microcosm experiments are at least targeting anammox organisms that are currently active. Furthermore, the general impact of the introduction of large quantities of organic substrates to the anammox community (in terms of increase or decrease in biomass or transcription) should also be investigated prior to more in

depth studies. By adopting such approaches and obtaining a more lucid understanding of the environmental controls of anammox bacteria, future investigations into the potential for organammox may have greater success than the explorational investigation presented in this thesis.

Regardless of the potential caveats and limitations associated with the conclusions drawn from the data obtained in this investigation, this thesis represents a significant increase to our knowledge and understanding of the ecology of anammox organisms. It provides an unprecedented level of detail as to the diversity and distribution of anammox organisms across environmental gradients from two environmentally important environments. It further speculates and provides some (however limited) data on the potential for novel metabolisms of these organisms which would revolutionise the understanding of the roles of anammox bacteria in relation to global biogeochemical cycling.

8. References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R. and Stahl, D.A. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and Environmental Microbiology* **56**, 1919-1925.

Amano, T., Yoshinaga, I., Okada, K., Yamagishi, T., Ueda, S., Obuchi, A., Sako, Y. and Suwa, Y. (2007). Detection of anammox activity and diversity of anammox bacteria-related 16S rRNA genes in coastal marine sediment in japan. *Microbes and Environments* **22**, 232-242.

Amano, T., Yoshinaga, I., Yamagishi, T., Chu, V.T., Pham, T.T., Ueda, S., Kato, K., Sako, Y. and Suwa, Y. (2011). Contribution of Anammox Bacteria to Benthic Nitrogen Cycling in a Mangrove Forest and Shrimp Ponds, Haiphong, Vietnam. *Microbes and Environments* **26**, 1-6.

An, S.M. and Gardner, W.S. (2002). Dissimilatory nitrate reduction to ammonium (DNRA) as a nitrogen link, versus denitrification as a sink in a shallow estuary (Laguna Madre/Baffin Bay, Texas). *Marine Ecology-Progress Series* **237**, 41-50.

Arbeli, Z. and Fuentes, C.L. (2007). Improved purification and PCR amplification of DNA from environmental samples. *Fems Microbiology Letters* **272**, 269-275.

Bae, H., Chung, Y.C. and Jung, J.Y. (2010). Microbial community structure and occurrence of diverse autotrophic ammonium oxidizing microorganisms in the anammox process. *Water Science and Technology* **61**, 2723-2732.

Baerlocher, G.M., Vulto, I., de Jong, G. and Lansdorp, P.M. (2006). Flow cytometry and FISH to measure the average length of telomeres (flow FISH). *Nature protocols* **1**, 2365-2376.

Bauer, S., Hitchcock, G.L. and Olson, D.B. (1991). Influence of monsoonally-forced Ekman dynamics upon surface layer depth and plankton biomass distribution in the Arabian Sea. *Deep Sea Research Part A. Oceanographic Research Papers* **38**, 531-553.

Boumann, H.A., Longo, M.L., Stroeve, P., Poolman, B., Hopmans, E.C., Stuart, M.C.A., Damste, J.S.S. and Schouten, S. (2009). Biophysical properties of membrane lipids of anammox bacteria: I. Ladderane phospholipids form highly organized fluid membranes. *Biochimica Et Biophysica Acta-Biomembranes* **1788**, 1444-1451.

Brakenhoff, R.H., Schoenmakers, J. and Lubsen, N.H. (1991). Chimeric cDNA clones: a novel PCR artifact. *Nucleic acids research* **19**, 1949.

Brandsma, J., van de Vossenberg, J., Risgaard-Petersen, N., Schmid, M.C., Engstrom, P., Eurenius, K., Hulth, S., Jaeschke, A., Abbas, B., Hopmans, E.C., Strous, M., Schouten, S., Jetten, M.S.M. and Damste, J.S.S. (2011). A multi-proxy study of anaerobic ammonium oxidation in marine sediments of the Gullmar Fjord, Sweden. *Environmental Microbiology Reports* **3**, 360-366.

Britten, R.J., Rowen, L., Williams, J. and Cameron, R.A. (2003). Majority of divergence between closely related DNA samples is due to indels. *Proceedings of the National Academy of Sciences* **100**, 4661-4665.

Brock, T.D. (1966) *Principles of Microbial Ecology*. ed. Englewood Cliffs, NJ, USA: Prentice-Hall, Inc.

Broda, E. (1977). 2 Kinds of Lithotrophs Missing in Nature. *Zeitschrift Fur Allgemeine Mikrobiologie* **17**, 491-493.

Bulow, S.E., Rich, J.J., Naik, H.S., Pratihary, A.K. and Ward, B.B. (2010). Denitrification exceeds anammox as a nitrogen loss pathway in the Arabian Sea oxygen minimum zone. *Deep-Sea Research Part I-Oceanographic Research Papers* **57**, 384-393.

Byrne, N., Strous, M., Crepeau, V., Kartal, B., Birrien, J.-L., Schmid, M., Lesongeur, F., Schouten, S., Jaeschke, A., Jetten, M., Prieur, D. and Godfroy, A. (2008). Presence and activity of anaerobic ammonium-oxidizing bacteria at deep-sea hydrothermal vents. *Isme Journal* **3**, 117-123.

Cadisch, G., Espana, M., Causey, R., Richter, M., Shaw, E., Morgan, J.A.W., Rahn, C. and Bending, G.D. (2005). Technical considerations for the use of N-15-DNA stable-isotope probing for functional microbial activity in soils. *Rapid Communications in Mass Spectrometry* **19**, 1424-1428.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J. and Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335-336.

Chang, B.X., Devol, A.H. and Emerson, S.R. (2012). Fixed nitrogen loss from the eastern tropical North Pacific and Arabian Sea oxygen deficient zones determined from measurements of N-2:Ar. *Global Biogeochemical Cycles* **26**,

Chen, Y. and Murrell, J.C. (2010). When metagenomics meets stable-isotope probing: progress and perspectives. *Trends in Microbiology* **18**, 157-163.

Clark, I., Timlin, R., Bourbonnais, A., Jones, K., Lafleur, D. and Wickens, K. (2008). Origin and fate of industrial ammonium in anoxic ground water—15N evidence for anaerobic oxidation (anammox). *Ground Water Monitoring & Remediation* **28**, 73-82.

Dale, O.R., Tobias, C.R. and Song, B. (2009). Biogeographical distribution of diverse anaerobic ammonium oxidizing (anammox) bacteria in Cape Fear River Estuary. *Environmental microbiology* **11**, 1194-1207.

Dalsgaard, T., Canfield, D.E., Petersen, J., Thamdrup, B. and Acuna-Gonzalez, J. (2003). N_2 production by the anammox reaction in the anoxic water column of Golfo Dulce, Costa Rica. *Nature* **422**, 606-608.

Dalsgaard, T. and Thamdrup, B. (2002). Factors controlling anaerobic ammonium oxidation with nitrite in marine sediments. *Applied and Environmental Microbiology* **68**, 3802-3808.

Dalsgaard, T., Thamdrup, B., Farias, L. and Revsbech, N.P. (2012). Anammox and denitrification in the oxygen minimum zone of the eastern South Pacific. *Limnology and oceanography* **57**, 1331-1346.

Dang, H.Y., Chen, R.P., Wang, L., Guo, L.Z., Chen, P.P., Tang, Z.W., Tian, F., Li, S.Z. and Klotz, M.G. (2010). Environmental Factors Shape Sediment Anammox Bacterial Communities in Hypernutrified Jiaozhou Bay, China. *Applied and Environmental Microbiology* **76**, 7036-7047.

Devol, A.H. (2003). Nitrogen cycle - Solution to a marine mystery. *Nature* **422**, 575-576.

Devol, A.H., Naqvi, S.W.A. and Codispoti, L.A. (2006) Nitrogen cycling in the suboxic waters of the Arabian Sea. In *Past and Present Water Column Anoxia*, vol. 64, pp. 283-310. Editted by Neretin, L.N.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **14**, 927-930.

Dong, L.F., Smith, C.J., Papaspyrou, S., Stott, A., Osborn, A.M. and Nedwell, D.B. (2009). Changes in benthic denitrification, nitrate ammonification, and anammox process rates and nitrate and nitrite reductase gene abundances along an estuarine nutrient gradient (the Colne Estuary, United Kingdom). *Applied and Environmental Microbiology* **75**, 3171-3179.

Dong, L.F., Sobey, M.N., Smith, C.J., Rusmana, I., Phillips, W., Stott, A., Osborn, A.M. and Nedwell, D.B. (2011). Dissimilatory reduction of nitrate to ammonium, not denitrification or anammox, dominates benthic nitrate reduction in tropical estuaries. *Limnology and oceanography* **56**, 279-291.

Dowd, S.E., Callaway, T.R., Wolcott, R.D., Sun, Y., McKeehan, T., Hagevoort, R.G. and Edrington, T.S. (2008). Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *Bmc Microbiology* **8**,

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797.

Engstrom, P., Dalsgaard, T., Hulth, S. and Aller, R.C. (2005). Anaerobic ammonium oxidation by nitrite (anammox): Implications for N-2 production in coastal marine sediments. *Geochimica Et Cosmochimica Acta* **69**, 2057-2065.

Engstrom, P., Penton, C.R. and Devol, A.H. (2009). Anaerobic ammonium oxidation in deep-sea sediments off the Washington margin. *Limnology and oceanography* **54**, 1643-1652.

Fitzsimons, M.F., Kahni-danon, B. and Dawitt, M. (2001). Distributions and adsorption of the methylamines in the inter-tidal sediments of an East Anglian Estuary. *Environmental and Experimental Botany* **46**, 225-236.

Francis, C.A., Beman, J.M. and Kuypers, M.M.M. (2007). New processes and players in the nitrogen cycle: the microbial ecology of anaerobic and archaeal ammonia oxidation. *Isme Journal* **1**, 19-27.

Fu, L., Niu, B., Zhu, Z., Wu, S. and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152.

Fuerst, J.A. (2005). Intracellular compartmentation in planctomycetes. *Annual Review of Microbiology* **59**, 299-328.

Fuerst, J.A. and Sagulenko, E. (2011). Beyond the bacterium: planctomycetes challenge our concepts of microbial structure and function. *Nature Reviews Microbiology* **9**, 403-413.

Galan, A., Molina, V., Belmar, L. and Ulloa, O. (2012). Temporal variability and phylogenetic characterization of planktonic anammox bacteria in the coastal upwelling ecosystem off central Chile. *Progress in Oceanography* **92-95**, 110-120.

Galan, A., Molina, V., Thamdrup, B., Woebken, D., Lavik, G., Kuypers, M.M.M. and Ulloa, O. (2009). Anammox bacteria and the anaerobic oxidation of ammonium in the oxygen minimum zone off northern Chile. *Deep-Sea Research Part Ii-Topical Studies in Oceanography* **56**, 1125-1135.

Gibb, S.W., Mantoura, R.F.C. and Liss, P.S. (1995). Analysis of ammonia and methylamines in natural waters by flow injection gas diffusion coupled to ion chromatography. *Analytica Chimica Acta* **316**, 291-304.

Gori, F., Tringe, S.G., Kartal, B., Machiori, E. and Jetten, M.S.M. (2011). The metagenomic basis of anammox metabolism in Candidatus 'Brocadia fulgida'. *Biochemical Society Transactions* **39**, 1799-1804.

Gruber, N. and Sarmiento, J.L. (1997). Global patterns of marine nitrogen fixation and denitrification. *Global Biogeochemical Cycles* **11**, 235-266.

Güven, D., Dapena, A., Kartal, B., Schmid, M.C., Maas, B., van de Pas-Schoonen, K., Sozen, S., Mendez, R., Op den Camp, H.J.M., Jetten, M.S.M., Strous, M. and Schmidt, I. (2005). Propionate oxidation by and methanol inhibition of anaerobic ammonium-oxidizing bacteria. *Applied and Environmental Microbiology* **71**, 1066-1071.

Hamersley, M.R., Lavik, G., Woebken, D., Rattray, J.E., Lam, P., Hopmans, E.C., Damste, J.S.S., Kruger, S., Graco, M., Gutierrez, D. and Kuypers, M.M.M. (2007). Anaerobic ammonium oxidation in the Peruvian oxygen minimum zone. *Limnology and oceanography* **52**, 923-933.

Hamersley, M.R., Woebken, D., Boehrer, B., Schultze, M., Lavik, G. and Kuypers, M.M.M. (2009). Water column anammox and denitrification in a temperate permanently stratified lake (Lake Rassnitzer, Germany). *Systematic and Applied Microbiology* **32**, 571-582.

Harhangi, H.R., Le Roy, M., van Alen, T., Hu, B.L., Groen, J., Kartal, B., Tringe, S.G., Quan, Z.X., Jetten, M.S.M. and Op den Camp, H.J.M. (2012). Hydrazine Synthase, a Unique Phylomarker with Which To Study the Presence and Biodiversity of Anammox Bacteria. *Applied and Environmental Microbiology* **78**, 752-758.

Hira, D., Toh, H., Migita, C.T., Okubo, H., Nishiyama, T., Hattori, M., Furukawa, K. and Fujii, T. (2012). Anammox organism KSU-1 expresses a NirK-type copper-containing nitrite reductase instead of a NirS-type with cytochrome cd(1). *Febs Letters* **586**, 1658-1663.

Hirsch, M.D., Long, Z.T. and Song, B. (2011). Anammox Bacterial Diversity in Various Aquatic Ecosystems Based on the Detection of Hydrazine Oxidase Genes (hzoA/hzoB). *Microbial Ecology* **61**, 264-276.

Hu, Z., Speth, D.R., Francoijs, K.-J., Quan, Z.-X. and Jetten, M.S.M. (2012). Metagenome Analysis of a Complex Community Reveals the Metabolic Blueprint of Anammox Bacterium "Candidatus Jettenia asiatica". *Frontiers in microbiology* **3**, 366-366.

Huang, W.E., Stoecker, K., Griffiths, R., Newbold, L., Daims, H., Whiteley, A.S. and Wagner, M. (2007). Raman‐FISH: combining stable‐isotope Raman spectroscopy and fluorescence in situ hybridization for the single cell analysis of identity and function. *Environmental microbiology* **9**, 1878-1889.

Huber, T., Faulkner, G. and Hugenholtz, P. (2004). Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**, 2317-2319.

Humbert, S., Tarnawski, S., Fromin, N., Mallet, M.P., Aragno, M. and Zopfi, J. (2010). Molecular detection of anammox bacteria in terrestrial ecosystems: distribution and diversity. *Isme Journal* **4**, 450-454.

Humbert, S., Zopfi, J. and Tarnawski, S.-E. (2012). Abundance of anammox bacteria in different wetland soils. *Environmental Microbiology Reports* **4**, 484-490.

Huse, S.M., Dethlefsen, L., Huber, J.A., Welch, D.M., Relman, D.A. and Sogin, M.L. (2008). Exploring Microbial Diversity and Taxonomy Using SSU rRNA Hypervariable Tag Sequencing. *PLoS Genet* **4**, e1000255.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L. and Mark Welch, D. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology* **8**,

Huse, S.M., Welch, D.M., Morrison, H.G. and Sogin, M.L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental microbiology* **12**, 1889-1898.

Jaeschke, A., den Camp, H., Harhangi, H., Klimiuk, A., Hopmans, E.C., Jetten, M.S.M., Schouten, S. and Damste, J.S.S. (2009a). 16S rRNA gene and lipid biomarker evidence for anaerobic ammonium-oxidizing bacteria (anammox) in California and Nevada hot springs. *Fems Microbiology Ecology* **67**, 343-350.

Jaeschke, A., Hopmans, E.C., Wakeham, S.G., Schouten, S. and Damste, J.S. (2007). The presence of ladderane lipids in the oxygen minimum zone of the Arabian Sea indicates nitrogen loss through anammox. *Limnology and oceanography* **52**, 780-786.

Jaeschke, A., Rooks, C., Trimmer, M., Nicholls, J.C., Hopmans, E.C., Schouten, S. and Damste, J.S.S. (2009b). Comparison of ladderane phospholipid and core lipids as indicators for anaerobic ammonium oxidation (anammox) in marine sediments. *Geochimica Et Cosmochimica Acta* **73**, 2077-2088.

Jetten, M., Schmid, M., van de Pas-Schoonen, K., Damste, J.S. and Strous, M. (2005) Anammox organisms: Enrichment, cultivation, and environmental analysis. In *Environmental Microbiology*, vol. 397, pp. 34-57. Editted by.

Jetten, M.S.M., Sliekers, O., Kuypers, M., Dalsgaard, T., van Niftrik, L., Cirpus, I., van de Pas-Schoonen, K., Lavik, G., Thamdrup, B., Le Paslier, D., Op den Camp, H.J.M., Hulth, S., Nielsen, L.P., Abma, W., Third, K., Engstrom, P., Kuenen, J.G., Jorgensen, B.B., Canfield, D.E., Damste, J.S.S., Revsbech, N.P., Fuerst, J., Weissenbach, J., Wagner, M., Schmidt, I., Schmid, M. and Strous, M. (2003). Anaerobic ammonium oxidation by marine and freshwater planctomycete-like bacteria. *Applied Microbiology and Biotechnology* **63**, 107-114.

Jetten, M.S.M., van Niftrik, L., Strous, M., Kartal, B., Keltjens, J.T. and Op den Camp, H.J.M. (2009). Biochemistry and molecular biology of anammox bacteria. *Critical Reviews in Biochemistry and Molecular Biology* **44**, 65-84.

Karlsson, R., Karlsson, A., Bäckman, O., Johansson, B.R. and Hulth, S. (2009). Identification of key proteins involved in the anammox reaction. *Fems Microbiology Letters* **297**, 87-94.

Karstensen, J., Stramma, L. and Visbeck, M. (2008). Oxygen minimum zones in the eastern tropical Atlantic and Pacific oceans. *Progress in Oceanography* **77**, 331-350.

Kartal, B., Koleva, M., Arsov, R., van der Star, W., Jetten, M.S.M. and Strous, M. (2006). Adaptation of a freshwater anammox population to high salinity wastewater. *Journal of Biotechnology* **126**, 546-553.

Kartal, B., Kuypers, M.M.M., Lavik, G., Schalk, J., Op den Camp, H.J.M., Jetten, M.S.M. and Strous, M. (2007a). Anammox bacteria disguised as denitrifiers: nitrate reduction to dinitrogen gas via nitrite and ammonium. *Environmental microbiology* **9**, 635-642.

Kartal, B., Maalcke, W.J., de Almeida, N.M., Cirpus, I., Gloerich, J., Geerts, W., den Camp, H.J.O., Harhangi, H.R., Janssen-Megens, E.M. and Francoijs, K.-J. (2011a). Molecular mechanism of anaerobic ammonium oxidation. *Nature* **479**, 127-130.

Kartal, B., Maalcke, W.J., de Almeida, N.M., Cirpus, I., Gloerich, J., Geerts, W., Op den Camp, H.J.M., Harhangi, H.R., Janssen-Megens, E.M., Francoijs, K.-J., Stunnenberg, H.G., Keltjens, J.T., Jetten, M.S.M. and Strous, M. (2011b). Molecular mechanism of anaerobic ammonium oxidation. *Nature* **advance online publication**,

Kartal, B., Rattray, J., van Niftrik, L.A., van de Vossenberg, J., Schmid, M.C., Webb, R.I., Schouten, S., Fuerst, J.A., Damste, J.S., Jetten, M.S.M. and Strous, M. (2007b). Candidatus "*Anammoxoglobus propionicus*" a new propionate oxidizing species of anaerobic ammonium oxidizing bacteria. *Systematic and Applied Microbiology* **30**, 39-49.

Kartal, B., van Niftrik, L., Rattray, J., de Vossenberg, J., Schmid, M.C., Damste, J.S.S., Jetten, M.S.M. and Strous, M. (2008). Candidatus 'Brocadia fulgida': an autofluorescent anaerobic ammonium oxidizing bacterium. *Fems Microbiology Ecology* **63**, 46-55.

King, G.M. (1984). Metabolism of trimethylamine, choline, and glycine betaine by sulfate-reducing and methanogenic bacteria in marine sediments. *Appl Environ Microbiol* **48**, 719-725.

Klotz, M.G., Schmid, M.C., Strous, M., den Camp, H., Jetten, M.S.M. and Hooper, A.B. (2008). Evolution of an octahaem cytochrome c protein family that is key to aerobic and anaerobic ammonia oxidation by bacteria. *Environmental microbiology* **10**, 3150-3163.

Klotz, M.G. and Stein, L.Y. (2008). Nitrifier genomics and evolution of the nitrogen cycle. *Fems Microbiology Letters* **278**, 146-156.

Kopczynski, E.D., Bateson, M.M. and Ward, D.M. (1994). Recognition of chimeric small-subunit ribosomal DNAs composed of genes from uncultivated microorganisms. *Applied and Environmental Microbiology* **60**, 746-748.

Kuenen, J.G. (2008). Anammox bacteria: from discovery to application. *Nature Reviews Microbiology* **6**, 320-326.

Kunin, V., Engelbrektson, A., Ochman, H. and Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental microbiology* **12**, 118-123.

Kuypers, M.M.M., Lavik, G. and Thamdrup, B. (2006) Anaerobic ammonium oxidation in the marine environment. In *Past and Present Water Column Anoxia*, vol. 64, pp. 311-335. Editted by Neretin, L.N.

Kuypers, M.M.M., Lavik, G., Woebken, D., Schmid, M., Fuchs, B.M., Amann, R., Jorgensen, B.B. and Jetten, M.S.M. (2005). Massive nitrogen loss from the Benguela upwelling system through anaerobic ammonium oxidation. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 6478-6483.

Kuypers, M.M.M., Sliekers, A.O., Lavik, G., Schmid, M., Jorgensen, B.B., Kuenen, J.G., Damste, J.S.S., Strous, M. and Jetten, M.S.M. (2003). Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. *Nature* **422**, 608-611.

Lam, P., Lavik, G., Jensen, M.M., van de Vossenberg, J., Schmid, M., Woebken, D., Dimitri, G., Amann, R., Jetten, M.S.M. and Kuypers, M.M.M. (2009).

Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 4752-4757.

Li, H., Chen, S., Mu, B.Z. and Gu, J.D. (2010a). Molecular Detection of Anaerobic Ammonium-Oxidizing (Anammox) Bacteria in High-Temperature Petroleum Reservoirs. *Microbial Ecology* **60**, 771-783.

Li, M., Ford, T., Li, X.Y. and Gu, J.D. (2011). Cytochrome cd1-Containing Nitrite Reductase Encoding Gene nirS as a New Functional Biomarker for Detection of Anaerobic Ammonium Oxidizing (Anammox) Bacteria. *Environmental Science & Technology* **45**, 3547-3553.

Li, M., Hong, Y.G., Klotz, M.G. and Gu, J.D. (2010b). A comparison of primer sets for detecting 16S rRNA and hydrazine oxidoreductase genes of anaerobic ammonium-oxidizing bacteria in marine sediments. *Applied Microbiology and Biotechnology* **86**, 781-790.

Li, W., Jaroszewski, L. and Godzik, A. (2001). Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17**, 282-283.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., Konig, A., Liss, T., Lussmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A. and Schleifer, K.H. (2004). ARB: a software environment for sequence data. *Nucleic acids research* **32**, 1363-1371.

Luesken, F.A., Sanchez, J., van Alen, T.A., Sanabria, J., Op den Camp, H.J.M., Jetten, M.S.M. and Kartal, B. (2011). Simultaneous Nitrite-Dependent Anaerobic Methane and Ammonium Oxidation Processes. *Applied and Environmental Microbiology* **77**, 6802-6807.

Manahan, S. (2005) *Environmental chemistry*. CRC press.

Manefield, M., Whiteley, A.S., Griffiths, R.I. and Bailey, M.J. (2002). RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. *Applied and Environmental Microbiology* **68**, 5367-5373.

McArthur, J.V. (2006) *Microbial ecology: an evolutionary approach*. Academic press.

McDaniel, L.D., Young, E., Delaney, J., Ruhnau, F., Ritchie, K.B. and Paul, J.H. (2010). High Frequency of Horizontal Gene Transfer in the Oceans. *Science* **330**, 50-50.

Mohamed, N.M., Saito, K., Tal, Y. and Hill, R.T. (2010). Diversity of aerobic and anaerobic ammonia-oxidizing bacteria in marine sponges. *Isme Journal* **4**, 38-48.

Moore, T.A., Xing, Y.P., Lazenby, B., Lynch, M.D.J., Schiff, S., Robertson, W.D., Timlin, R., Lanza, S., Ryan, M.C., Aravena, R., Fortin, D., Clark, I.D. and Neufeld, J.D. (2011). Prevalence of Anaerobic Ammonium-Oxidizing Bacteria in Contaminated Groundwater. *Environmental Science & Technology* **45**, 7217-7225.

Mulder, A., Vandegraaf, A.A., Robertson, L.A. and Kuenen, J.G. (1995). Anaerobic Ammonium Oxidation Discovered in a Denitrifying Fluidized-Bed Reactor. *Fems Microbiology Ecology* **16**, 177-183.

Nagel, B., Emeis, K.-C., Flohr, A., Rixen, T., Schlarbaum, T., Mohrholz, V. and van der Plas, A. (2013). N-cycling and balancing of the N-deficit generated in the oxygen minimum zone over the Namibian shelf—An isotope-based approach. *Journal of Geophysical Research: Biogeosciences* **118**, 361-371.

Neef, A., Amann, R., Schlesner, H. and Schleifer, K.H. (1998). Monitoring a widespread bacterial group: in situ detection of planctomycetes with 16S rRNA-targeted probes. *Microbiology-Uk* **144**, 3257-3266.

Neufeld, J.D., Vohra, J., Dumont, M.G., Lueders, T., Manefield, M., Friedrich, M.W. and Murrell, J.C. (2007). DNA stable-isotope probing. *Nature protocols* **2**, 860-866.

Neumann, S., Jetten, M.S.M. and van Niftrik, L. (2011). The ultrastructure of the compartmentalized anaerobic ammonium-oxidizing bacteria is linked to their energy metabolism. *Biochemical Society Transactions* **39**, 1805-1810.

Nicholls, J.C., Davies, C.A. and Trimmer, M. (2007). High resolution profiles and nitrogen isotope tracing reveal a dominant source of nitrous oxide and multiple pathways of nitrogen gas formation in the central Arabian Sea. *Limnology and oceanography* **52**, 156-168.

Nicholls, J.C. and Trimmer, M. (2009). Widespread occurrence of the anammox reaction in estuarine sediments. *Aquatic Microbial Ecology* **55**, 105-113.

Nielsen, L.P. (1992). Denitrification in Sediment Determined From Nitrogen Isotope Pairing. *Fems Microbiology Ecology* **86**, 357-362.

Niu, B.F., Fu, L.M., Sun, S.L. and Li, W.Z. (2010). Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *Bmc Bioinformatics* **11**,

Nogales, B., Timmis, K.N., Nedwell, D.B. and Osborn, A.M. (2002). Detection and diversity of expressed denitrification genes in estuarine sediments after reverse transcription-PCR amplification from mRNA. *Applied and Environmental Microbiology* **68**, 5017-5025.

Oakley, B.B., Carbonero, F., Dowd, S.E., Hawkins, R.J. and Purdy, K.J. (2012). Contrasting patterns of niche partitioning between two anaerobic terminal oxidizers of organic matter. *Isme Journal* **6**, 905-914.

Oakley, B.B., Dowd, S.E. and Purdy, K.J. (2011). ThermoPhyl: a software tool for selecting phylogenetically optimized conventional and quantitative-PCR taxon-targeted assays for use with complex samples. *Fems Microbiology Ecology* **77**, 17-27.

Ochman, H., Lawrence, J.G. and Groisman, E.A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299-304.

Ogilvie, B., Nedwell, D.B., Harrison, R.M., Robinson, A. and Sage, A. (1997). High nitrate, muddy estuaries as nitrogen sinks: The nitrogen budget of the River Colne estuary (United Kingdom). *Marine Ecology Progress Series* **150**, 217-228.

Paulmier, A. and Ruiz-Pino, D. (2009). Oxygen minimum zones (OMZs) in the modern ocean. *Progress in Oceanography* **80**, 113-128.

Penton, C.R., Devol, A.H. and Tiedje, J.M. (2006). Molecular evidence for the broad distribution of anaerobic ammonium-oxidizing bacteria in freshwater and marine sediments. *Applied and Environmental Microbiology* **72**, 6829-6832.

Podlaska, A., Wakeham, S.G., Fanning, K.A. and Taylor, G.T. (2012). Microbial community structure and productivity in the oxygen minimum zone of the

eastern tropical North Pacific. *Deep-Sea Research Part I-Oceanographic Research Papers* **66**, 77-89.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W.G., Peplies, J. and Glockner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids research* **35**, 7188-7196.

Purdy, K.J. (2005) Nucleic acid recovery from complex environmental samples. In *Environmental Microbiology*, vol. 397, pp. 271-292. Editted by.

Purdy, K.J., Embley, T.M., Takii, S. and Nedwell, D.B. (1996). Rapid extraction of DNA and rRNA from sediments by a novel hydroxyapatite spin-column method. *Applied and Environmental Microbiology* **62**, 3905-3907.

Quan, Z.X., Rhee, S.K., Zuo, J.E., Yang, Y., Bae, J.W., Park, J.R., Lee, S.T. and Park, Y.H. (2008). Diversity of ammonium-oxidizing bacteria in a granular sludge anaerobic ammonium-oxidizing (anammox) reactor. *Environmental microbiology* **10**, 3130-3139.

Quince, C., Curtis, T.P. and Sloan, W.T. (2008). The rational exploration of microbial diversity. *Isme Journal* **2**, 997-1006.

Quince, C., Lanzen, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M., Read, L.F. and Sloan, W.T. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods* **6**, 639-U627.

Rattray, J.E., van de Vossenberg, J., Hopmans, E.C., Kartal, B., van Niftrik, L., Rijpstra, W.I.C., Strous, M., Jetten, M.S.M., Schouten, S. and Damste, J.S.S. (2008). Ladderane lipid distribution in four genera of anammox bacteria. *Archives of Microbiology* **190**, 51-66.

Redfield, A.C. (1934). On the proportions of organic derivations in sea water and their relation to the composition of plankton. *University Press of Liverpool*,

Rich, J., Dale, O., Song, B. and Ward, B. (2008). Anaerobic ammonium oxidation (anammox) in Chesapeake Bay sediments. *Microbial Ecology* **55**, 311-320.

Richards, F.A. (1965) Anoxic basins and Fjords. In *Chemical Oceanography*, pp. 611-643. Editted by Ripley, J.P. and Skirrow, G.: Academic Press.

Richards, F.A., Anderson, J.J. and Cline, J.D. (1971). CHEMICAL AND PHYSICAL OBSERVATIONS IN GOLFO DULCE, AN ANOXIC BASIN ON PACIFIC COAST OF COSTA-RICA. *Limnology and oceanography* **16**, 43-&.

Risgaard-Petersen, N., Meyer, R.L., Schmid, M., Jetten, M.S.M., Enrich-Prast, A., Rysgaard, S. and Revsbech, N.P. (2004). Anaerobic ammonium oxidation in an estuarine sediment. *Aquatic Microbial Ecology* **36**, 293-304.

Risgaard-Petersen, N., Nielsen, L.P., Rysgaard, S., Dalsgaard, T. and Meyer, R.L. (2003). Application of the isotope pairing technique in sediments where anammox and denitrification co-exist. *Limnology and Oceanography: Methods* **1**, 63-73.

Rooks, C., Schmid, M.C., Mehsana, W. and Trimmer, M. (2012). The depth-specific significance and relative abundance of anaerobic ammonium-oxidizing bacteria in estuarine sediments (Medway Estuary, UK). *Fems Microbiology Ecology* **80**, 19-29.

Rush, D., Wakeham, S.G., Hopmans, E.C., Schouten, S. and Damste, J.S.S. (2012). Biomarker evidence for anammox in the oxygen minimum zone of the Eastern Tropical North Pacific. *Organic Geochemistry* **53**, 80-87.

Russ, L., Harhangi, H.R., Schellekens, J., Verdellen, B., Kartal, B., Op den Camp, H.J.M. and Jetten, M.S.M. (2012). Genome analysis and

heterologous expression of acetate-activating enzymes in the anammox bacterium Kuenenia stuttgartiensis. *Archives of Microbiology* **194**, 943-948.

Rysgaard, S. and Glud, R.N. (2004). Anaerobic N-2 production in Arctic sea ice. *Limnology and oceanography* **49**, 86-94.

Rysgaard, S., Glud, R.N., Risgaard-Petersen, N. and Dalsgaard, T. (2004). Denitrification and anammox activity in Arctic marine sediments. *Limnology and oceanography* **49**, 1493-1502.

Sambrook, J., Fritsch, E. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sato, Y., Ohta, H., Yamagishi, T., Guo, Y., Nishizawa, T., Rahman, M.H., Kuroda, H., Kato, T., Saito, M., Yoshinaga, I., Inubushi, K. and Suwa, Y. (2012). Detection of Anammox Activity and 16S rRNA Genes in Ravine Paddy Field Soil. *Microbes and Environments* **27**, 316-319.

Schafer, H. and Muyzer, G. (2001) Denaturing gradient gel electrophoresis in marine microbial ecology. In *Methods in Microbiology, Vol 30*, vol. 30, pp. 425-468. Editted by.

Schalk, J., de Vries, S., Kuenen, J.G. and Jetten, M.S.M. (2000). Involvement of a novel hydroxylamine oxidoreductase in anaerobic ammonium oxidation. *Biochemistry* **39**, 5405-5412.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J. and Weber, C.F. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology* **75**, 7537-7541.

Schmid, M., Twachtmann, U., Klein, M., Strous, M., Juretschko, S., Jetten, M., Metzger, J.W., Schleifer, K.H. and Wagner, M. (2000). Molecular evidence for genus level diversity of bacteria capable of catalyzing anaerobic ammonium oxidation. *Systematic and Applied Microbiology* **23**, 93-106.

Schmid, M., Walsh, K., Webb, R., Rijpstra, W.I.C., van de Pas-Schoonen, K., Verbruggen, M.J., Hill, T., Moffett, B., Fuerst, J., Schouten, S., Damste, J.S.S., Harris, J., Shaw, P., Jetten, M. and Strous, M. (2003). Candidatus "*Scalindua brodae*", sp nov., Candidatus "*Scalindua wagneri*", sp nov., two new species of anaerobic ammonium oxidizing bacteria. *Systematic and Applied Microbiology* **26**, 529-538.

Schmid, M.C., Hooper, A.B., Klotz, M.G., Woebken, D., Lam, P., Kuypers, M.M.M., Pommerening-Roeser, A., op den Camp, H.J.M. and Jetten, M.S.M. (2008). Environmental detection of octahaem cytochrome *c* hydroxylamine/hydrazine oxidoreductase genes of aerobic and anaerobic ammonium-oxidizing bacteria. *Environmental microbiology* **10**, 3140-3149.

Schmid, M.C., Risgaard-Petersen, N., van de Vossenberg, J., Kuypers, M.M.M., Lavik, G., Petersen, J., Hulth, S., Thamdrup, B., Canfield, D., Dalsgaard, T., Rysgaard, S., Sejr, M.K., Strous, M., Op den Camp, H.J.M. and Jetten, M.S.M. (2007). Anaerobic ammonium-oxidizing bacteria in marine environments: widespread occurrence but low diversity. *Environmental microbiology* **9**, 1476-1484.

Schubert, C.J., Durisch-Kaiser, E., Wehrli, B., Thamdrup, B., Lam, P. and Kuypers, M.M.M. (2006). Anaerobic ammonium oxidation in a tropical

freshwater system (Lake Tanganyika). *Environmental microbiology* **8**, 1857-1863.

Seitzinger, S., Harrison, J.A., Böhlke, J., Bouwman, A., Lowrance, R., Peterson, B., Tobias, C. and Drecht, G.V. (2006). Denitrification across landscapes and waterscapes: a synthesis. *Ecological Applications* **16**, 2064-2090.

Seitzinger, S.P., Sanders, R.W. and Styles, R. (2002). Bioavailability of DON from natural and anthropogenic sources to estuarine plankton. *Limnology and oceanography* **47**, 353-366.

Selenska, S. and Klingmuller, W. (1991). DNA recovery and direct detection of Tn5 sequences from soil. *Letters in Applied Microbiology* **13**, 21-24.

Shimamura, M., Nishiyama, T., Shigetomo, H., Toyomoto, T., Kawahara, Y., Furukawa, K. and Fujii, T. (2007). Isolation of a multiheme protein with features of a hydrazine-oxidizing enzyme from an anaerobic ammonium-oxidizing enrichment culture. *Applied and Environmental Microbiology* **73**, 1065-1072.

Sliekers, A.O., Haaijer, S., Schmid, M., Harhangi, H., Verwegen, K., Kuenen, J.G. and Jetten, M.S.M. (2004). Nitrification and Anammox with urea as the energy source. *Systematic and Applied Microbiology* **27**, 271-278.

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. *Applied and Environmental Microbiology* **73**, 3612-3622.

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., Arrieta, J.M. and Herndl, G.J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America* **103**, 12115-12120.

Song, B. and Tobias, C.R. (2011) MOLECULAR AND STABLE ISOTOPE METHODS TO DETECT AND MEASURE ANAEROBIC AMMONIUM OXIDATION (ANAMMOX) IN AQUATIC ECOSYSTEMS. In *Methods in Enzymology, Vol 46: Research on Nitrification and Related Processes, Pt B*, vol. 496, pp. 63-89. Editted by.

Stackebrandt, E. and Goebel, B. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic Bacteriology* **44**, 846-849.

Stewart, F.J. (2011). Dissimilatory sulfur cycling in oxygen minimum zones: an emerging metagenomics perspective. *Biochemical Society Transactions* **39**, 1859-1863.

Strous, M., Fuerst, J.A., Kramer, E.H.M., Logemann, S., Muyzer, G., van de Pas-Schoonen, K.T., Webb, R., Kuenen, J.G. and Jetten, M.S.M. (1999). Missing lithotroph identified as new planctomycete. *Nature* **400**, 446-449.

Strous, M., Heijnen, J., Kuenen, J. and Jetten, M. (1998). The sequencing batch reactor as a powerful tool for the study of slowly growing anaerobic ammonium-oxidizing microorganisms. *Applied Microbiology and Biotechnology* **50**, 589-596.

Strous, M. and Jetten, M.S.M. (2004). Anaerobic oxidation of methane and ammonium. *Annual Review of Microbiology* **58**, 99-117.

Strous, M., Pelletier, E., Mangenot, S., Rattei, T., Lehner, A., Taylor, M.W., Horn, M., Daims, H., Bartol-Mavel, D., Wincker, P., Barbe, V., Fonknechten, N., Vallenet, D., Segurens, B., Schenowitz-Truong, C.,

Medigue, C., Collingro, A., Snel, B., Dutilh, B.E., Op den Camp, H.J.M., van der Drift, C., Cirpus, I., van de Pas-Schoonen, K.T., Harhangi, H.R., van Niftrik, L., Schmid, M., Keltjens, J., van de Vossenberg, J., Kartal, B., Meier, H., Frishman, D., Huynen, M.A., Mewes, H.W., Weissenbach, J., Jetten, M.S.M., Wagner, M. and Le Paslier, D. (2006). Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440**, 790-794.

Strous, M., vanGerven, E., Kuenen, J.G. and Jetten, M. (1997). Effects of aerobic and microaerobic conditions on anaerobic ammonium-oxidizing (Anammox) sludge. *Applied and Environmental Microbiology* **63**, 2446-2448.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596-1599.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28**, 2731-2739.

Thamdrup, B. and Dalsgaard, T. (2002). Production of N-2 through anaerobic ammonium oxidation coupled to nitrate reduction in marine sediments. *Applied and Environmental Microbiology* **68**, 1312-1318.

Thamdrup, B., Dalsgaard, T., Jensen, M.M., Ulloa, O., Farias, L. and Escribano, R. (2006). Anaerobic ammonium oxidation in the oxygen-deficient waters off northern Chile. *Limnology and oceanography* **51**, 2145-2156.

The Environment Agency. (2005) Medway Catchment Abstraction Management Strategy.

The Environment Agency. (2013) Medway Abstraction Licensing Strategy. *LIT 1995*.

Trimmer, M., Nicholls, J.C. and Deflandre, B. (2003). Anaerobic Ammonium Oxidation measured in sediments along the Thames estuary, United Kingdom. *Applied and Environmental Microbiology* **69**, 6447-6454.

Trimmer, M., Nicholls, J.C., Morley, N., Davies, C.A. and Aldridge, J. (2005). Biphasic behavior of Anammox regulated by nitrite and nitrate in an estuarine sediment. *Applied and Environmental Microbiology* **71**, 1923-1930.

Trimmer, M. and Purdy, K.J. (2012). Evidence for the direct oxidation of organic nitrogen to N-2 gas in the Arabian Sea. *Isme Journal* **6**, 1798-1800.

Ulloa, O., Canfield, D.E., DeLong, E.F., Letelier, R.M. and Stewart, F.J. (2012). Microbial oceanography of anoxic oxygen minimum zones. *Proceedings of the National Academy of Sciences* **109**, 15996-16003.

Van de Graaf, A.A., de Bruijn, P., Robertson, L.A., Jetten, M.S.M. and Kuenen, J.G. (1996). Autotrophic growth of anaerobic ammonium-oxidizing micro-organisms in a fluidized bed reactor. *Microbiology* **142**, 2187-2196.

van de Graaf, A.A., deBruijn, P., Robertson, L.A., Jetten, M.S.M. and Kuenen, J.G. (1997). Metabolic pathway of anaerobic ammonium oxidation on the basis of N-15 studies in a fluidized bed reactor. *Microbiology-Uk* **143**, 2415-2421.

van de Graaf, A.A., Mulder, A., Debruijn, P., Jetten, M.S.M., Robertson, L.A. and Kuenen, J.G. (1995). Anaerobic oxidation of ammonium is a biologically mediated process. *Applied and Environmental Microbiology* **61**, 1246-1251.

van de Vossenberg, J., Rattray, J.E., Geerts, W., Kartal, B., van Niftrik, L., van Donselaar, E.G., Damste, J.S.S., Strous, M. and Jetten, M.S.M. (2008). Enrichment and characterization of marine anammox bacteria associated with global nitrogen gas production. *Environmental microbiology* **10**, 3120-3129.

van de Vossenberg, J., Woebken, D., Maalcke, W.J., Wessels, H.J.C.T., Dutilh, B.E., Kartal, B., Janssen-Megens, E.M., Roeselers, G., Yan, J., Speth, D., Gloerich, J., Geerts, W., van der Biezen, E., Pluk, W., Francoijs, K.-J., Russ, L., Lam, P., Malfatti, S.A., Tringe, S.G., Haaijer, S.C.M., Op den Camp, H.J.M., Stunnenberg, H.G., Amann, R., Kuypers, M.M.M. and Jetten, M.S.M. (2012). The metagenome of the marine anammox bacterium 'Candidatus Scalindua profunda' illustrates the versatility of this globally important nitrogen cycle bacterium. *Environmental microbiology*, no-no.

van Niftrik, L., Geerts, W.J., van Donselaar, E.G., Humbel, B.M., Yakushevska, A., Verkleij, A.J., Jetten, M.S. and Strous, M. (2008). Combined structural and chemical analysis of the anammoxosome: a membrane-bounded intracytoplasmic compartment in anammox bacteria. *Journal of structural biology* **161**, 401-410.

van Niftrik, L. and Jetten, M.S.M. (2012). Anaerobic Ammonium-Oxidizing Bacteria: Unique Microorganisms with Exceptional Properties. *Microbiology and Molecular Biology Reviews* **76**, 585-+.

van Niftrik, L., van Helden, M., Kirchen, S., van Donselaar, E.G., Harhangi, H.R., Webb, R.I., Fuerst, J.A., Op den Camp, H.J.M., Jetten, M.S.M. and Strous, M. (2010). Intracellular localization of membrane-bound ATPases in the compartmentalized anammox bacterium 'Candidatus Kuenenia stuttgartiensis'. *Molecular Microbiology* **77**, 701-715.

Viancelli, A., Kunz, A., Esteves, P.A., Bauermann, F.V., Furukawa, K., Fujii, T., Antônio, R.V. and Vanotti, M. (2011). Bacterial biodiversity from an anaerobic up flow bioreactor with ANAMMOX activity inoculated with swine sludge. *Brazilian Archives of Biology and Technology* **54**, 1035-1041.

Wang, J. and Gu, J.-D. (2013). Dominance of Candidatus Scalindua species in anammox community revealed in soils with different duration of rice paddy cultivation in Northeast China. *Applied Microbiology and Biotechnology* **97**, 1785-1798.

Ward, B.B., Capone, D.G. and Zehr, J.P. (2007). What's New in the Nitrogen Cycle? *Oceanography* **20**, 101-109.

Ward, B.B., Devol, A.H., Rich, J.J., Chang, B.X., Bulow, S.E., Naik, H., Pratihary, A. and Jayakumar, A. (2009). Denitrification as the dominant nitrogen loss process in the Arabian Sea. *Nature* **461**, 78-81.

Wharfe, J.R. (1977). INTERTIDAL SEDIMENT HABITATS OF LOWER MEDWAY ESTUARY, KENT. *Environmental Pollution* **13**, 79-91.

Woebken, D., Fuchs, B.M., Kuypers, M.M.M. and Amann, R. (2007). Potential interactions of particle-associated anammox bacteria with bacterial and archaeal partners in the Namibian upwelling system. *Applied and Environmental Microbiology* **73**, 4648-4657.

Woebken, D., Lam, P., Kuypers, M.M.M., Naqvi, S.W.A., Kartal, B., Strous, M., Jetten, M.S.M., Fuchs, B.M. and Amann, R. (2008). A microdiversity study of anammox bacteria reveals a novel *Candidatus* Scalindua phylotype in marine oxygen minimum zones. *Environmental microbiology* **10**, 3106-3119.

Yamagishi, H., Westley, M.B., Popp, B.N., Toyoda, S., Yoshida, N., Watanabe, S., Koba, K. and Yamanaka, Y. (2007). Role of nitrification and denitrification on the nitrous oxide cycle in the eastern tropical North Pacific and Gulf of California. *Journal of Geophysical Research-Biogeosciences* **112**,

Yoshinaga, I., Amano, T., Yamagishi, T., Okada, K., Ueda, S., Sako, Y. and Suwa, Y. (2011). Distribution and Diversity of Anaerobic Ammonium Oxidation (Anammox) Bacteria in the Sediment of a Eutrophic Freshwater Lake, Lake Kitaura, Japan. *Microbes and Environments* **26**, 189-197.

You, F.M., Huo, N.X., Gu, Y.Q., Luo, M.C., Ma, Y.Q., Hane, D., Lazo, G.R., Dvorak, J. and Anderson, O.D. (2008). BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *Bmc Bioinformatics* **9**,

Zanardini, E., may, E., Moussard, H., Inkpen, R., Cappitelli, F., Murrell, J. and Purdy, K. (Unpublished). Investigations of the diversity of archaeal and bacterial communities on exfoliated sandstone from Porchester Castle (UK). *Fems Microbiology Ecology*,

Zehr, J.P. and Ward, B.B. (2002). Nitrogen cycling in the ocean: New perspectives on processes and paradigms. *Applied and Environmental Microbiology* **68**, 1015-1024.

Zhang, Z., Schwartz, S., Wagner, L. and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7**, 203-214.

Zheng, D.D., Alm, E.W., Stahl, D.A. and Raskin, L. (1996). Characterization of universal small-subunit rRNA hybridization probes for quantitative molecular microbial ecology studies. *Applied and Environmental Microbiology* **62**, 4504-4513.

Zhu, B.L., Sanchez, J., van Alen, T.A., Sanabria, J., Jetten, M.S.M., Ettwig, K.F. and Kartal, B. (2011). Combined anaerobic ammonium and methane oxidation for nitrogen and methane removal. *Biochemical Society Transactions* **39**, 1822-1825.