COMPARATIVE ANALYSIS OF *NITROSOCOCCUS* GENOMES; ISOLATION, PHYSIOLOGICAL AND GENOMIC CHARACTERIZATION OF THE NEW AMMONIA-OXIDIZING SPECIES *NITROSOCOCCUS WARDIAE*, STRAIN D1FHS^{T}; AND MOLECULAR STUDIES TOWARDS A BETTER UNDERSTANDING OF AMMONIA-DEPENDENT CHEMOLITHOTROPHY

by

Lin Wang

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Biology

Charlotte

2016

Approved by:

_____
Dr. Molly C. Redmond

_____
Dr. Martin G. Klotz

_____
Dr. Todd R. Steck

_____
Dr. Matthew W. Parrow

_____
Dr. Anthony A. Fodor

ABSTRACT

LIN WANG. Comparative analysis of *Nitrosococcus* genomes; isolation, physiological and genomic characterization of the new ammonia-oxidizing species *Nitrosococcus wardiae*, strain D1FHS^T; and molecular studies towards a better understanding of ammonia-dependent chemolithotrophy. (Under the direction of DR. MOLLY C. REDMOND and DR. MARTIN G. KLOTZ)

Obligate aerobic chemolithoautotrophic ammonia-oxidzing bacteria (AOB) derive energy and reductant needed for growth solely from the oxidation of ammonia to nitrite thereby facilitating the process of nitritation. A strain of an ammonia-oxidizing Gammaproteobacterium, D1FHS, was isolated into pure culture from an enrichment of sediment sampled from Jiaozhou Bay, China. Critical bioinformatics analyses of the whole genome sequence assigned D1FHS^T as the type-strain to a new *Nitrosococcus* species, named *Nitrosococcus wardiae*, and revealed unique genetic features and its archetypal metabolic capacity in the ammonia-oxidizing gammaproteobacterial genus *Nitrosococcus*. Growth-physiological studies of its salt, ammonium and thermal tolerance confirmed that *N. wardiae* represented by strain D1FHS^T is distinct from other known *Nitrosococcus* species, including *N. oceani*, *N. halophilus* and *N. watsonii*.

Presently, only *Nitrosococcus oceani* is represented by multiple strains isolated into pure culture from different oceanic gyres, several of which have been genome-sequenced. Comparative analysis of genome-sequenced strains of *Nitrosococcus oceani* from four different oceanic locations suggests that they have evolved by genome economization while maintaining a high level of identity in sequence and synteny, which correlates with their molecular defense capacity.

Chemolithoautotrophic ammonia-oxidizing bacteria facilitate the process of nitritation as their sole catabolic activity. The Gammaproteobacterium *Methylococcus capsulatus Bath* (MCB) is also capable of oxidizing ammonia and hydroxylamine to nitrite but cannot support growth on this nitritation process. MCB, a nitrifying and denitrifying obligate methanotroph, is thus the best thinkable host for investigating the molecular basis of ammonia-dependent chemolithotrophy. Quantitative real-time PCR studies of steady-state mRNA levels revealed that genes involved in nitrification (*hao*A, *hao*B, *cyt*L), denitrification (*norC*, *nor*B, *cyt*S) and ammonification (*nir*B, *nas*A) were differentially expressed with in the presence of ammonium, which is consistent with physiological observations.

It is known that the Epsilonmicrobium *Nautilia profundicola* is able to use nitrate as a terminal electron acceptor and as a source of nitrogen in the absence of ammonium; however, it lacks any known ammonium-forming nitrite reductase. Because involvement of reversely operating nitritation inventory, the reverse "hydroxylamine ubiquinone redox module" (HURM), has been demonstrated to contribute to ammonification and respiration in *Nautilia* and other bacteria, expression constructs were designed to assess the function of the reverse-HURM pathway with further investigations ongoing.

ACKNOWLEDGMENTS

I would like to thank my dissertation advisors, Dr. Molly C. Redmond and Dr. Martin G. Klotz, for their guidance, advice and support. Dr. Redmond has been patient, helpful and supportive of my own ideas and conveyed confidence in my abilities. She has been a strong guiding force in finalizing my dissertation. Dr. Klotz has had a tremendous influence on me, both personally and professionally, and has trained me to grow as a researcher. He is the one introduced me into this fascinating world of "nitrifying and denitrifying" microorganisms. Without his guidance and persistent help, none of this work would have been possible.

I also wish to thank my committee members – Dr. Todd R. Steck, Dr. Matthew W. Parrow, and Dr. Anthony A. Fodor for dedicating their time, expertise and support throughout my graduate studies.

I would like to extend a special thanks to my former labmate and friend, Dr. Chee Kent Lim, for his help and guidance with genomic data analysis. I would also like to extend my gratitude to Dr. Thomas E. Hanson, University of Delaware, for facilitating the PacBio sequencing of the D1FHS genome.

I am grateful to my parents for their understanding, emotional support, and continuous encouragement, and to all my friends, especially Dr. Kevin C. Lambirth, departmental colleagues, faculty and staff for their assistance throughout my studies.

I would also like to recognize the financial support of funding from the US National Science Foundations (a grant awarded to MGK), the tuition assistance by GASP

form the UNC Charlotte Graduate School, and support from Department of Biological Sciences.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 The Global Nitrogen Cycle

The global nitrogen (N) cycle (Fig.1.1), traditionally composed of nitrogen fixation (Fig.1.1, reaction1), nitrification (Fig.1.1, reaction 3,4,5) and denitrification (Fig.1.1, reaction 6), connects pathways of biotic and abiotic transformations of various nitrogenous compounds (Stein and Klotz 2016). The biotic processes are exclusively facilitated by microorganisms and microbes play thus also a primary role in marine N cycle. The bioavailability of nitrogenous nutrients is a critical limiting factor that controls biological productivity, biomass and eutrophication, and it consequently affects the rates of carbon assimilation and carbon loss in the coastal marine environment (Ryther and Dunstan 1971). Due to the fact that many protein complexes facilitating the transformation of intermediates in the N cycle are homologs of protein complexes that are active in the Carbon and Sulfur cycles and are dependent of redox-active metals such as Iron and Copper, the N cycle is linked to other major marine biogeochemical cycles.

Human activities have indisputably impacted the global N cycle (Vitousek et al., 1997). Most significantly, the Haber-Bosch process invented more than 100 years ago provided for the industrial fixation of gaseous dinitrogen (N_2) into ammonia (NH_3), a costly process that was quickly amended by industrial production of urea because urea is easier to store and transport; both processes provided for the nearly unlimited application of vast amounts of fixed nitrogen anywhere on the planet. This accelerated input of fixed

nitrogen to terrestrial, estuary, and concomitantly to open ocean habits also affected the diversity and distribution of microorganisms in those ecosystems (Nevison and Holland 1997, Karl et al., 2002, Doney et al., 2007, Dang et al., 2008, Duce et al., 2008, Galloway et al., 2008, Dang et al., 2009, 2010a,b, Stein and Klotz 2016).

1.1.1 Nitrogen Compound Transformations to Support Chemolithotrophy

Nitrogen compounds exist in a wide range of oxidation states, from -3, as in ammonium/ammonia and amines, to +5, as in the nitrate anion of nitric acid. The transformations between these nitrogen compounds are always coupled to the transfer of electrons between an electron donor and an electron acceptor, reactions that can occur spontaneously, are facilitated by various enzymes in single reactions or they are modular components of complex pathways. Ammonia-oxidizing chemolithotrophic bacteria (AOB) are capable of extracting electrons from reduced nitrogen compounds, relay them to the quinone pool and generate a proton motive force (pmf) to support ATP synthesis as well as the capture of useable reducing power for chemolithoautotrophic growth. In AOB, ammonia monooxygenase (AMO, EC 1.14.99.39) oxidizes ammonia (NH_3; but not its protonated form, ammonium) to hydroxylamine (NH_2OH), which requires the participation of two protons and two electrons, hypothesized in the literature to come from the quinone pool (Fig.1.2). NH_2OH represents the internal reductant useable for the extraction of electrons.

In AOB, the oxidation of hydroxylamine to NO_2^- is facilitated by hydroxylamine dehydrogenase (HAO, EC 1.7.2.6), an octaheme cytochrome *c* protein that extracts four electrons in the process (Fig.1.2). Given that 2 electrons per oxidation of ammonia must be provided to AMO, only two of these four extracted electrons remain available in the

quinone pool for participation in the respiratory electron flow to oxygen (O_2), which serves as the obligatory terminal electron acceptor, or to NAD to produce useable reducing power. The start and end points of the oxidative branch of the electron transport chain in AOB, quinol oxidation by ubiquinone: cytochrome *c* oxidoreductase (a.k.a., Complex III) and oxygen reduction by terminal oxidase (usually an A-type heme-copper oxidase, a.k.a. Complex IV; (Garcia-Horsman et al., 1994)), respectively, generate the proton motive force (PMF) that drives the synthesis of ATP, to provide the energy needed for growth. NADH: ubiquinone oxidoreductase complex (a.k.a., Complex I) will produce NADH to satisfy the need of reducing power for biosynthesis, while depleting PMF. AOB thus juggle their quinone/quinol pool at a fine line and ongoing catabolic activity, measurable as nitrite accumulation, does not necessarily result in cellular and population growth.

In AOB, electrons extracted by oxidation of hydroxylamine are relayed to the quinone pool by a hydroxylamine: ubiquinone redox module (HURM; Klotz and Stein 2008, Simon and Klotz 2013), which consists of hydroxylamine dehydrogenase (HAO = $HaoA_3$, encoded by *hao*A) connected to a cytochrome *c* quinone reductase (c_M552, encoded by *cyc*B). It has not been experimentally established whether HAO and c_M552 are interacting directly or connect via a cytochrome *c* protein electron shuttle such as *c*554 (encoded by *cyc*A) (Klotz et al., 2008, Campbell et al., 2011b). $HaoA_3$ (HAO) is a cross-linked, circular-symmetric trimeric enzyme complex that is mandatorily functional in all AOB (Hooper et al., 1983). Cytochrome *c*554 proteins are members of a large multiheme cytochrome *c* protein superfamily that also includes the HaoA protein family (Kern et al., 2011) and it has been experimentally established as a functional redox

partner of HAO (Arciero et al., 1991a, Collins et al., 1991, McTavish et al., 1995, Iverson et al., 1998, 2001, Kim et al., 2010). Biochemical experiments verified that cytochrome *c*554 reduces HAO by directing the extracted electrons to an artificial electron acceptor thereby allowing hydroxylamine oxidation to continue to work (Arciero et al., 1991b). While cytochrome *c*554 is a good candidate for accepting electrons from hydroxylamine, its participation in relaying electrons to the quinone pool via cytochrome c_M552 has never been successfully tested; instead, it was suggested that *c*554 might function as an electron flow switch that directs electrons to other acceptors in the periplasm (Stein et al., 2013). Cytochrome c_M552 is a membrane-bound tetraheme protein in the NrfH/NapC/NirT protein superfamily (Elmore et al., 2007) that functions as a quinone reductase in AOB; experimentation has shown that c_M552 interacts with quinone pool directly (Kim et al., 2008). However and again, whether and, if so, how cytochromes *c*554 and c_M552 interact as a serial redox shuttle between HAO and the quinone pool has never been experimentally tested as it is possible that cytochrome *c*554 relays electrons from the HAO complex to other electron acceptors including cytochrome *c*522 or terminal oxidase. Genomes of several *Epsilonproteobacteria* such as *Nautilia profundicola* do not encode any homologues of cytochrome *c*554, but they contain the genes that encode the HaoA protein and a member of the NapC/NrfH/c_M552 protein family, which is absent from the genomes of most *Epsilonproteobacteria* (Campbell et al., 2009, Hanson et al. 2013). *Nautilia profundicola* is able to use nitrate as a sole source of nitrogen as well as an anaerobic respiratory terminal electron acceptor, producing nitrite, but lacks any kind of gene encoding known ammonium-forming nitrite reductases. Therefore, a reverse-HURM pathway has been proposed to facilitate nitrite reduction, which is consistent with

experimental evidence (Campbell et al., 2009, Hanson et al., 2013). The HURM operating reversely in *Nautilia* generates hydroxylamine and ammonium from nitrite (Campbell et al., 2009, Hanson et al., 2013, Simon and Klotz 2013). The HaoA protein in these *Epsilonmicrobia* lacks the critical tyrosine protein ligand effecting the assembly of HaoA proteins into a different quaternary structure, which apparently prevents disproportionation of hydroxylamine and, instead, converts it to a reversely acting HAO complex functioning as a nitrite reductase; it is therefore referred to as reverse HAO (HAO'; Hanson et al., 2013). Because the electrons required for nitrite reduction by reverse HAO are supplied from the oxidation of quinol by a member of the NapC/NrfH/c_M552 protein superfamily, HAO and cytochrome c_M552 appear to be direct functional redox partners (Hanson et al., 2013). This NapC/NrfH/c_M552 is functionally analogous and structurally homologous to the pentaheme cytochrome *c* nitrite reductase complex that facilitates respiratory ammonification (Simon and Klotz 2013).

1.1.2 The Aerobic Oxidation of Ammonium, Starting the Process of Nitrification

Nitrification (Fig.1.1, reaction 3,4,5) is the process of oxidation of ammonia to nitrate via nitrite ($NH_3 \rightarrow NO_2^- \rightarrow NO_3^-$) (Camargo and Alonso 2006). Aerobic oxidation of ammonia to nitrite is catalyzed by AOB and ammonia-oxidizing Thaumarchaeota (AOA) (Konneke et al., 2005). Analyzed genome sequences indicated that all AOA express an ammonia monooxygenase, a key inventory of the bacterial nitrification pathway, which is only distantly related to the bacterial AMO (~ 40% protein sequence similarity (Hallam et al., 2012, Stahl and De la Torre 2012, Tavormina et al., 2011, Walker et al., 2010). In contrast, bacterial particulate methane monooxygenase (pMMO), encoded by *pmo* genes in the genome of methanotrophs (Methane-oxidizing bacteria,

MOB), which are homologues of *amo* genes (Klotz and Norton 1998), and bacterial AMO share a protein sequence similarity of approximately 74% (Sayavedra-Soto et al., 2011a). MOB are "quasi" chemoorganotrophs because they are acquiring energy, reductant and carbon by oxidizing methane (CH_4). MOB also engage in a diverse array of nitrogen compound transformations; for instance, MOB such as *Methylococcus capsulatus* can nitrify (transform ammonia to nitrite); however, they cannot extract energy or useable reductant through this process (Stein and Klotz 2011). Therefore, nitrifying MOB and AOB perform ecologically analogous roles in the global nitrogen cycle because they initiate the biological conversion of the most reduced N species to more oxidized forms (Prosser 1989). MOB, AOB and AOA are however not the only organisms that are able to oxidize ammonia; it has been shown that pMMO in methanotrophs, and AMO are homologous enzymes (Klotz and Norton 1998) and members of a large Cu-dependent membrane monooxygenase (Cu-MMO) superfamily that oxidizes ammonia and short-chain hydrocarbon compounds in other bacteria that do not utilize ammonia-oxidation in their catabolic activities (Sayavedra-Soto et al., 2011a, Tavormina et al., 2011, Coleman et al., 2012).

The second step of nitrification is the oxidation of nitrite to nitrate, also known as "nitratation", performed by obligate aerobic nitrite-oxidizing bacteria (NOB) (Prosser 1989). However, certain strains in the *Nitrospira* genus are capable of facilitating a complete oxidation of ammonia to nitrate (comammox; Fig.1.1, reactions 3A, 3B and 4; (Daims et al., 2015, Van Kessel et al., 2015)). The genomes of these *Nitrosospira* strains encode enzymes needed for the ammonia oxidation and nitrite oxidation processes.

1.1.3 The Reduction of Nitrite, Starting the Process of Denitrification

Denitrification is the process that recycles fixed nitrogen, nitrite and nitrate, back to gaseous dinitrogen (Fig.1.1 reaction 6). The process exists as canonical denitrification (NO^{3-} to N_2) of heterotrophic bacteria in anoxic environments and thus include anaerobic respiration (Zumft 1997). The process also exists as incomplete or nitrifier denitrification (Lipschultz et al., 1981) by reductive detoxification of nitrite to nitrous oxide (N_2O) in aerobic environments. AOB can facilitate the nitrifier denitrification process and produce nitric oxide and nitrous oxide through reduction in amounts that depend on the external supply of ammonium. AOB genome sequences indicate that both copper-containing nitrite reductase (NirK; *nir*K) and membrane-bound cytochrome *c* nitric oxide reductase (cNOR; *nor*CBQD) genes are present. Cytochromes *c*'-beta and P460 (encoded by *cyt*S and *cyt*L, respectively) can bind to and transform NO to detoxify it by reduction and oxidation, respectively (Elmore et al., 2007). Periplasmic mono-heme and di-heme cytochrome *c*552 proteins are expressed in all AOB, which appear to transfer electrons to multiple electron sinks in the periplasm, including but not restricted to NirK, cNor and complex IV. The finding that ammonium and hydroxylamine induce the expression of genes encoding cytochrome P460 and cytochrome *c*552 suggests a role of these proteins in the catabolic electron flow of *Nitrosococcus* and other AOB (Stein et al., 2013). According to recent genome analyses, MOB express a sizeable soluble periplasmic cytochrome *c* complement, consisting of cytochromes *c*551/*c*552/*c*553 and *c*555, that deliver electrons to NirK, cNor and cytochrome *c*'-beta for the reduction of NO_x (Fig.1.2; Poret-Peterson et al., 2008).

1.1.4 The Assimilation of Nitrogen into Biomass

Nitrogen fixation, as a major form of assimilatory ammonification, is the transformation of inert dinitrogen gas (N_2) to a biologically accessible because highly reduced NC: ammonium (NH_4^+). This complex, energy-costly reduction process is restricted to diazotrophic microorganism including nitrogen-fixing bacteria (Postgate 1970) and methanogenic archaea (Leigh 2000) that encode various forms of nitrogenase enzyme complexes. The ammonium produced during this process is either assimilated into biomass or is catabolized by aerobic (Fig.1.1, reaction 3) and anaerobic ammonia-oxidizing (Fig.1.1, reaction 7) microbes to conserve energy and reductant as discussed in text above. During the process of nitrogen assimilation into biomass, ammonium is assimilated into glutamate via glutamine by the glutamine synthetase-glutamine oxoglutarate aminotransferase (GS-GOGAT) complex in low ammonium environments (a.k.a. high affinity ammonia assimilation) or by glutamate dehydrogenase (GDH) in high ammonium environments (a.k.a. low affinity ammonia assimilation) and then synthesized into proteins and nucleic acids (Fig.1.3). Ammonium transporters are usually needed for acquisition of ammonium from the environment for assimilation. Based on analyses of their genome sequences, AOB seem to have both GD and GS-GOGAT (Arp et al., 2007, Klotz et al., 2006) or GD only (Chain et al., 2003) for ammonium assimilation.

In addition to the nitrogen compound transformations described above, anaerobic ammonia-oxidizing (anammox) bacteria perform a coupled nitrification–denitrification process that was first reported by Strous in 1999 (Strous et al., 1999). In anoxic environments, this process couples the oxidation of ammonium to the subsequent reduction of nitrite and NO to produce hydrazine, which is then dehydrogenated to

dinitrogen by a homologue in the HaoA protein family (Schmidt et al., 2001, Geerts et al. 2011). Anammox is accomplished within a specialized organelle called the anammoxosome in chemolithoautotrophic anaerobic ammonia-oxidizing bacteria in the *Brocadiaceae* family of the order *Planctomycetales* (Strous and Jetten 2004, Niftrik et al., 2009). The Anammox reaction has been reported as a major nitrogen cycling process in numerous natural systems such as the Black Sea and many marine oxygen-minimum zones (Kuypers 2003, 2005). Anammox is also ecologically beneficial for wastewater treatment as it removes both nitrite and ammonium simultaneously without releasing toxic gas N_2O. A very recent report described a hydroxylamine-dependent anammox process, in which hydroxylamine instead of NO is the key reactant for the oxidation of NH_4^+ (Oshiki et al., 2016).

Our current understanding of the nitrogen cycle can be summarized in five N transformation pathways: ammonification, including nitrogen fixation, ANRA and DNRA (Fig.1.1, reaction 1 and 2); nitrification, including ammonia oxidation and nitrite oxidation (Fig.1.1, reactions 3A, 3B and 4); denitrification, including complete or canonical denitrification and nitrifier denitrification (Fig. 1.2, reactions 6A–D); and nitrite–nitrate interconversion (Fig.1.2, reactions 4 and 5).

1.2 Ammonia-oxidizing Bacteria (AOB)

Obligate aerobic chemolithotrophic ammonia-oxidizing bacteria (AOB) utilize ammonium as the sole source for energy and reducing power for growth by oxidizing its non-protonated form, ammonia, to nitrite with hydroxylamine as the intermediate. While initially defined primarily by differences in cell morphology and physiological characteristics (Watson and Mandel 1971), the advent of 16S rRNA gene phylogeny led

to the definition of three genera of AOB (Head et al., 1993), which was largely confirmed by the sequencing of available representatives in pure culture (Arp et al., 2007, Nyerges et al., 2011). While the betaproteobacterial AOB (Beta-AOB) in the family *Nitrosomonadaceae* represent 2 genera, *Nitrosomonas* and *Nitrospira* with numerous species (Urakawa et al., 2015), *Nitrosococcus* is the only genus of gammaproteobacterial AOB (Gamma-AOB; (Campbell et al., 2011a)). In the past, research of the ecology and phylogeny of AOB has almost entirely focused on the Beta-AOB because this group is abundant in all oxic environments, soil, freshwater and marine, Beta-AOB usually grow faster than Gamma-AOB, making them more amenable to growth in the laboratory (Kowalchuk and Stephen 2001).

1.2.1 Diversity and Abundance of AOB

Because it is difficult to isolate AOB into pure culture, we can infer AOB community structure (abundance and distribution) by determining the abundance and diversity of marker genes with culture-independent molecular methods, such as clone library analysis, amplification of target genes by quantitative PCR and detection using gene microarrays (Ward et al., 2007). Because ammonia monooxygenase (*amo*A) and 16S rRNA gene-based phylogenies are congruent (Purkhold et al., 2003) and because the analysis of sequence identity and relatedness of *amo*A genes has higher resolution than that of 16S rRNA genes (Rotthauwe et al., 1997), AOB abundance has been mainly assessed by analysis of the *amo*A gene using quantitative real-time PCR (Limpiyakorn et al., 2005, Li et al. 2010, Mosier and Francis 2011).

Environmental conditions, including pH, salinity, ammonium, nitrite and oxygen concentrations, influence the diversity and spatial distribution of AOB (Bernhard et al.,

2005, 2007, Bollmann et al., 2002, De Bie et al., 2001) in estuary and coastal systems. Increasing salinity has been implicated in the loss of observed AOB diversity in the Plum Island Sound estuary system (Bernhard et al., 2005). In anthropogenic activity-impacted coastal areas, such as Jiaozhou Bay, China, where we sampled from, ammonium, nitrite-nitrogen from the nearby wastewater treatment plants and polluted rivers have most notable impact on the diversity and distribution of the AOB community than other environmental factors (Dang et al., 2010b).

Our understanding of the role environmental factors on AOB community composition is generally still limited, and much less is known about the abundance and functional diversity of Gamma-AOB. *Nitrosococcus oceani* has since been detected in many marine environments by immunofluorescence at concentrations of 10^3 to 10^4 cells ml^{-1} (Ward 1982, 2000, Ward and Carlucci 1985, Voytek 1995, Zaccone et al., 1995). Sequences of 16S rRNA and *amo*A genes cloned from environmental DNA from open ocean samples are closely related to those several cultured *Nitrosococcus oceani* strains (Ward and O'Mullan 2002), while some sequences indicating that these marine environments harboring unidentified new *Nitrosococcus* species. It implied that *N. oceani* strains are omnipresent in marine environments (Ward and O'Mullan 2002).

1.2.2 Cultivation of AOB in Pure Culture

AOB grow very slowly and are considered fastidious bacteria that are very difficult to isolate into pure culture. However, working with pure isolates is still necessary in order to conduct the physiological and bioenergetics studies needed to understand the biological basis of the marine nitrogen cycle.

The recalcitrance of Gamma-AOB to cultivation is the main reason for impeded progress in identifying more species. The conventional enrichment technique pioneered more than 100 years ago by the Russian microbiologist Sergey Winogradsky is utilized here. AOB were grown in 100 mL batch cultures, in 250ml Erlenmeyer flasks, at 30 ºC, in the dark, without shaking, in modified artificial sea water [700 mM NaCl, 12.5 mM $(NH4)_2SO_4$, 30 mM $MgSO_4$, 20 mM $MgCl_2$, 10 mM $CaCl_2$, 10 mM KCl, 0.2 mM $NaCO_3$, 3.0 mM $NaHCO_3$, 0.09 mM K_2HPO_4, 3 µM chelated iron, 0.4 µM $NaMoO_4$ 1.0 µM $MnCl_2$, 0.008 µM $CoCl_2$, 0.35 µM ZnSO4, 0.08 µM $CuSO_4$, phenol red 0.5%] with the addition of 0.1 mM NH_2OH.HCl solution (Campbell et al., 2011a).The production of nitrite from ammonia oxidation, will lower the pH of the medium. To maintain a suitable pH conditions from AOB to grow, the pH of the medium has to be adjusted daily to 7.5-8.0 using 0.25M K_2CO3. For maintenance, the culture was transferred as one-fifth of the original volume into a fresh marine medium once a month.

We recently reported the successful isolation into pure culture of a new Gamma-AOB, *Nitrosococcus wardiae* (Wang et al., 2016), which represents only the second such isolation of a Gamma-AOB since 1990 (Campbell et al., 2011a).

1.2.3 Gammaproteobacterial AOB

Gamma-AOB in the genus *Nitrosococcus* are adapted to and restricted to marine environments (Table 1.1; (Ward and O'Mullan 2002)) and strains representing three species are being maintained in pure culture (Campbell et al., 2011a). The first Gamma-AOB was isolated from seawater in 1965 by Watson (Watson 1965) and was called *Nitrosocystis oceanus*, now known as *Nitrosococcus oceani.* In 1976, Koops reported the isolation of a novel Gamma-AOB species, *Nitrosococcus mobilis* (Koops 1976) but it

was later shown to be a Beta-AOB, *Nitrosomonas mobilis* (Campbell et al., 2011a). To date, *Nitrosococcus oceani* (Murray and Watson 1962), *Nitrosococcus halophilus* (Koops et al., 1990) and *Nitrosococcus watsonii* (Campbell et al., 2011a) are the only the only species of marine gammaproteobacterial AOB with representative type strains in pure culture.

It is known that AOB have the ability to oxidize methane utilizing AMO; apparently, the AMO in gamma-AOB strain *Nitrosococus oceanii* has a similar affinity to ammonia and methane, while the AMO in *Nitrosomonas* has a much higher affinity to ammonia than methane (Hyman and Wood 1983). The sequence difference between Beta-*amoA* (from *Nitrosococcus oceani*) and Gamma-*amoA* genes (from *Nitrosomonas*) is congruent with the difference between Betaproteobacteria and Gammaproteobacteria based on ribosomal RNA gene sequences (Murrell and Holmes 1996). Interestingly, sequence of *amoA* of Gamma-AOB is more similar to *pmoA* of the Gammaproteobacteria than *amoA* of Beta-AOB (Hooper et al., 1997).

1.3 Jiaozhou Bay, a Marginal Sea at the Eastern Seaboard of China

1.3.1 Geographical Description of the Sampling Sites

Jiaozhou Bay (36°7′24.44″N 120°14′44.3″E) (Fig.1.4) is a "trumpet-shaped" semi-enclosed water body with a surface area of about 390 km^2 and an average water depth of 7 m, located on the Southwestern coast of Shandong Peninsula, China. It connects to the Chinese Yellow Sea with a narrow mouth, about 2.5 km wide (Shen et al., 2006). The bay is bordered by Qingdao City, a large industrial and agricultural city in North China with a population of approximately 9 million residents. Freshwater is provided to Jiaozhou Bay by numerous tributaries (Yang River, Dagu River, Moshui

River, Baishahe River, Wantou River, Loushan River, Banqiaofang River, Licun River, Haipo River. etc). Most of these rivers have become the canals for industrial, agricultural and urban wastes and are thus highly enriched in nutrients including concentrated dissolved inorganic nitrogen and inorganic phosphate sources (Buick et al., 2001).

Sediment samples were collected from 14 stations (including Station D1) of Jiaozhou Bay on October 26, 2008 (Fig.1.4). Stations A3, A5, J1, P5 and Y1 are located at the deep inner side of the bay with the least water exchange with the Yellow Sea while stations Dx and D5 are the closest ones to the Yellow Sea. Stations D1 and P1 located at the narrow mouth of the "trumpet", most frequent water exchange between Jiaozhou Bay and Yellow Sea. D1 (Fig.1.4) is close to Huangdao industrial park, which includes one of the three largest crude oil storage facilities of China with the capacity to stockpile more than 3 million tons of crude oil as well as a large oil-refining facility with a designed processing capacity of 10 million tons per year (Shen et al., 2006, Wang et al. 2009, Dang et al., 2010a,b). Huangdao industrial park also contains one of the largest sea terminals for the shipment of crude and refined oil. Despite the advanced eutrophic state of the water column, the much less impacted sediments revealed a highly diverse community of ammonia-oxidizing bacteria that facilitate aerobic (Dang et al., 2010a) and anaerobic (Dang et al., 2010a,b) removal of ammonium. This was the impetus for us to use sediments from this unique environment to search for unknown representatives of marine ammonia-oxidizing bacteria.

1.3.2 Characterization of the Ammonia/Ammonium-oxidizing Bacterial Community at the Sampling Sites

AOB and anammox bacteria have both been detected in sediment samples from

most of sampling stations (A3, A5, B2, C4, D1, D5, Dx and Y1) of Jiaozhou Bay. AOB marker gene, *amo*A (Dang et al., 2010b), and anammox bacterial maker gene, *hzo* (Dang et al., 2010a) were reproduced with the primer pairs amoA-1F and amoA-2R (Rotthauwe et al., 1997) and hzoF1 and hzoR1 (Li et al., 2010), respectively. However, the *amo*A primers employed in the above studies were suited to target only Beta-AOB but not Gamma-AOB. Campbell et al. (2011a) firstly reported the presence of Gamma-AOB in Jiaozhou Bay. The diversity of AOB and anammox were assessed via gene clone library and RFLP analysis, resulting of 43 unique *amo*A operational taxonomic units (OTUs), 41 anammox 16S rRNA OTUs and 56 *hzo* OTUs (Dang et al., 2010b , Li et al. 2010). Results from previous studies indicate that Jiaozhou Bay is a reservoir of diverse ammonia-oxidizing phylotypes. Station D1 had the lowest gene abundance for both Beta-AOB and anammox bacteria, but had the highest richness of OTUs (Dang et al., 2010a) among all 8 stations studied, indicating D1 is distinct from those of other stations. Environmental factors, especially OrgC/orgN ratio might have an important impact on it (Dang et al., 2010a).

1.4 Molecular Genetic Methods for Nucleic Acid Characterization

1.4.1 Sanger Sequencing and NGS Technology for DNA Analysis

DNA sequencing is the process of determination of base sequences in nucleic acids. There are two major types of sequencing technologies used today: Sanger sequencing and next-generation sequencing (NGS).

Sanger sequencing, also referred to as chain-termination or dideoxy method, is invented by Frederick Sanger and colleagues in the mid-1970s (Sanger et al., 1977) based on the use of dideoxynucleotides triphosphates (ddNTPs) in addition to the normal

deoxynucleotides triphosphates (dNTPs) found in DNA. The classical Sanger sequencing could only read about 80~120 base pairs per reaction. The industrial automation of the classical Sanger sequencing, has dramatically increased the sequencing efficiency and accuracy. Today's capillary electrophoresis (CE) sequencing instrument, such as ABI 3500 series Genetic Analyzers with 8-capillary (3500 System) or 24-capillary (3500xL System) array and polymer, can typically generate 750 to 1,000 base pairs of sequence from a single reaction and handle even more than 1,272 samples per day (http://www.thermofisher.com/us/en/home/mk-ab-moved.html).

Sanger sequencing has been the dominant sequencing technique until the emergence of the next generation sequencing (NGS) in 2005. NGS is also called high-throughput sequencing, utilizing massively parallel sequencing to generate thousands of mega bases of sequence data per day on a single chip whereas only one read (up to ~1kb) can be taken at a time in Sanger sequencing. Currently, there are four typical DNA sequencing methods applied in NGS systems: pyrosequencing; sequencing by synthesis, sequencing by ligation and Ion semiconductor sequencing (Fig.1.4). The major difference between NGS and Sanger sequencing is the combination of chemical reaction with the signal detection in NGS, whereas in Sanger sequencing these are two separate processes. All four sequencing methods generally follow the similar procedures: library preparation, colony generation, sequencing, and data analysis. Library preparation is the fragmentation of DNA/ cDNA samples or ligation with customized adaptor depending on the platform used. It is the underpinning of obtaining high-quality NGS results. Colony generation is to magnify library signals by amplification of each library fragment on a solid surface with covalently attached linkers to generate distinct and colonial clusters.

After obtained the raw sequencing data, usually we will analyze them through assembly, alignment and annotation of the sequences.

In this study, we used a new sequencing technology sometimes called third-generation sequencing, the Pacific Biosciences' (PacBio) Single Molecule Real-Time (SMRT) sequencing (Eid J et al. 2009), to obtain the genome sequence of *Nitrosococcus wardiae* D1FHS. The SMRT technology produces long reads with average lengths of 3,000 to 5,000 bp, and up to over 20,000 bp. It is performed on SMRT cells (Fig.1.7). Each SMRT cell is patterned with 150,000 Zero-Mode Waveguide (ZMW) reaction wells (Levene et al., 2003) which is now increased to one million by PacBio. It is necessary to check the sample purity before the start of sequencing (Fig.1.5). The input sample quality will severely affect sequencing results because PacBio sequencing doesn't use amplification technology in sample preparation (Fig.1.6). Nanodrop spectrophotometry is used for quantitative assessment while gel electrophoresis (Fig.1.5) is utilized for qualitative assessment in this study to ensure the DNA sample is clean and the amount is appropriate for sequencing. Sample preparation (Fig.1.6) is somewhat similar to NGS technologies. DNA samples (double-stranded DNA here as example) have to be fragmented on a DNA shearing device (i.e. Covaris® g-TUBE®) generating DNA fragments lengths of 6 kb to 20 kb depending on the quality and concentration of DNA samples. In a next step, the fragmented DNA are trimmed and ligated to the specific adaptors therefore generated structurally linear and topologically circular SMRTbell templates (Fig.1.6). The adaptor designed with a universal primer binding site and a short initiation sequence will help to protect the ends of the DNA fragments from damaging. The SMRTbell templates annealed with sequencing primers are bound to DNA

polymerase afterwards. The polymerase-template complexes are loaded into SMRT cell and immobilized at the ZMWs bottom (Fig.1.7) via Megabead methods. Ideally, each ZMW would have one polymerase-template complex bond at the bottom. It is ready to start sequencing after all reagents needed adding to the ZMWs (Fig.1.7). The DNA polymerase could incorporate bases at a speed of tens per second continuously. Similar to the "sequencing by synthesis" method, SMRT sequencing could detect signal occurs across all of the ZMWs on the SMRT chip simultaneously and continuously while DNA is synthesizing by polymerase. The output per day is ~ 20Gb with Sequel (http://allseq.com/knowledge-bank/sequencing-platforms/pacific-biosciences/). It is notable that SMRT sequencing single pass accuracy is relatively low (~86%) compared with other high-throughput sequencing methods (http://allseq.com/knowledge-bank/sequencing-platforms/pacific-biosciences/). However, because the raw read errors are distributed randomly, this can be overcome by generating sufficient sequencing passes. A 99.999% accuracy could be easily achieved with sufficient depth of coverage (http://allseq.com/knowledge-bank/sequencing-platforms/pacific-biosciences/).

1.4.2 Assembly and Annotation of Genomic DNA

The current whole-genome sequencing platforms are based on shotgun sequencing method generating a number of short fragments (reads). Shotgun sequencing can generate single-end or paired-end reads (Fig.1.8). Single-end reads are the fragments that are sequenced from one direction, whereas paired-end reads are sequenced from both 5' and 3' end. The reads need to be either aligned to a reference genome sequence if available or assembled *de novo* without reference input as desired to reconstruct the genome (Fig.1.8). Genome assembly in brief is to find overlaps between reads and joint

the overlaps into contigs (contiguous sequences) which can be linked to scaffolds (composed of contigs and gaps) to generate intact assemblies. Though the NGS technologies are faster and much less expensive, reads are shorter (50~500bp) than Sanger sequencing (750~1000bp). The advantage of long-read SMRT sequencing is outstanding in regard of genome assembly. Errors occurred in assembly, for example, reads are disregarded by mistakes or repeats or incorrectly joined up, could be minimized or avoided by collecting sufficient coverage. There are also software tools designed to correct those errors, such as NxRepair (Murphy et al., 2015). Here, sequence reads generated from PacBio SMRT sequencing were *de novo* assembled with the latest SMRT analysis algorithms, RS_HGAP_Assembly protocol v.3. The Hierarchical Genome Assembly Process (HGAP) generates comprehensive, highly accurate *de novo* assemblies using the non-hybrid approach (Chin et al., 2013). HGAP consists of pre-assembly, *de novo* assembly with Celera Assembler, and assembly polishing with Quiver.

Finding ways to assess the quality and validation of the assemblers could greatly improve its completeness and identification in general. In this study, the assembled genome was assessed by BLASTN and TBLASTN searches (CLC Main Workbench, Qiagen, Denmark) using the D1FHS 16S rRNA sequence and a collection of 31 single-copy genes (details in Chapter 3; Wu and Eisen 2008). The assembled draft genome was visualized in Mauve (multiple genome alignment) (Rissman et al., 2009, Darling et al., 2010), which can also help to identify rearrangements, deletions and insertions by mapping against the reference genome, *Nitrosococcus halophilus* Nc4.

Once the genome was assembled, the next step was to annotate the draft genome (Fig.1.8). Annotation based on the identification of protein-coding DNA sequences

(CDS) in bacterial genomes is not complicated by exon-intron organization because protein-encoding genes in bacteria are by and large intron-free. Once assembled, the genome sequence can be submitted to an automated web-based tool such as the RAST annotation server (Aziz et al., 2008). There are also many command-line annotation tools available (Fig.1.8). The annotation of all CDS in key pathways, including those for ammonia oxidation, respiration and carbon fixation, however, need to be manually refined by searching against KEGG (Kyoto Encyclopedia of Genes and Genomes; (Kanehisa and Goto 2000)), and the NCBI database on an individual gene-by-gene basis.

1.4.3 Real-Time Quantitative Polymerase Chain Reaction for mRNA Analysis

Genome sequencing and thorough genome annotation provide genetic information about the metabolic potential of an organism, while mRNA analysis reveals an overview of actively expressed genes and the stability of transcripts under various conditions. Essentially, there are three techniques suited for transcript analysis: gene expression microarrays (Tarca et al., 2006), RNA-seq (Wang et al., 2009) and the real-time quantitative Polymerase Chain Reaction (qPCR) (Higuchi et al., 1992). Microarray and RNA-seq offer a genome-wide gene expression profile, but microarrays can only detect selective sequences dependent on primers and probes used on the array while RNA-Seq provides unbiased, comprehensive data about the expression of every single transcript (Wang et al., 2009). In contrast, the qPCR technique is targeting a relatively small number of transcripts in a large set of samples at a high sensitive level, which is dependent on the specificity of primers. It combines the amplification and detection steps of the PCR reaction simultaneously using fluorescent dyes as reporters (Higuchi et al., 1993). Real-time reverse transcriptional PCR (RT-qPCR) can be performed as either one-

step or two-step reaction. One-step means that the entire reaction, from reverse transcription (mRNA transcribed to cDNA) to PCR amplification occurred in a single tube, whereas these two reactions are separate in the two-step reaction. We used two-step reaction in this study. Generally, there are absolute and relative quantification of qPCR. Absolute quantification is to determine the copy number of specific genes by comparing to a pertinent standard curve or by using digital PCR, which can quantify gene copies by number of digital PCR replicates. Relative quantification is to analyze the changes in gene expression in a given sample relative to the reference sample (i.e. nontreatment control sample) or an internal control, also known as a calibrator (i.e., a housekeeping gene). The standard curve method and comparative Ct ($2^{-\Delta\Delta Ct}$) method are the current two common methods used for relative quantification. As for using the standard curve method, the quantity of all experimental samples are first determined using a standard curve, which is then being divided by the target quantity of the calibrator sample. The calibrator is the 1-fold sample (here is the non-treatment control sample) whereas all experimental samples are expressed as an n-fold difference relative to the calibrator. The comparative Ct method compares the Ct value of one target to another and calculates changes in gene expression as a relative fold difference between an experimental sample and the calibrator. Here, we used comparative Ct method in order to study the differentially expressed genes in D1FHS under various treatments (see chapter 3).

The workflow of performing RT-qPCR to analyze gene expression levels can be simplified in four steps: RNA extraction, cDNA synthesis, PCR experiment, and data analysis (Fig.1.9; (Nolan et al., 2006)). The procedure is straightforward, however, it is essential to minimize variability and maximize reproducibility by quality-assessing

throughout the entire procedure to obtain reliable results. A similar sample size for all treatments and repeats is the first measure for reducing experimental error. This can be monitored by using near-identical culture volumes and cell pellet masses after centrifugation. The step of RNA assessment such as checking quantity and integrity is necessary after extracting mRNA using commercial kit, to ensure that a high quality RNA sample is being used for the RT-qPCR assay and also for data normalization (Perez-Novo et al., 2005). RNA integrity can be analyzed by examining samples using a Bioanalyzer or by agarose gel electrophoresis. Quantity and purity of RNA samples can be assessed using a Nanodrop spectrophotometer. Contamination with genomic DNA can be eliminated by performing an on-column DNase treatment directly following the RNA extraction (Qiagen, Germany). A reference gene (i.e. housekeeping gene) must be included in the RT-qPCR assay as a control to identify errors in the reaction and to detect differences in the RNA amounts used in the reverse transcription step (Pfaffl 2004).

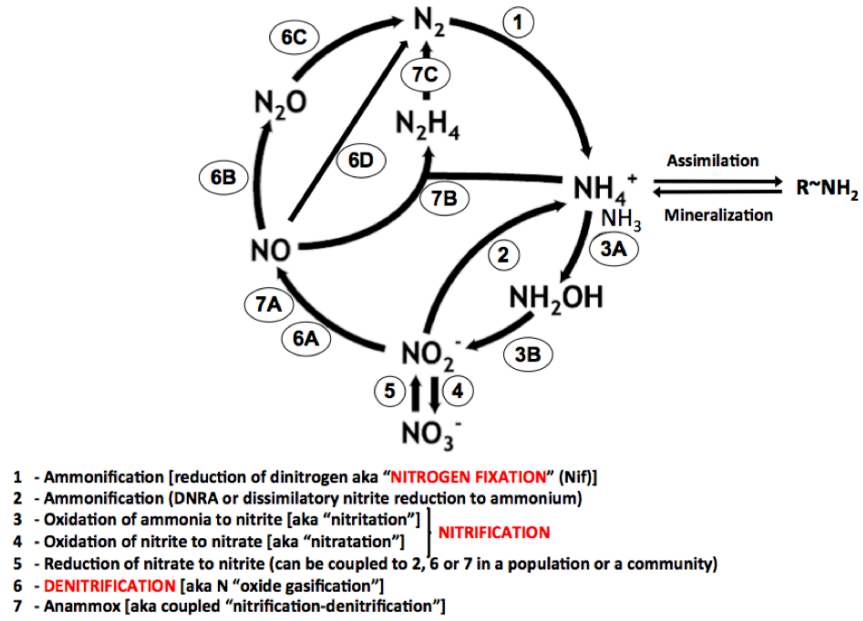


Figure 1.1: Overview of Nitrogen Cycle from (Stein and Klotz 2016)

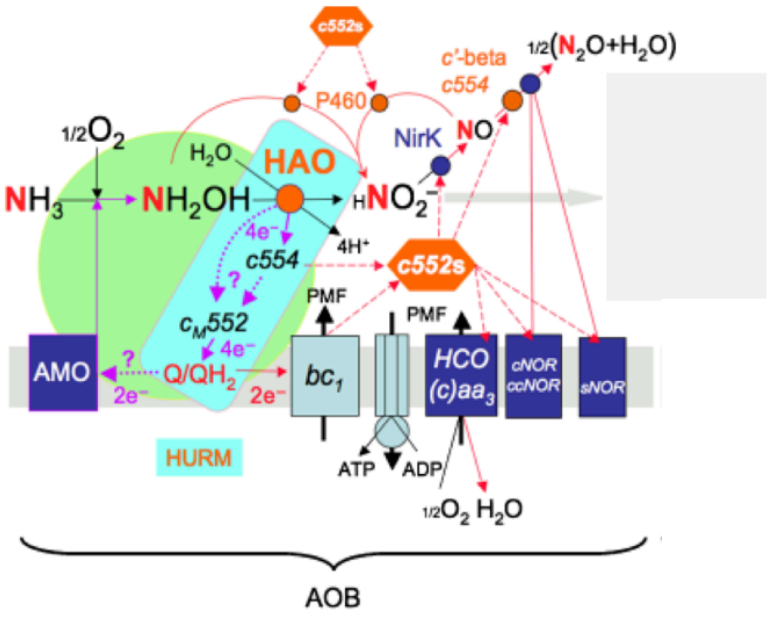


Figure 1.2: Flow of electron and reductant during nitrification process in ammonia-oxidizing bacteria (AOB). Solid lines show experimentally verified reactions, dotted lines with question marks indicate lack of experimental verification (Hooper, Arciero et al. 2005, Arp, Chain et al. 2007). REDOX reactions via dashed lines indicate that multiple non-identical copies of *c*552 are required to interact with a diverse set of reaction partners. The circuit for reverse electron flow was omitted. Colored backgrounds were used to highlight the cyclic electron flow module as well as the hydroxylamine/hydrazine-ubiquinone-redox-module, HURM. Blue-colored boxes indicate copper-dependent enzyme complexes. *(c)aa*$_3$, cytochrome *(c)aa*$_3$; *bc*$_1$, cytochrome *bc*$_1$ (complex III); NirK, Cu-dependent nitrate reductase; *c*'-β, cytochrome *c*'-β; c550, cytochrome *c*550; *c*552, cytochrome *c*552; *c*$_M$552, cytochrome *c*$_M$552; *c*554, cytochrome *c*554; NXR, nitrite oxidoreductase; P460, cytochrome P460; PMF, proton-motive force; Q/QH2, ubiquinone-ubiquinol pool; sNOR, cNOR, ccNOR, nitric oxide reductase with differing electron acceptor mechanisms [From (Klotz and Stein 2008)].

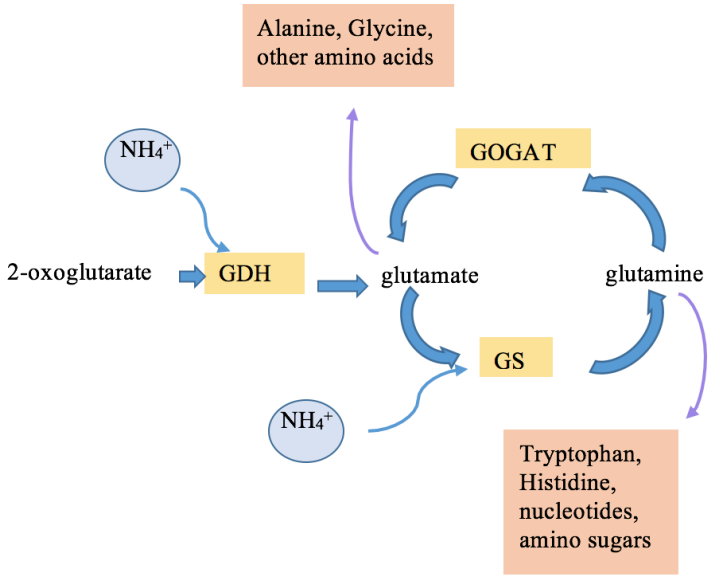


Figure 1.3: Nitrogen assimilation pathway in bacteria

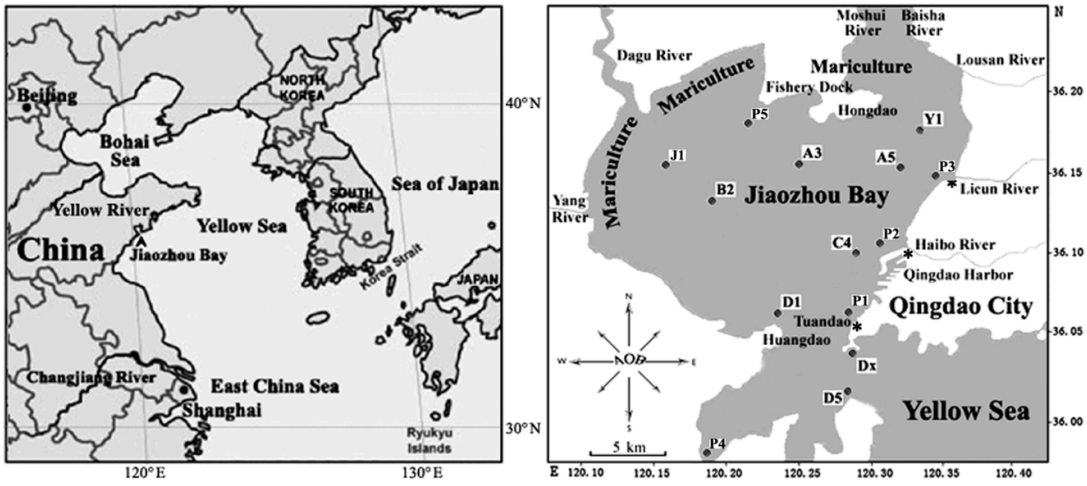


Figure 1.4: Map of Jiaozhou Bay and sampling stations

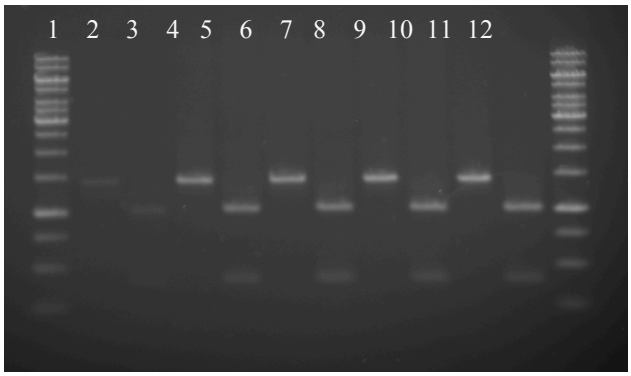


Figure 1.5: Assessment of sample purity using PCR and gel electrophoresis. Results of 16S PCR (primers set is 27F and 1406R) of original DNA prepared by multiple displacement amplification (El Rayes et al., 1991) based on published protocols (Zhang et al., 2006). Half of the PCR assays were subjected to digestion with PstI (a unique restriction site in the sequence). Both undigested and PstI-cut samples for all samples were loaded on the gel. Lanes 1 & 12 are the DNA ladder standard; Lanes 2 & 3 contain the original genomic DNA; the other lanes (lane 4 to lane 11) contain four pairs of MDAs

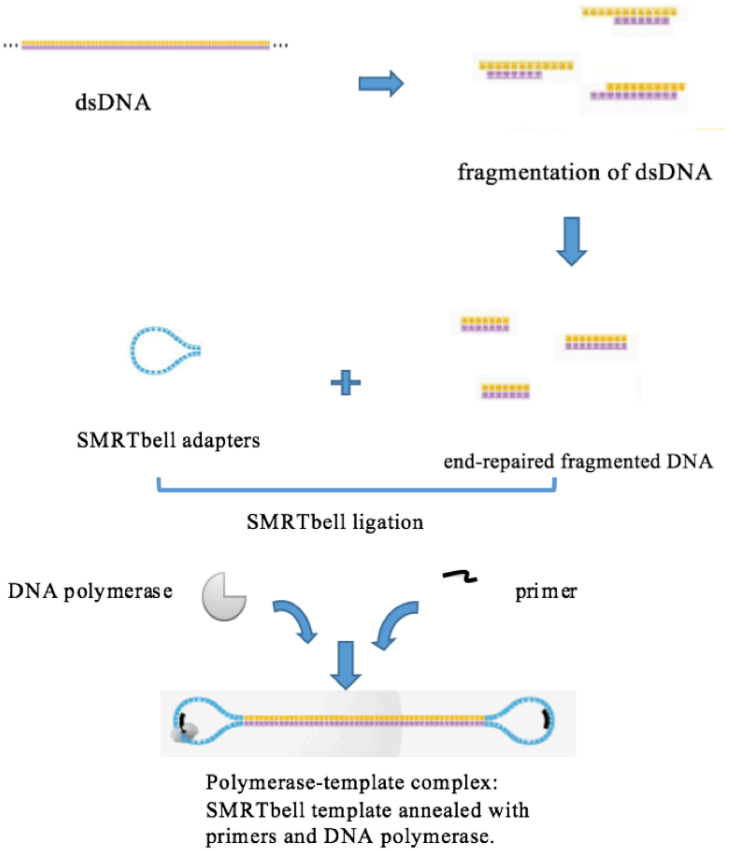


Figure 1.6: SMRT sequencing sample preparation procedure. Double-stranded DNA (dsDNA) is sheared into DNA fragments. The ends of DNA fragments are prepared to ligate with SMRT bell adapters. The SMRT bell templates are then annealed to primers, and the annealed templates are bound to DNA polymerase (Guide - Pacific Biosciences Template Preparation and Sequencing from http://www.pacb.com).

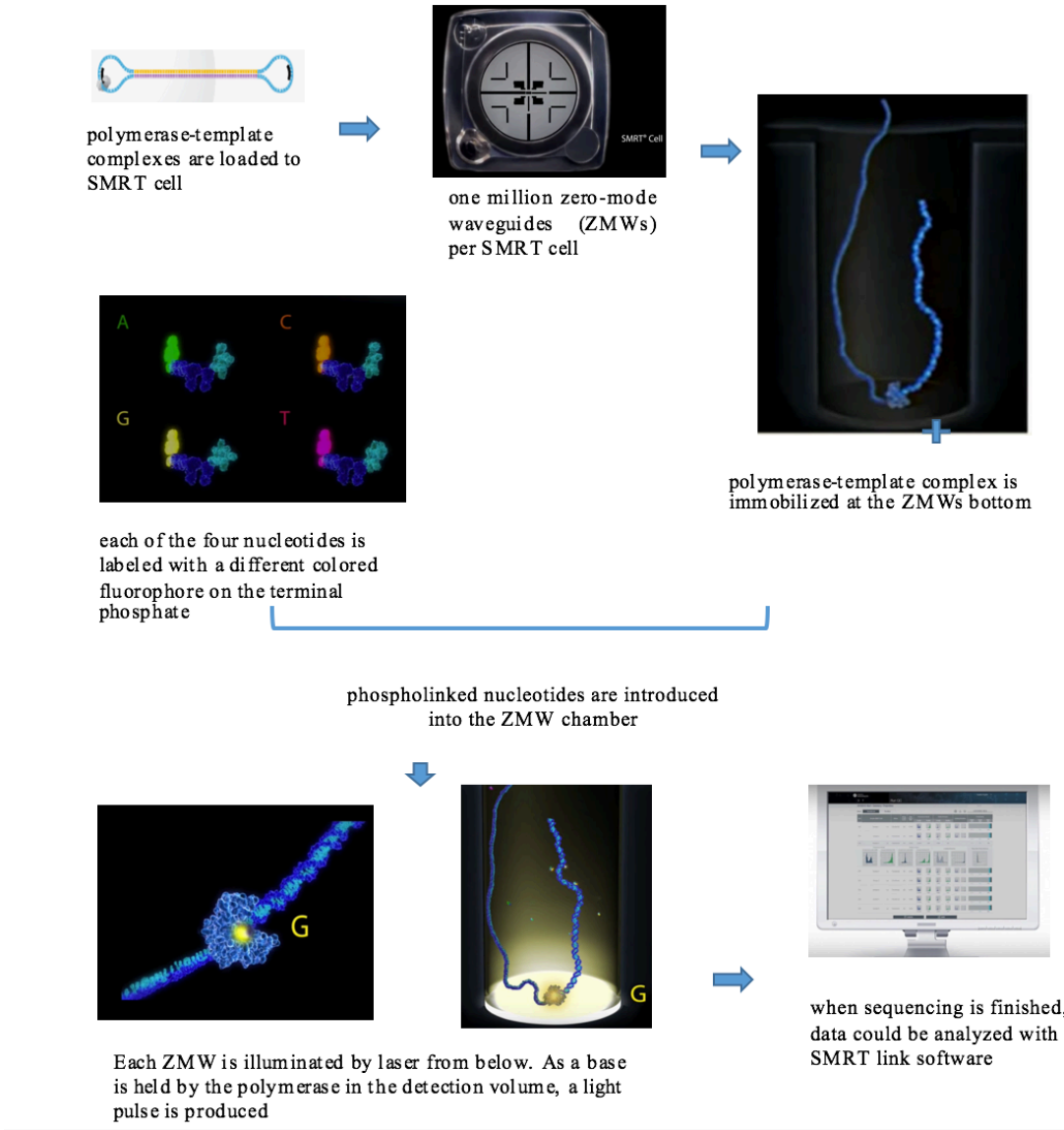


Figure 1.7: Workflow of PacBio SMRT sequencing. Polymerase-template complexes are immobilized to the ZMWs bottom surface. As each ZMW is illuminated from below, the wavelength of the light is too large to allow it to pass efficiently through the waveguide. As a base is held by the polymerase, it takes several milliseconds to incorporate, longer than the time it takes a nucleotide to diffuse in and out of the detection volume. The fluorophore label is then cleaved off together with the phosphate. Sequencing data can be analyzed on a computer with designed software.

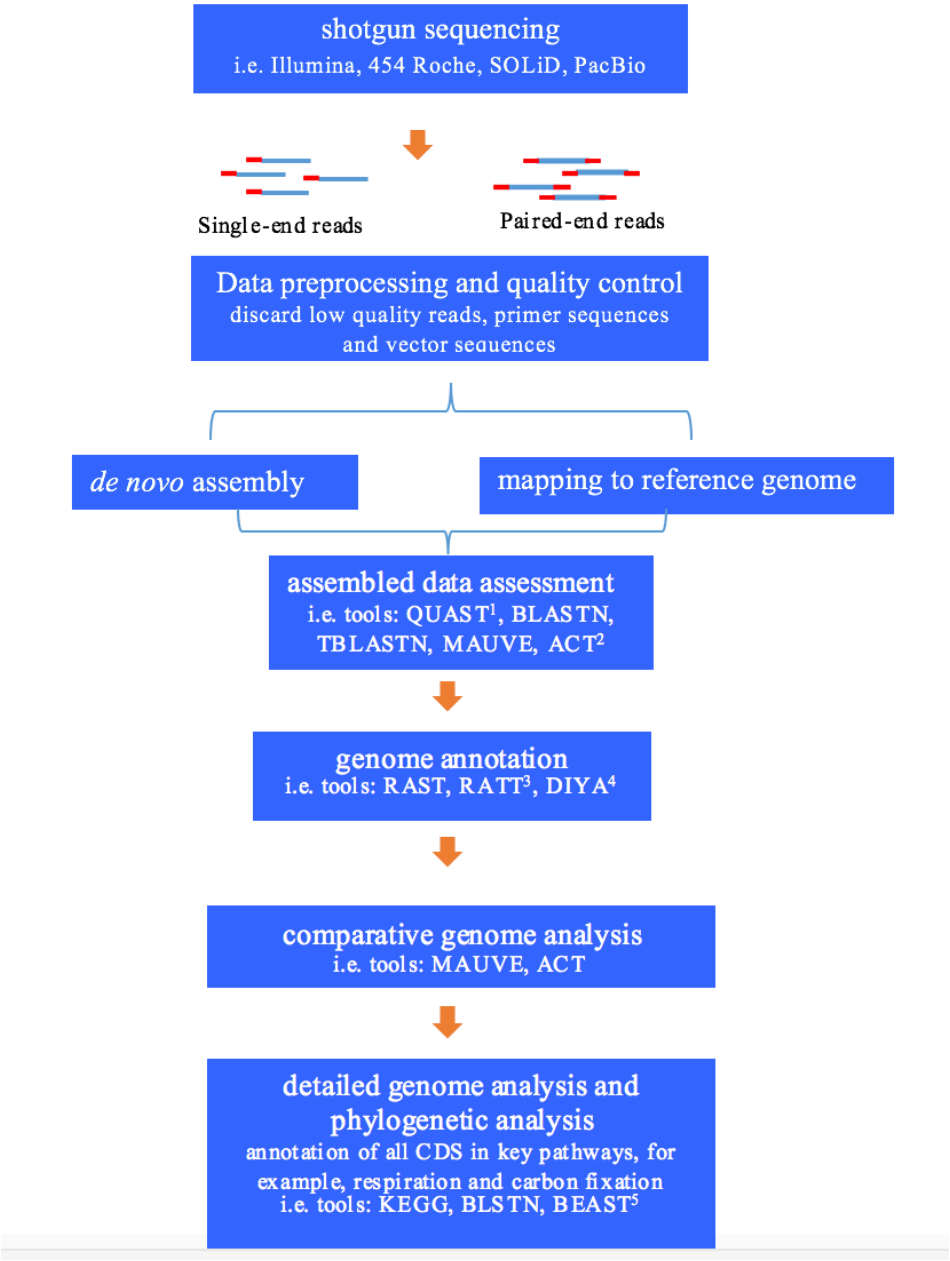


Figure 1.8: Workflow of genome assembly and annotation. [1]:QUAST: Quality assessment tool for genome assemblies (Gurevich et al., 2013); [2]:ACT: Artemis Comparison Tool. (Carver et al., 2005); [3]:RATT: Rapid Annotation Transfer Tool. (Sayavedra-Soto et al., 2011); [4]:DIYA: Do-It-Yourself Annotator (Stewart et al., 2009); [5]:BEAST v1.8.1 (Drummond et al., 2012).

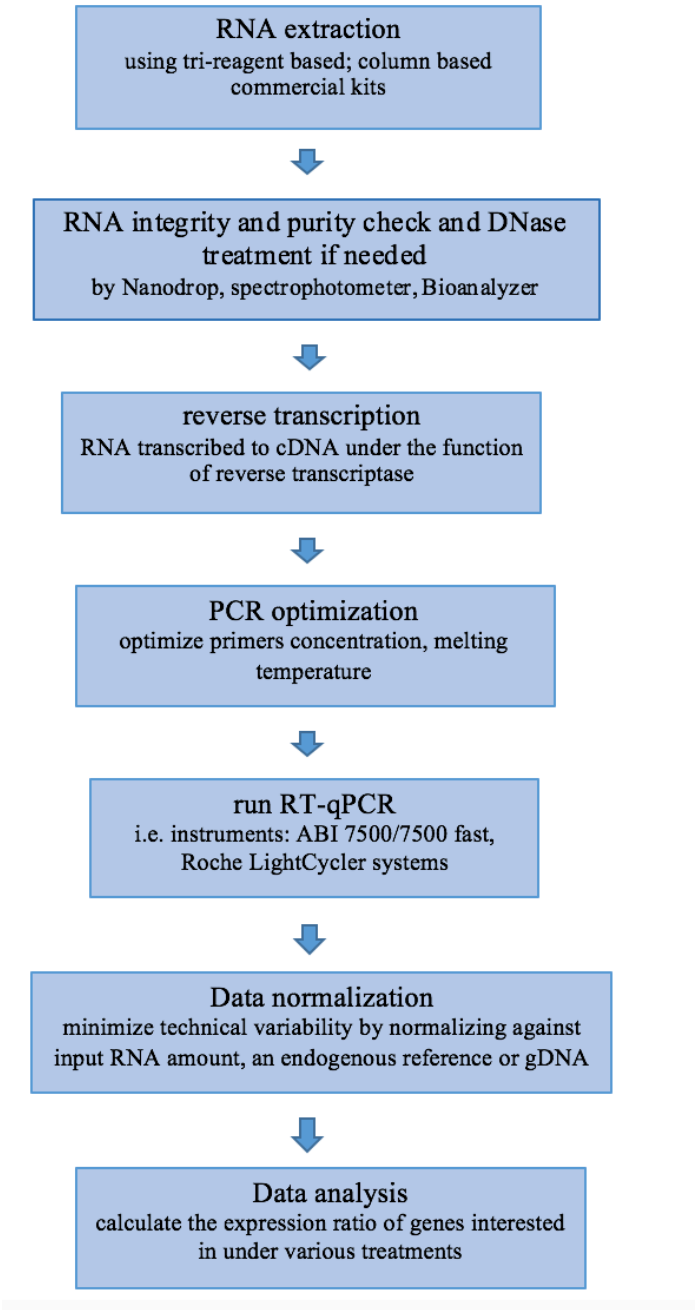


Figure 1.9: Workflow of RT-qPCR for mRNA analysis

CHAPTER 2: ENRICHMENT, ISOLATION, PHYLOGENETIC AND GROWTH PHYSIOLOGICAL CHARACTERIZATION OF A NOVAL AMMONIA-OXIDIZING BACTERIUM *NITROSOCOCCUS WARDIAE* D1FHS^T

2.1 Abstract

An ammonia-oxidizing bacterium, strain D1FHS, was enriched into pure culture from a sediment sample retrieved in Jiaozhou Bay, a hyper-eutrophic semi-closed water body hosting the metropolitan area of Qingdao, China. Based on initial 16S rRNA gene sequence analysis, strain D1FHS was classified in the genus *Nitrosococcus*, family *Chromatiaceae*, order *Chromatiales*, class *Gammaproteobacteria*; the 16S rRNA gene sequence with highest level of identity to that of D1FHS was obtained from *Nitrosococcus halophilus* Nc4^T. The average nucleotide identity between the genomes of strain D1FHS and *Nitrosococcus halophilus* strain Nc4 is 89.5%. Known species in the genus *Nitrosococcus* are obligate aerobic chemolithotrophic ammonia-oxidizing bacteria adapted to and restricted to marine environments. The optimum growth (maximum nitrite production) conditions for D1FHS in a minimal salts medium are: 50 mM ammonium and 700 mM NaCl at pH of 7.5 to 8.0 and at 37°C in dark. Because pertinent conditions for other studied *Nitrosococcus* spp. are 100-200 mM ammonium and <700 mM NaCl at pH of 7.5 to 8.0 and at 28-32°C, D1FHS is physiologically distinct from other *Nitrosococcus* spp. in terms of substrate, salt and thermal tolerance.

2.2 Introduction

Jiaozhou Bay (36°7′24.44″N 120°14′44.3″E) (Fig. 2.1) is a semi-enclosed water body with a surface area of about 390 km^2 and an average water depth of 7 m, located on the Southwestern coast of Shandong Peninsula, China. It connects to the Chinese Yellow Sea with a narrow mouth, about 2.5 km wide (Shen et al., 2006). The bay is bordered by Qingdao City, a large industrial and agricultural city in North China with a population of approximately 9 million residents. Freshwater is provided to Jiaozhou Bay by numerous tributaries (Yang River, Dagu River, Moshui River, Baishahe River, Wantou River, Loushan River, Banqiaofang River, Licun River, Haipo River. etc). Most of these rivers have become the canals for industrial, agricultural and urban wastes and are thus highly enriched in nutrients including concentrated dissolved inorganic nitrogen and inorganic phosphate sources (Shen et al., 2006). One of the frequently sampled locations is Station D1, (Fig. 2.1 and (Dang, et al. 2008, 2009b, 2010a)). D1 is close to Huangdao industrial park, which includes one of the three largest crude oil storage facilities of China with the capacity to stockpile more than 3 million tons of crude oil as well as a large oil-refining facility with a designed processing capacity of 10 million tons per year (Shen et al., 2006, Dang et al., 2009a, b, 2010a). Huangdao industrial park also contains one of the largest sea terminals for the shipment of crude and refined oil. Despite the advanced eutrophic state of the water column, the much less impacted sediments revealed a highly diverse community of ammonia-oxidizing bacteria that facilitate aerobic (Dang et al., 2010c) and anaerobic (Dang et al., 2010a) removal of ammonium. This was the impetus for us to use sediments from this unique environment to search for unknown representatives of marine ammonia-oxidizing bacteria.

Obligate aerobic chemolithotrophic ammonia-oxidizing bacteria (AOB) utilize ammonium as the sole source for energy and reducing power for growth by oxidizing its non-protonated form, ammonia, to nitrite with hydroxylamine as the intermediate. While initially defined primarily by differences in cell morphology and physiological characteristics (Watson and Mandel 1976), the advent of 16S rRNA gene phylogeny led to the definition of three genera of AOB (Head et al., 1993), which was largely confirmed by the sequencing of available representatives in pure culture (Arp et al., 2007, Campbell et al., 2011a). While the betaproteobacterial AOB in the family *Nitrosomonadaceae* represent 2 genera with numerous species (Urakawa et al., 2015), *Nitrosococcus* is the only genus of gammaproteobacterial AOB (gamma-AOB, (Campbell et al., 2011a)). In the past, research of the ecology and phylogeny of AOB has almost entirely focused on the beta-AOB because this group is abundant in all oxic environments, accessible to molecular ecological analysis because the evolutionary histories of enzymes that are key to ammonia catabolism are monophyletic and congruent with 16S rRNA gene-based phylogeny, and beta-AOB usually grow faster than gamma-AOB (Kowalchuk and Stephen 2001 , Campbell et al., 2011a). Gamma-AOB in the genus *Nitrosococcus* are adapted to and restricted to marine environments (Table 2.1; (Ward and O'Mullan 2002)) and strains representing three species are being maintained in pure culture (Campbell et al., 2011a).

As recently summarized (Campbell et al., 2011a), the type strain of the genus *Nitrosococcus* was originally described by Winogradskyi (1892), renamed ‘Micrococcus nitrosus’ by Migula (Migula 1990) and finally established as *Nitrosococcus nitrosus* (Buchanan 1925). The binomial was approved as valid by Skerman et al. (1980);

however, by then the culture was already lost thereby rendering the genus without a validly described type strain in culture (Campbell et al., 2011a). Because Winogradskyi enriched *Nitrosococcus nitrosus* reportedly from freshwater, marine strain D1FHS isolated into pure culture (from station **D1** in Jiaozhou Bay, **F**ermented in a bioreactor and cultured in the presence of **H**ydroxylamine at 700 mM NaCl instead of the typical 500 mM in mineral **S**alts media) cannot be assigned to the valid type species of the *Nitrosococcus* genus. The characterization of its properties indicates that D1FHS is physiologically and taxonomically distinct from other *Nitrosococcus* spp. leading us to propose it as the type strain of *Nitrosococcus wardiae* spec. nov. Strain D1FHS has been submitted to the Japan Collection of Microorganisms (JCM) and the American Type Culture Collection (ATCC) for deposit.

2.3 Materials and Methods

2.3.1 Sample Collection, Cultivation and Screening of the Enrichment

Sediment samples were collected from station D1 (120.16438°E, 36.00879°N) and from 13 other stations of Jiaozhou Bay (Fig. 2.1) on October 26, 2008, as described previously by Dang (Dang et al., 2009a). Briefly, surface sediment subcore samples were collected from14 stations including D1 (Fig. 2.1) in summer 2008 using a stainless steel 0.05-m^2 Gray O'Hara box corer. Replicate samples were taken from inside the box core reaching to a depth of 5 cm, placed in sterile Ziploc bags and stored on dry ice and at 80°C after returning to the laboratory. Water was collected at each of the 14 sampling sites and temperatures, DO, salinity and pH of all surface sediments were measured (Table 2.2. SI). Only sediment samples from station D1 were used for this study.

Enrichments were started by mixing 0.5 g of frozen sediment into 100 mL of

ammonia mineral salts media (AMS: 700 mM NaCl, 12.5 mM $(NH_4)_2SO_4$, 30 mM $MgSO_4$, 20 mM $MgCl_2$, 10 mM $CaCl_2$, 10 mM KCl, 0.2 mM $NaCO_3$, 3.0 mM $NaHCO_3$, 0.09 mM K_2HPO_4, 3 μM chelated iron, 0.4 μM $NaMoO_4$ 1.0 μM $MnCl_2$, 0.008 μM $CoCl_2$, 0.35 μM $ZnSO_4$, 0.08 μM $CuSO_4$, phenol red 0.5% w/v) amended with hydroxylamine at 100 μM final concentration (AMS-H) in a 250 ml Erlenmeyer flask (Pyrex, USA) and maintained as standing batch cultures at 30 ºC, in the dark, without shaking (Campbell et al., 2011a). Phenol-red was used as a pH indicator and pH was adjusted daily to 7.5~8.0 using 0.25 M K_2CO_3. Batch cultures were propagated monthly by transferring 5 mL of culture into 100 mL of fresh AMS-H medium using a sterile pipette (Corning, USA). Hydroxylamine was added weekly to a final concentration of 200 μM to eliminate susceptible bacteria and archaea (Campbell et al., 2011a). Nitrite (NO_2^-) was measured in spent medium samples (1.3 mL) as an indicator of catabolic activity of nitrifying bacteria by the sequential addition of 0.5 mL of 1% sulfanilamide and 0.5 mL of 0.02% N-(l-napthyl) ethylenediamine dihydrochloride using sterile medium as a control (Nicholas and Nason 1957, Koops et al., 1990). Following incubation for 20 min at room temperature, absorbance at 543 nm was measured using a Nanodrop 200C spectrophotometer (Thermo Fisher Scientific, USA). Contamination with heterotrophs potentially introduced during handling and transfer of cultures was assessed by plating an aliquot on LB agar and incubation at 35°C.

Biomass was harvested from 200 mL of D1 enrichment in exponential phase (based on NO_2^- measurement) by centrifugation at 4,300 g for 15 min using a Sorvall Evolution centrifuge and SS-34 rotor (Thermo Fisher Scientific, USA). The resulting cell pellet was resuspended in water to a volume of 200 μl for metagenomic DNA extraction

using the MP FastDNA Spin Kit (MP Biomedicals, USA) and a MP FastPrep®-24 Instrument (MP Biomedicals, USA) (Poret-Peterson et al., 2008). Genomic DNA, primers targeting universal and gammaproteobacteria-16S rRNA-specific sequences of the 16S rRNA gene (Campbell et al., 2011a) as well as primers specific for detection of the *hao*A gene from gamma-AOB (Schmid et al., 2008) and the polymerase chain reaction (PCR) were used to generate inserts for clone libraries (Table 2.3). Amplicons (universal 16S ~1470 bp, gamma-16S ~1500 bp and gamma-*hao*A ~1220 bp) were generated using GoTaq Green Master® Mix (Promega, USA) on a gradient mastercycler (Eppendorf, Germany) following an optimized PCR protocol (30 cycles of 1 min at 95 °C, 1 min at annealing temperature (T_m): 55 °C for universal 16S, 56 °C for gamma 16S and 60 °C for gamma *hao*A, 1 min extension time at 72 °C). For each 50-ul PCR reaction, 2 ul of extracted D1 metagenomic DNA was used as the template. The PCR using gamma-16S rRNA-gene primers generated products of the expected length; therefore, PCR products obtained with universal 16S rRNA and gamma-*hao*A gene primers were gel-purified (1.5% low-melt agarose) and ligated into pCR®2.1-TOPO TA vectors (Invitrogen, USA). The overnight ligation products were used to transform TOP10 competent cells for clone library construction. Forty randomly picked recombinants from each clone library were selected using X-Gal-LB plates with 25 μg/ml kanamycin (Dang et al., 2008, 2009b, 2011, Campbell et al., 2011a). Plasmids were extracted using the alkaline mini-prep method (Promega, USA) and inserts were sequenced with vector primers, M13F and M13R (Table 2.3; DNA Core Facility, University of Louisville). Insert sequences were trimmed, aligned using MUSCLE (Darling et al., 2010) provided at the EMBL-EBI webserver

(http://www.ebi.ac.uk/Tools/msa/muscle/) and analyzed for closest-matches with sequences deposited in GenBank using the BLASTN program (Altschul et al., 1997).

2.3.2 Isolation of D1FHS into Pure Culture, DNA Extraction, Genome Sequencing, and Annotation

Sample aliquots of 0.1 ml were sequentially diluted, plated on sterile semi-solid (0.3% agar) AMS-H medium and incubated in the dark at 30 °C (Tittsler and Sandholzer 1936). After two months, visible single colonies were picked and transferred into a 6-well plate with 5 ml liquid AMS-H media. Subsamples (2.5 ml) of wells that turned yellow, indicating acidification due to nitrous acid production, were transferred into 250 ml Erlenmeyer flasks with 100 ml AMS-H media. Enriched culture (20 mL) was used to inoculate a benchtop bioreactor (BioFlo® / CelliGen® 115, New Brunswick/Eppendorf, Germany) with 1 L working volume of AMS-H media sterilized by autoclaving (121 °C, 20 min, 100 Kpa, 15 psi). The chemostat, operated in batch culture mode, was incubated in the dark at 28 °C with agitation of 50 rpm and 1.0 LPM flow of 0.22 μm filtered air; pH was maintained at 8.0 by automated addition of sterile 0.5 M K_2CO_3. Cell density (OD_{600}) and nitrite concentration were measured daily by withdrawing a 5-ml sample from the reactor. Continuous culture for 7-10 days was alternated with batch culture (500 mL in 2-L Erlenmeyer flask) several times; each time an aliquot of the extracted genomic DNA was used to build a universal 16S rRNA cloning library for insert sequencing and analysis of sequence from 40 randomly picked clones (see text above) after each alternation. The alignments were inspected for single nucleotide polymorphisms. Once a single 16S rRNA sequence was observed across all clones in multiple subcultures, the culture was deemed pure and labeled “D1FHS.”

To inspect morphological homogeneity of the D1FHS culture, cells from 50 ml of a 10-day old culture were harvested by centrifugation (4,300 g, 15 mins). The cell pellet was washed twice with AMS medium and resuspended in 10 µl of AMS medium. A 2-µl aliquot of cell suspension was spread on a glass slide and covered with a cover slip to observe cell motility with a light microscope (Olympus, Japan) at 400x magnification. An 8-µl aliquot of cell suspension was heat-fixed on a slide for Gram staining (Sigma-Aldrich, USA) and observed using a light microscope (Olympus, Japan) at 1000x magnification.

For genome sequencing, ~1.2 µg of gDNA were isolated from a 500-mL mid-exponential-phase culture. To obtain sufficient DNA for Pacific Biosciences (PacBio) single molecule sequencing, the purified genomic DNA was amplified by multiple displacement amplification (El Rayes et al., 1991) based on published protocols (Zhang et al., 2006). Briefly, 100 ng amounts of DNA were amplified in five MDA reactions (REPLI-g Mini, Qiagen). MDA products from each reaction (~4 µg per reaction) were sequentially treated with 10 U Phi29 (Fermentas) and 100 U S1 nuclease (Thermo Fisher Scientific, USA) followed by precipitation (SureClean, Bioline) and resuspension in sterile 18 MΩ water. The identity of the amplified DNA was verified by PCR amplification with universal 16S rRNA primers. Half of the PCR product was digested with *Pst*I (Fermentas) to verify that they contained a single site as predicted and the remainder was directly sequenced with the amplification primers to verify that the sequence matched that from D1FHS. Following confirmation, the amplified DNA (~15 µg total) was submitted for PacBio sequencing at the University of Delaware Genotyping and Sequencing Core Facility, University of Delaware. Sequence reads were assembled

using RS_HGAP_Assembly protocol v.3 on a computational cluster maintained by the University of Delaware Center for Bioinformatics and Computational Biology. Completeness of the assembled genome was assessed by BLASTN and TBLASTN searches (CLC Main Workbench, Qiagen, Danmark) using the D1FHS 16S rRNA sequence and a collection of 31 single-copy genes (Wu and Eisen 2008). Biosample (SAMN04324281) and Bioproject (PRJNA305330) information has been submitted to the NCBI.

2.3.3 HaoA Protein Sequence Analysis

HaoA protein sequences from *Nitrosococcus* species*,* including *N. oceani* strains C-107, C-27, AFC132, and AFC27, *N. watsonii* C-113, and *N. halophilus* Nc4, the deduced HaoA sequence from D1FHS and pertinent sequences of the "HAO" clade of octaheme cytochrome *c* proteins were aligned using MUSCLE and manually refined by comparison with previously published results from phylogenetic analyses including structural and protein sequence-analytical features (El Sheikh and Klotz 2008, Kern et al., 2011). N- and C-terminally extending sequences beyond the first and last heme-binding motif (CxxCH), respectively, were trimmed and the final alignment was subjected to a Bayesian inference of phylogeny using the BEAST package (v1.8.1 of BEAUti, BEAST and TreeAnnotator; FigTree v.1.4; (Drummond et al., 2012)). By utilizing unique sites, tree likelihoods (ignoring ambiguities) were determined for the alignment by creating a Monte-Carlo Markov Chain (10,000,000 generations) in three independent runs. The searches were conducted assuming an equal distribution of rates across sites, sampling every 1000th generation and using the WAG empirical amino acid substitution model

(Whelan and Goldman 2001). The resulting 10,000 trees (omitting the first 350 trees as burn-in) were used to construct a phylogenetic consensus tree.

2.3.4 Average Nucleotide Identity Between Bacterial Genomes

The Average Nucleotide Identity (ANI) calculated for pairwise compared genomes is an advanced method developed to identify species boundaries for bacteria and archaea (Konstantinidis and Tiedje 2005). The species delineation cut-off point is an ANI of 94% (Konstantinidis and Tiedje 2005, Goris et al., 2007), which corresponds to the traditional species cut-off of 70% using DNA-DNA hybridization (Wayne et al., 1987). The high quality draft genome sequence of strain D1FHS (single chromosome, no plasmid) was compared with the genome sequences of *N. halophilus* Nc4 (CP001798; plasmid: CP001799), *N. oceani* ATCC19707 (CP000127; plasmid: CP000126), *N. watsonii* C-113 [CP002086; plasmids: CP002087 & CP002088] as well as those of other Chromatiaceae, including *N. oceani* strain AFC-27 [ABSG00000000], *N. oceani* strain AFC132 [SRS650332], *N. oceani* strain C-27 [SRS642301], *Allochromatium vinosum* strain DSM 180 [CP001896; plasmids: CP001897 & CP001898], *Halorhodospira halophila* strain SL1 [CP000544], *Alkalimnicola ehrlichei* strain MLHE-1 [CP000453], *Thioalkalivibrio* sp. strains HL-EbGR7 [CP001339] and K90mix [CP001905; plasmid: CP00196] and *Halothiobacillus neopolitanus* strain c2 [CP001801] as previously done for the identification of *Nitrosococcus watsonii* as a species (Campbell et al., 2011a). The ANI values reported in this paper were calculated using the tool developed and provided online by Luis M. Rodriguez-R and Kostas T. Konstantinidis (http://enve-omics.ce.gatech.edu/ani/) using default parameters.

2.3.5 Physiological Assays and Phenotypic Assessment

One-ml aliquots of mid-exponential-phase "D1FHS" pure culture were inoculated into 25 ml HEPES-buffered AMS-H media using 125 ml flasks and incubated in the dark at 30 °C without shaking to determine growth as a function of time, salt and ammonium tolerance as well as optimum pH and temperature. A growth curve was estimated by measuring nitrite concentrations at eight times (days 0, 1, 3, 5, 6, 8, 10 and 12) after inoculation in AMS-H standard media (700 mM NaCl and 12.5 mM $(NH_4)_2SO_4$ at a set pH of 8). AMS-H media at 8 different NaCl concentrations ranging from 100 mM to 1600 mM were used to determine the optimum NaCl concentration for the growth of D1FHS. Growth of the D1FHS culture was also tested at eight pH conditions (pH of 4, 5, 6,7, 7.5, 8, 9 and 10) and at five temperatures (4 °C, 20 °C, 28 °C, 37 °C and 45 °C). Ammonium tolerance was examined by testing ten ammonium concentrations (using $(NH_4)_2\,SO_4$) ranging from 0 to 600 mM. Cell cultures were monitored for nitrite production routinely. All tests were performed with three technical replicates and repeated with two or more biological replicates. Like *N. halophilus* Nc4 and unlike cultures of *N. oceani* C-107 and *N. watsonii* C-113 (Campbell et al., 2011a), D1FHS did not grow on urea when provided as the sole source of ammonium (N, energy and reductant). Flagellar motility was observed under a light microscope at 1000x magnification.

2.4 Results and Discussion

2.4.1 Characterization of The Enrichment and Pure Isolate Cultures

PCR performed with metagenomic DNA extracted from D1 enrichment culture targeting at universal 16S rRNA, gamma 16S rRNA and gamma *haoA* genes, yielded

correctly sized fragments (universal 16S ~1470 bp, gamma-16S ~1500 bp and gamma-haoA ~1220 bp, respectively), indicating that the D1 enrichment culture contained gamma-AOB. Universal 16S rRNA gene sequences were analyzed based on BlastN searches of the NCBI database; 29 of the 40 sequences were best hits to 16S rRNA gene sequences from *Nitrosococcus* species: 87%~98% identity to *Nitrosococcus halophilus* Nc4; 93%~98% identity to *Nitrosococcus oceani* C-107, and 93%~98% identity to *Nitrosococcus watsonii* C-113. The rest of the sequences had best hits to sequences from *Mesorhizobium*, *Janibacter* sp. or uncultured bacteria.

Following the repeated enrichment procedure described in the experimental procedures that involved selection of colonies from semi-solid agar, alternating batch and continuous culture and the application of functional pressure by exposure to hydroxylamine and high sodium salt concentration, isolated genomic DNA from the final culture labeled "D1FHS" was, again, investigated using primers designed to detect universal 16S rRNA, gamma 16S rRNA and gamma *haoA* genes and the PCR. Analysis of the obtained amplicon sequences using BlastN consistently revealed highest identity with sequences from *Nitrosococcus halophilus* Nc4 and no clone had a best hit with sequences from bacteria outside of the *Nitrosococcus* genus.

2.4.2 Genome Sequencing and Characteristics

PacBio-based high-throughput sequencing of D1FHS DNA produced 2,267,036,463 bp of sequence in 191,850 reads (N50 read length = 16,723 bp, mean read length = 11,816 bp). These were assembled into a high quality draft sequence containing one contig of 4,022,640 base pairs; the genome has a GC content of 50.7%. Homologs were detected for each of 31 universal single copy genes (Wu and Eisen 2008) by

TBLASTN searches and two copies of the D1FHS 16S rRNA sequence by BLASTN indicating that the assembled contig contains a complete bacterial genome. The sequence of the D1FHS genome was analyzed using the RAST annotation server (Aziz et al., 2008) and found to contain 4,128 protein-coding DNA sequences, 51 RNA genes including two 16S-23S-5S rRNA operons (RAST-ID: 6666666.126356). The genome sequence of D1FHS included genetic markers known from other *Nitrosococcus* genomes: the genome of D1FHS encodes a complete set of protein inventories required for ammonia-dependent chemolithotropic growth, including ammonia monooxygenase (AMO, EC 1.14.99.39), hydroxylamine dehydrogenase (HAO, EC 1.7.2.6), cytochromes *c*554 and c_M552 as well as nitrosocyanin, although the functions of cytochromes *c*554 and nitrosocyanin are still elusive (Stein et al., 2013). Based on the identification of signal peptides, all of these proteins operate and have access to their reactants within the periplasm or at the periplasmic side of the plasma membrane (Arp et al., 2007). Consistent with the inability of D1FHS to grow on urea, the genome did not encode urea hydrolase.

2.4.3 Average Nucleotide Identity and HaoA Protein Phylogenetic Analysis

The ANI values for the comparison of the D1FHS genome with the genomes from Proteobacteria outside the genus *Nitrosococcus* were low as previously reported for the analysis of the *Nitrosococcus watsonii* C-113 genome (Campbell et al., 2011a). ANI values calculated between the D1FHS genome and all available *Nitrosococcus* genomes including *N. halophilus* strain Nc4 (89.5%), *N. watsonii* C-113 (78.2%) and *N. oceani* (C-107, 78.4%; C-27, 78.2%; AFC27, 78.4%; AFC132, 78.2%) are significantly below the cut-off for the delineation of species (Konstantinidis et al., 2006), which suggests that strain D1FHS is representative of a species distinct from *N. halophilus*, *N. watsonii* and

N. oceani most closesly related to *N. halophilus*. We propose to assign D1FHS as the type strain of *Nitrosococcus wardiae* spec. nov.

Phylogenetic analysis of protein sequences confirmed that the HaoA proteins from D1FHS and *N. halophilus* strain Nc4 are more closely related to one another than either is to HaoA proteins from strains of *N. watsonii* or *N. oceani* (Fig. 2.2). Identical results were obtained from phylogenetic analysis of the other proteins implicated in ammonia-dependent chemolithotrophy (Arp et al., 2007, Stein et al., 2013): ammonia monooxygenase, cytochrome c_M552 and nitrosocyanin (data not shown). The phylogram in Fig. 2.2 supports the conclusion that the ancestor of the *Nitrosococcus* genus diverged into two phylotypes that gave rise to *N. wardiae* spec. nov. and *N. halophilus* as well as *N. watsonii* and *N. oceani*. It was previously hypothesized that the *hao* genes encoded on the chromosome (soil methane-oxidizing) or a plasmid (marine sulfur-oxidizing) of Alphaproteobacteria and the *hao* genes in the genomes of betaproteobacterial AOB were obtained by horizontal transfer from Gammaproteobacteria (Klotz et al., 2008, Kern et al., 2011, Stein et al., 2011). The position of the Hao protein sequence from the "comammox" bacterium *Nitrospira inopinata* ERN4 capable of oxidizing ammonia all the way to nitrate (Daims et al., 2015) included in this analysis suggests that the *hao* gene cluster has been horizontally transferred from Gammaproteobacteria into the genomes of nitrite-oxidizing *Nitrospira* bacteria ancestral to the comammox lineage as well. Preliminary analysis of the D1FHS genome suggests that its genome is likely most representative of the ancestral *Nitrosococcus* genome; a detailed analysis of the genome is forthcoming.

2.4.4 Growth-Physiological Characterization of D1FHS

In batch culture, cells of D1FHS routinely exhibited a lag time of 5 days before entering exponential growth phase. Concomitantly, cells started dividing and Gram-staining revealed that cells appear as short rods, arranged singly or in pairs but not in tetrads (Fig. 2.3). The stain also confirmed the existence of only one cellular morphotype. In agreement with strains of other *Nitrosococcus* species, D1FHS contained intracellular membrane stacks, was lophotrichous and capable of flagellar motility (data not shown).

The nitrite production rate was used to assess growth physiological characteristics of D1FHS ("optimum catabolism"), which was calculated based on a nitrite standard curve (calculated as μmol per L of media). The pH, temperature, NaCl and ammonium tolerances were determined by measuring the maximum nitrite production of log-phase cultures (day 5 to day 8; Fig. 2.4a). Based on the observed maximum nitrite production rate, the optimum growth temperature for D1FHS is 37 °C (Fig. 2.4b), which is much higher than those reported (also based on nitrite production) for other strains of the genus *Nitrosococcus* (Table 5; (Koops et al., 1990, Purkhold et al., 2000, Campbell et al., 2011a)). There is presently no published information for any *Nitrosococcus* strain about the need for maintenance energy at different temperatures. No growth (no changes in nitrite levels) was observed at 4 °C and 52 °C, but D1FHS tolerated 45°C (Fig. 2.4b). D1FHS was able to grow at a pH ranging from 5 to 9 with an optimum pH between 7.5 and 8.0 (Fig. 2.4c), which is in the same range reported for other strains of the genus *Nitrosococcus* (Table 5). D1FHS was not able to tolerate ammonium concentrations higher than 300 mM, which represents a much lower tolerance than what is known for other *Nitrosococcus* strains (Table 2.4, Fig. 2.4d; (Koops et al., 1990, Purkhold et al.,

2000, Campbell et al., 2011a)); consequently, the optimum ammonium concentration of 50 mM is also the lowest known optimum for all characterized strains of *Nitrosococcus*. D1FHS exhibited a NaCl tolerance similar to that of *Nitrosococcus halophilus* as it can grow at salt concentrations up to 1600 mM; the optimum salt concentration is 700 mM (Fig. 2. 4e). Taken together, *Nitrosococcus wardiae* D1FHS represents a new species that is characterized by a distinct growth physiology from other *Nitrosococcus* species.

2.4.5 Emended Description of the Genus *Nitrosococcus*

(GenBank *Taxonomy ID:* 1227) synonym: *Nitrosococcus* Winogradsky 1892 (Winogradsky 1892); ex "*Micrococcus nitrosus*" Migula (Migula 1990); ex *"Nitrosococcus nitrosus*" Buchanan (Buchanan 1925 , Editorial_Board 1955, Commission 1958). Name approved by Skerman et al. (Skerman et al., 1980), but the culture had been lost decades before. The etymology of *Nitrosococcus* is catalogued using digital optical identifiers as follows: *Genus* (doi://dx.doi.org/10.1601/nm.2107), Family *Chromatiaceae* (doi://dx.doi.org/10.1601/tx.2070; (Bavendamm 1924)), Order *Chromatiales* (doi://dx.doi.org/10.1601/nm.2069; (Imhoff 2005; List_Editor 2005)), Class *Gammaproteobacteria* (doi://dx.doi.org/10.1601/tx.2068), phylum *Proteobacteria* (doi://dx.doi.org/10.1601/tx.808; (Garrity and Holt 2001)).

2.4.6 Description of *Nitrosococcus wardiae* spec. nov.

ETYMOLOGY: *Nitrosus* (Latin masculine adjective): nitrous; *coccus*: (Latin masculine adjective): sphere; *wardiae* (ward.i'ae. N.L. fem. gen. n. wardiae): of Ward, named after the American microbiologist Bess B. Ward for her work on the marine Nitrogen cycle including the study of ammonia-oxidizing bacteria in the genus *Nitrosococcus*.

LOCALITY AS WELL AS CULTURE HISTORY: Collected on October 26, 2008, at station D1 in Jiaozhou Bay, a marginal sea bay in the Yellow Sea of China (36.067°N, 120.230°E) by Hongyue Dang and Martin G. Klotz. Maintained as enrichment culture D1 in the Klotz Lab (University of Louisville). Identified as a gammaproteobacterial AOB in an enrichment-culture in 2011 by Lin Wang and Martin G. Klotz and later purified using a hydroxylamine treatment regime by Lin Wang, Chee Kent Lim and Martin G. Klotz (at UNC Charlotte).

Appearing as large cocci or very short rods. Cells contain a well-developed intra-cellular membrane system of an arrangement that appears as one stack of membrane vesicles packed mainly in the center of the cell. Polar flagella allow for motility; the genotype is chemotaxis-positive. Light sensitive; Strictly aerobic; Moderately alkaliphilic and mesophilic with an optimum growth temperature at 37 °C; Grows in the presence of sodium salts between 200 to 800 mM with an optimum at 700 mM; Grows at ammonium concentrations below 300 mM with an optimum at 50 mM; The size of the genome is 4,022,640 base pairs with a GC content of 50.7%. No plasmid was identified. The genome of *Nitrosococcus wardiae* strain D1FHS^{T} (GenBank *Taxonomy ID:* 1814290) is presently being annotated and analyzed. MIGS information (Chain et al., 2003) to correlate the genotype with the environment of its isolation is provided in Table 2.4.

2.5 Acknowledgments

Pertinent taxonomic information was accessed through the “Names for Life” online tool (http://namesforlife.com). Nitrogen cycle research in the Klotz lab has been supported during the duration of this project by the NSF (EF-0412129, EF-0541797, MCD1202648) the Gordon and Betty Moore Foundation, the China National Science

Foundation (Grants 91328209 and 41076091), a distinguished fellowship at State Key Laboratory of Marine Environmental Science at Xiamen University, China, and institutional incentive funds from the University of Louisville (Office of the EVP for Research). Lin Wang and Chee Kent Lim were supported by incentive funds from UNC Charlotte (Office of the VC for Research & the College of Liberal Arts & Sciences). Hongyue Dang was supported by China MOST 973 grant 2013CB955700, NSFC grants 91328209, 41076091 and 91428308, and SOA grant GASI-03-01-02-05.

Table 2.1: Genome-sequenced gammaproteobacterial AOB species

	Species/Strain	Date of Isolation	Isolated by	Origin of site
Gamma-AOB	*Nitrosococcus oceani* C-107	1957	S.W. Watson	North Atlantic
	AFC 132	1966	A.F. Carlucci	South Pacific (13°S,76°W)
	AFC27	1964	A.F. Carlucci	North Pacific (43°N,155°W)
	C-27	1964	S.W. Watson	Barbados Harbor
	Nitrosococcus halophilus Nc4	1990	Harms & Koops	Salt lagoon off the coast of Sardinia
	Nitrosococcus watsonii C-113	1966	S.W. Watson	Red Sea

Table 2.2: Environmental factors for Jiaozhou Bay sampling on 2008-10-26

	D1	J1	P5	A5	P3
Time	**10:45**	12:00	12:30	13:40	14:35
Longitude (^{0}E)	**120.13418**	120.09000	120.11576	120.18332	120.20463
Latitude (^{0}N)	**36.04186**	36.09500	36.11724	36.08859	36.09851
Water depth (m)	**8.0**	2.0	3.2	6.3	5.1
Surface seawater Temperature (℃)	**18.5**	16.2	18.5	18	20
Surface seawater Salinity (%)	**3.5**	3.0	3.2	3.3	3.3
Sediment temperature (℃)	**14.8**	16	16.8	16.8	17
pH (surface seawater)	**8.13**	8.09	8.13	8.0	7.95
DO (surface seawater) (mg/L)	**11.31**	10.85	10.96	10.29	12.33

Table 2.3: Primers used in this study

Name	Sequence (5'-3')	Used for	Reference
M13F	GTAAAACGACGGCCAG	Sequencing	-
M13R	CAGGAAACAGCTATGA	Sequencing	-
27F	AGAGTTTGATCMTGGCTCAG	PCR	Lane et al. (1991)
1492R	GGTTACCTTGTTACGACTT	PCR	Turner et al. (1999)
gamma-haoA fwd	YTGYCAYAAYGGRGYNGAYCAYAAY GAGT	PCR	Campbell et al. (2011a)
gamma-haoA rev	TTRTARWGCTKGAKSANRTGMTGYT CCCACAT	PCR	Campbell et al. (2011a)
gamma-16S fwd (F116s)	CAATCTGAGCCATGATCAAAC	PCR	this study
gamma-16S rev (R16s2)	CCTACGGCTACCTTGTTACG	PCR	this study

Table 2.4: Classification and general features of *Nitrosococcus wardiae* D1FHS according to MIGS Recommendations (Chain, et al. 2009)

MIGS ID	Property	Term	Evidence code
	Current classification	Domain *Bacteria* Phylum *Proteobacteria* Class *Gammaproteobacteria* Order *Chromatiales* Family *Chromatiaceae* Genus *Nitrosococcus* Species *wardiae* Strain D1FHS	IDA
	Gram stain	negative	IDA
	Cell shape	Coccus or short Rod	IDA
	Motility	motile	IDA
	Sporulation	non-sporulating	IDA
	Temperature range	mesophilic	IDA
	Optimum temperature	37°C	IDA
	Salinity	700 mM optimum	IDA
MIGS-22	Oxygen requirement	aerobic	IDA
	Carbon source	carbon dioxide (Calvin Benson Basham cycle)	IDA
	Energy source	ammonia	IDA
	Energy metabolism	chemolithotroph	IDA
MIGS-23	Isolation and growth conditions	Isolation after enrichment in ammonium minimal salts medium amended with hydroxylamine	IDA
MIGS-6	Habitat	Marine sediment	IDA
MIGS-15	Biotic relationship	free-living	NAS
MIGS-14	Pathogenicity	non-pathogenic	NAS
	Biosafety level	1	NAS
MIGS-4	Geographic location	Yellow Sea/ Jiaozhou Bay, China	IDA
MIGS-4.1	Latitude	36.067°N	IDA
MIGS-4.2	Longitude	120.230°E	IDA
MIGS-4.3	Depth	Sediment, 8 m depth	IDA
MIGS-4.4	Altitude	Not reported	NAS
MIGS-5	Sample collection time	2008 Oct 26	IDA

Evidence codes – IDA: Inferred from Direct Assay (first time in publication); TAS: Traceable Author Statement (i.e. a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e. not directly observed for living, isolated sample, but

based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from a living isolate cultured by these authors.

Table 2.5: Characteristics of Type Strains of species in the genus *Nitrosococcus*

	Nitrosococcus wardiae D1FHS	*Nitrosococcus halophilus* Nc4	*Nitrosococcus watsonii* C-113	*Nitrosococcus oceani* C-107
Habitat	Marine	Marine	Marine	Marine
16S rRNA gene clusters	2	2	2	2
Plasmids	0	1	2	1
Genome size	4,022,640 bp	4,079,427 bp	3,328,579 bp	3,481,691bp
G+C content	50.7%	51.6 %	50.1%	50.4%
Cell shape and arrangement	Coccus or short Rod Singles & Pairs	Coccus or short Rod Singles & Pairs	Coccus or short Rod Singles & Pairs	Coccus or short Rod Singles & Pairs
Temperature optimum	37 °C	28 - 32 °C	28 - 32 °C	28 - 32 °C
Salinity optimum	700 mM	800 mM	600 mM	500 mM
pH optimum	7.6 - 8.0	7.6 - 8.0	7.6 - 8.0	7.6 - 8.0
Ammonium tolerance	<300 mM	< 600 mM	< 1600 mM	< 1200 mM
Ammonium optimum	50 mM	100 mM	100 - 200 mM	100 mM
Use of urea	-	-	+	+

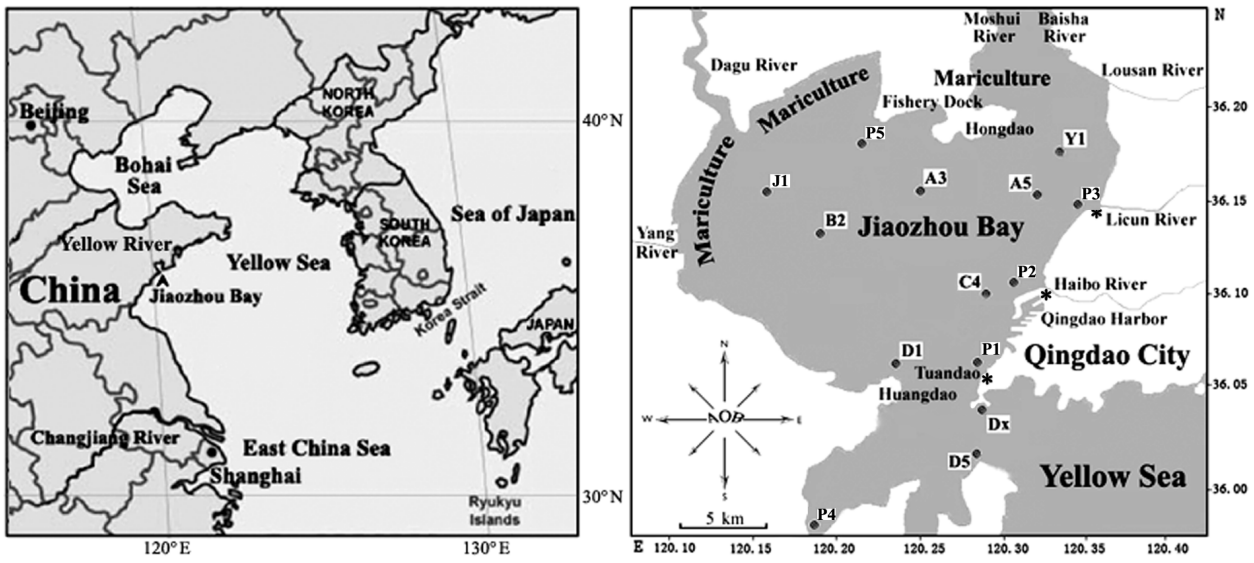


Figure 2.1: Map of Jiaozhou Bay and sampling stations.

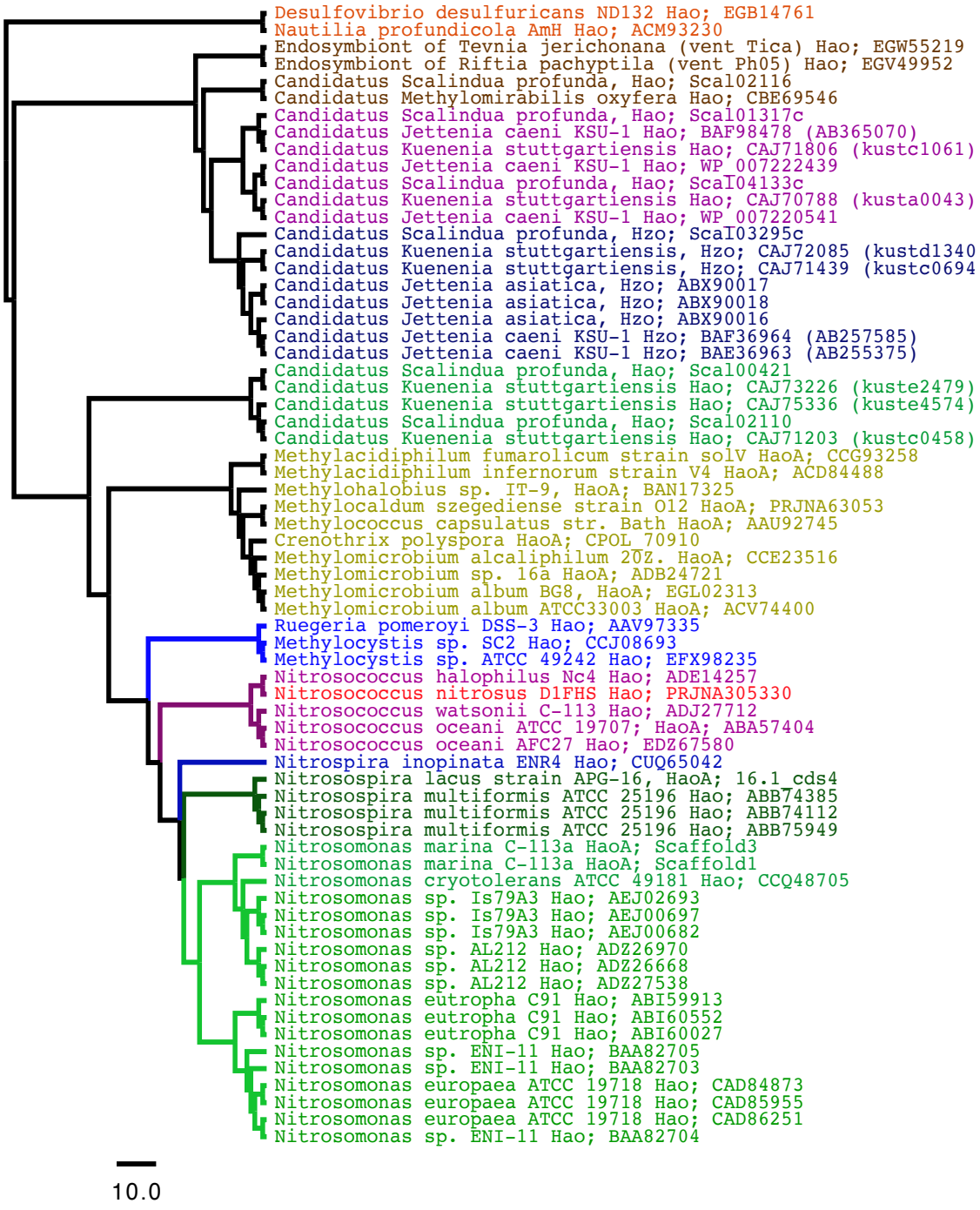


Figure 2.2: Consensus tree constructed after Bayesian inference of phylogeny from the MUSCLE alignments of sequences from octaheme cytochrome c proteins in the "Hao cluster" (Klotz et al., 2008, Kern et al., 2011) using the sequence from "reverse Hao" proteins as the out group (Hanson et al., 2013). The N- and C-terminal sequences outside the heme *c* binding motifs were eliminated from the final alignment that was subjected to Bayesian inference of phylogeny using the BEAST package (see *Experimental procedures*). Posterior probability values of all nodes were ≥0.99. Mean

branch lengths are characterized by a scale bar indicating the evolutionary distance between the proteins (changes per amino acid position). The branches are annotated with labels indicating the source organism and the protein sequence accession number.

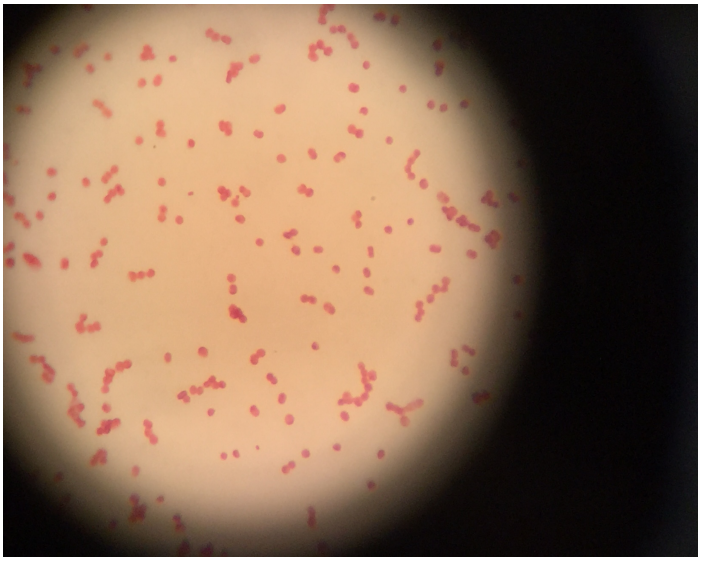
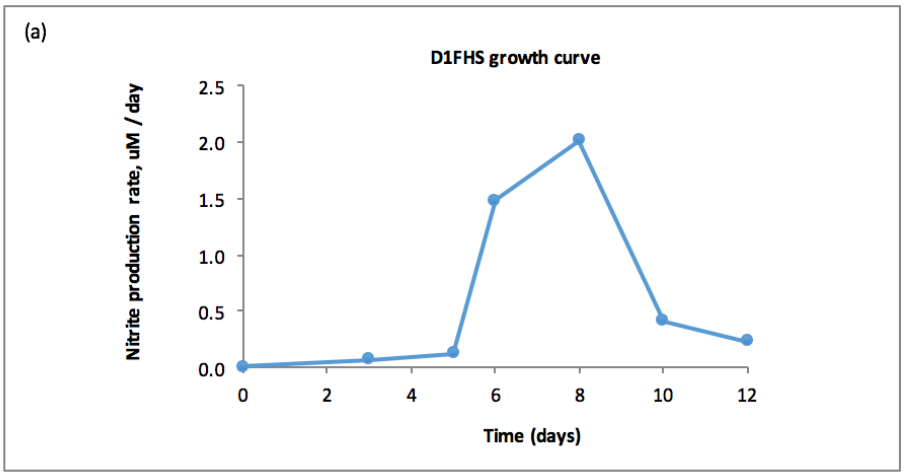
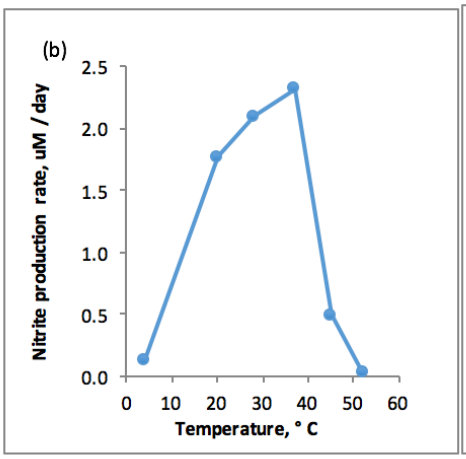
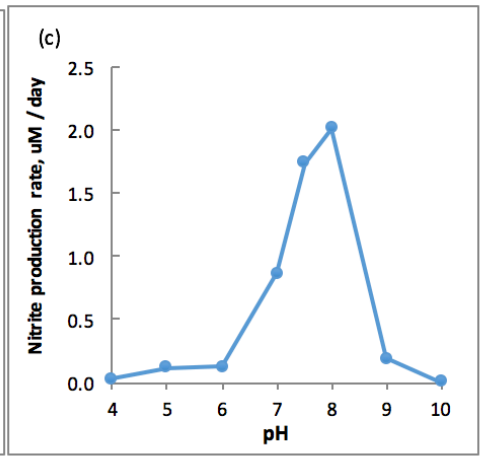
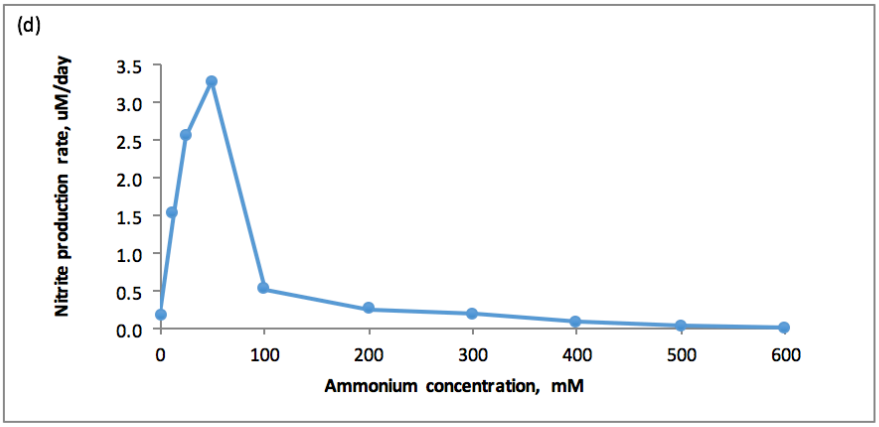


Figure 2.3: Gram Stain, 1000X magnification using a light microscope. Cell shape is short rods, arranged singly or in pairs.


(a)
D1FHS growth curve
Nitrite production rate, uM / day
2.5
2.0
1.5
1.0
0.5
0.0
0
2
4
6
8
10
12
Time (days)


(b)
Nitrite production rate, uM / day
2.5
2.0
1.5
1.0
0.5
0.0
0
10
20
30
40
50
60
Temperature, ° C


(c)
Nitrite production rate, uM / day
2.5
2.0
1.5
1.0
0.5
0.0
4
5
6
7
8
9
10
pH



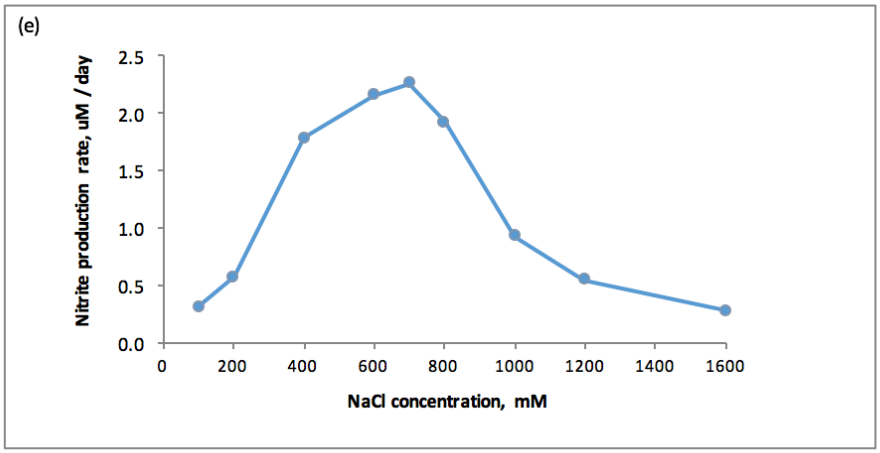


Figure 2.4: Figure 2.4a. Growth of D1FHS (assessed by maximum nitrite production rate) in AMS-H standard medium with 700mM NaCl and 12.5 mM $(NH_4)_2SO_4$ at pH 8, grown at 28° C in the dark. Figure 2.4b, c. Physiological tests: Growth of D1FHS in AMS-H in the dark (assessed by maximum nitrite production rate) was characterized as a function of (B) temperature and (C) pH conditions. Data points are mean values calculated from 3 biological replicates. Figure 2.4d. Physiological tests: Growth of D1FHS in AMS-H in the dark (assessed by maximum nitrite production rate) was characterized as a function of ammonium concentration. Data points are mean values calculated from 3 biological replicates. Figure 2.4e. Physiological tests: Growth of D1FHS in AMS-H in the dark (assessed by maximum nitrite production rate) was characterized as a function of NaCl concentration. Data points are mean values calculated from 3 biological replicates.

CHAPTER 3: THE GENOME SEQUENCE OF *NITROSOCOCCUS WARDIAE* STRAIN D1FHS^T REVEALS THE ARCHETYPAL METABOLIC CAPACITY OF AMMONIA-OXIDIZING GAMMAPROTEOBACTERIA.

3.1 Introduction

Ammonia-oxidizing bacteria (AOB) are obligate chemolithoautotrophic organisms catalyzing the first step of nitrification, the oxidation of ammonia to nitrite. To date, *Nitrosococcus oceani* (Murray and Watson 1962), *Nitrosococcus halophilus* (Koops et al., 1990), *Nitrosococcus watsonii* (Campbell et al., 2011a) and *Nitrosococcus wardiae* (Wang et al. 2016) are the only species of marine gammaproteobacterial AOB with representative type strains in pure culture. The genome sequences of the type strains from the first three species are publically accessible from the NCBI/EMBL and JGI/IMG/M databases; however, only the genome sequence of *N. oceani* strain C-107 (ATCC19707) has been annotated and analyzed (Klotz et al., 2006). Strain D1FHS of a new species classified as "*Nitrosococcus wardiae*" in the genus *Nitrosococcus*, family *Chromatiaceae*, order *Chromatiales*, class *Gammaproteobacteria*, was recently isolated into pure culture by enrichment from a sediment sampled from Jiaozhou Bay (Wang et al., 2016), a semi-closed water body locating on the Southwestern coast of Shandong Peninsula, China (Dang et al., 2010). D1FHS forms Gram-negative short rods, occurring singly or in pairs (Table 1), obtains energy and reductant solely from ammonia oxidation and grows with CO_2 as the carbon source. Likewise as reported for other characterized *Nitrosococcus* species, *N. wardiae* D1FHS was able to grow at pH 5~9 and tolerant to

salt concentrations up to 1600 mM. However, D1FHS exhibited a lower ammonium optimum concentration (50 mM) than other known *Nitrosococcus* species (Koops et al., 1990, Campbell et al., 2011, Wang et al. 2016). Growth of D1FHS was observed at temperatures as high as 45°C with the optimum temperature of 37°C for nitrite production (Wang et al., 2016). Here we report the complete genome sequence of type strain D1FHS and provide insight into the molecular mechanisms leading to its distinct physiology enabling it to grow as an ammonia-dependent chemolithotroph in the marine environment. The *Nitrosococcus wardiae* D1FHS genome sequence was analyzed in context with the genomes of the other type strains in the *Nitrosococcus* genus: *Nitrosococcus oceani* C-107 (=ATCC19797), *Nitrosococcus halophilus* Nc4 and *Nitrosococcus watsonii* C113. Surprisingly, the genome of D1FHS revealed the presence of a cyanase gene (*cyn*), thereby making *Nitrosococcus wardiae* D1FHS the first AOB with the genetic potential to encode cyanate hydratase (EC 4.2.1.104). The ability to grow with cyanate as the sole source of energy and reductant was reported recently for an ammonia-oxidizing Thaumarchaeote, *Nitrososphaera gargensis* (Palatinszky et al., 2015), enriched from a hot spring (Hatzenpichler et al., 2008). In contrast, cyanase is encoded in all genome-sequenced nitrite oxidizers and has been reported to function in many bacteria and archaea (Palatinszky et al., 2015). Cyanase hydrolyzes cyanate produced within cells during carbmoyl phosphate metabolism and urea formation or from the outside environment into carbon dioxide and ammonium, which are then available for assimilation as carbon and nitrogen sources, respectively, to support growth (Luque-Almagro et al., 2008, Kamennaya et al., 2008, Kamennaya and Post 2013). To test whether the cyanase gene identified in the genome of strain D1FHS was functional, we

measured growth and *in vivo* steady-state mRNA levels of the *cyn* gene and other marker genes in the presence of cyanate and ammonium respectively.

3.2 Material and Methods

3.2.1 Genomic DNA Sequencing, Assembly and Annotation

High quality D1FHS pure culture genomic DNA was sequenced using the Pacific Biosciences (PacBio, USA) single molecule sequencing method. Sequence reads generated from PacBio sequencing were assembled using RS_HGAP_Assembly protocol v.3. Completeness of the assembled genome was assessed by BLASTN and TBLASTN searches (CLC Main Workbench, Qiagen, Denmark) using the D1FHS 16S rRNA sequence and a collection of 31 single-copy genes (Wu and Eisen 2008). The assembled genome sequence was submitted to the RAST (Rapid Annotation using Subsystem Technology) annotation server (Aziz et al., 2008) for the automatic prediction and annotation of protein-coding DNA sequences (CDS). The annotation of all CDS in key pathways, including those for ammonia oxidation, respiration and carbon fixation was manually refined by searching against KEGG (*K*yoto *E*ncyclopedia of *G*enes and *G*enomes), and NCBI on an individual gene-by-gene basis as needed. Whole Genome Shotgun (WGS) project (SUB1472135) information has been submitted to NCBI.

3.2.2 Growth of D1FHS on Cyanate

The cyanate hydratase (EC 4.2.1.104) positive strain *N. wardiae* D1FHS and the cyanase-negative strain *N. halophilus* Nc4 used in this experiment were maintained in marine ammonium mineral salts (AMS) medium (700 mM NaCl, 2.5 mM $(NH_4)_2SO_4$, 30 mM $MgSO_4$, 20 mM $MgCl_2$, 10 mM $CaCl_2$, 10 mM KCl, 0.2 mM $NaCO_3$, 3.0 mM $NaHCO_3$, 0.09 mM K_2HPO_4, 3 μM chelated iron, 0.4 μM $NaMoO_4$, 1.0 μM $MnCl_2$, 0.008

μM $CoCl_2$, 0.35 μM $ZnSO_4$, 0.08 μM $CuSO_4$, phenol red 0.5% w/v). Culture pH was adjusted regularly to ~8 with sterile 0.25 M K_2CO_3. Mid-exponential phase *N. wardiae* D1FHS and *N. halophilus* Nc4 cultures were treated with 1mM KOCN and 2.5 mM $(NH_4)_2SO_4$ for two days to induce the expression of the *cyn* gene (Palatinszky et al., 2015). The cyanate-induced cultures were then harvested by centrifugation at 5,000 × g for 15 min (Sorvall Evolution; SS-1500 rotor). The cell pellet was washed twice with ammonium-free MS medium and resuspended. 1% inoculum (20μl) was added to 20ml fresh media in a 50 ml Erlenmeyer flask (Pyrex) with one of the following media recipe: ammonium-free MS medium (nitrate is the only nitrogen source in the media), AMS (containing 2.5 mM $(NH_4)_2SO_4$) and 1mM KOCN (ammonium-free MS medium amended with 1mM KOCN). All incubations were performed in triplicate and three independent trials. Cultures and media were incubated at 30ºC in the dark without shaking for up to 9 days (Fig. 3.3). Nitrite and ammonium concentration were monitored routinely during this period (Hansen and Nielsen 1939, Nicholas and Nason 1957, Koops et al., 1976, Wang et al., 2016).

3.2.3 Gene Expression Induction

For RNA extraction, two independent one-liter mid-exponential phase D1FHS cultures were harvested by centrifugation (5,000 × g, 15 min). The resulting cell pellets were washed twice with ammonium-free MS medium and then resuspended and distributed to 250 ml of ammonium-free MS medium (none treatment), MS medium containing 2.5mM $(NH_4)_2SO_4$ (NH_3-only treatment), MS medium containing 2 mM KOCN (KOCN-only treatment) and MS medium supplemented with 2.5mM $(NH_4)_2SO_4$ and 2 mM KOCN (NH_3+ KOCN treatment). For gene expression induction, the above

four cultures were incubated for 24h (El Sheikh and Klotz 2008) at 30°C, in the dark, without shaking.

3.2.4 RNA Extraction and cDNA Synthesis

Cell pellets collected from the above four differentially treated cultures were resuspended in 0.5 ml corresponding medium and immediately stabilized with RNAprotect Bacteria Reagent (Qiagen, Germany) before total RNA extraction. RNA was then extracted using RNeasy Mini Kit (Qiagen, Germany) following the manufacturer's protocol and dissolved in 40 ul nuclease-free H_2O (provided in the kit). RNA samples were assessed for purify by measuring 260/280 and 230/260 ratios and quantified by absorbance at 260nm using a Nanodrop 200C spectrophotometer (Thermo Fisher Scientific, USA). Afterwards, RNA samples were treated with RQ1 RNase-free DNase (Promega, USA) based on the manufacturer's instruction. The DNase-treated RNA samples were cleaned up using the RNeasy mini kit following the cleanup protocol provided by the manufacturer and reexamined on spectrophotometer for quantity and purity. Appropriate amount of the clean RNA was used in 20 ul reaction to synthesized first-strand cDNA with Superscript III reverse transcriptase (Invitrogen, USA) at an extension temperature of 50°C for 60 min. The resulting products in the tube were diluted with sterile H_2O to 100ul (Poret-Peterson et al., 2008) and quantified using a Nanodrop 2000 spectrophotometer(Thermo Fisher Scientific, USA).

3.2.5 Real-time Quantitative PCR

The synthesized first-strand cDNA samples were used for qPCR to examine the differential gene expression (*amo*A, *cyn*, *hao*A, *hao*B) with the16S rRNA gene as the internal reference gene under the above growth conditions. It was performed in 10-ul

reactions prepared in triplicate for each primer set using iQ SYBR Green supermix (Bio-Rad Laboratories, Hercules, USA) following the manufacture's protocol. Each reaction includes 5 ul iQ SYBR Green supermix, 100 nM forward and reverse primers, ~100 ng cDNA template and dH_2O adding up to the total reaction volume of 10 ul. The reactions were run on ABI 7500 fast real-time PCR system (Thermo Fisher Scientific, USA) with the following parameters settings: initial denaturation at 95 °C for 3 min, 40 cycles consisting of 95 °C for 10 s, 50 °C for 30 s and 72 °C for 15 s; the melt curve was extended over the temperature range of 50-95 °C. Ct values were automatically generated at the specific fluorescence level in log linearization phase. Ct value of these genes in each technical replicate were normalized with that of the pertinent 16S rRNA gene using the 2 (-delta delta C(T)) quantitation method (Livak 2001, Graham et al., 2011). The averages of the normalized technical triplicates of the Ct values for detected *amo*A, *cyn*, *hao*A and *hao*B genes in the cDNAs in two independent trials were used to calculate the significance with the two-way analysis of variance (ANOVA) including Tukey's honest significant difference (HSD) test to compare treated and non-treatment control samples as a reflection of changes in steady-state mRNA levels in response to the treatments (Fig. 3.4).

3.3 Results and Discussion

3.3.1 Genome Properties

The *N. wardiae* D1FHS genome consists of a single circular chromosome (4,022,640 bp; the G+C content of 50.7%) and no plasmid has been identified. The genome contains 4,128 CDS (protein-coding DNA sequences), 45 tRNA genes and two 16S-23S-5S rRNA gene clusters (Table 3.1). The *rrn* operons are located on different

replicons, the 16S, 23S and 5S rRNA genes in the two clusters are 99% identical. Genetic markers known from other *Nitrosococcus* genomes were also identified in the D1FHS genome (see text below).

3.3.2 Chemolithotrophy

a) AMO-HAO

In AOB, ammonia monooxygenase (AMO) and hydroxylamine dehydrogenase (HAO) perform oxidation reactions that lead to the extraction of electrons that need to be relayed to the ubiquinone pool (Whittaker et al., 2000, Arp et al., 2002, Hooper et al., 2005) to support chemolithotrophic growth. The membrane-bound copper-containing enzyme, AMO, is responsible for the initiation of ammonia oxidation and releases hydroxylamine in the periplasm while consuming two electrons (Whittaker et al., 2000). Analysis of the genome of *N. wardiae* D1FSH revealed a cluster of three contiguous genes, *amo*CAB, encoding the subunits of AMO (186,919...187,695; 185,823...186,566; 184,527...185,759). An *amo*D gene (*orf*5; 561bp; 183,767 ... 184,327, 90% identical to NAV4) (Norton et al., 2002, El Sheikh et al., 2008) was found downstream of *amo*B, with a terminator in between. The *orf*M gene identified in all other AOB genomes was not detected in the D1FSH genome; however, additional non-clustered *amo* genes were absent as reported for all other Gamma-AOB and *amo* pseudogenes were not detected, which is consistent with genome analyses of genomes of Beta- and Gamma-AOB (Klotz et al., 2006, Arp et al., 2007, Norton et al., 2008, Bollmann et al., 2013, Urakawa et al., 2015). Consistent with genome content in *N. halophilus*, a *pan*I-type multicopper oxidase (Lawton et al., 2009) was found upstream of the *amo* operon as part of a gene cluster that includes the *nirK* gene. This clustered arrangement of the *panI* and *nirK* genes in the *N.*

wardiae/N. halophilus ecogenotype is different from the orphan arrangement of both genes in the genomes of the *N. oceani*/*N. watsonii* ecogenotype and resembles the arrangement of these genes in genomes of the betaproteobacterial AOB *N. europaea* and *N. eutropha* (Arp et al., 2007, Stein et al., 2007).

The periplasmic protein complex, HAO (HAO = $HaoA_3$, encoded by *hao*A; 3,774,474...3,776,216) consisting of three HaoA subunits cross-linked with a tyrosine protein ligand, oxidizes hydroxylamine to nitrite and releases electrons (Arciero and Hooper 1997) and thus constitutes the key enzyme in bacterial ammonia catabolism (Klotz and Stein 2008, Simon and Klotz 2013, Stein and Klotz 2016). The *hao*A-associated gene *hao*B (3,773,418... 3,774,467) was found downstream adjacent to *hao*A in the same operon as it has been experimentally verified for other nitrifying bacteria (Poret-Peterson et al., 2008, Stein et al., 2013). To support growth via chemolithotrophic catabolism, electrons extracted during hydroxylamine oxidation to nitrite must be directed to the ubiquinone pool. Cytochrome c_M552 (*cyc*B, 3,772,665...3,773,387) has been shown to function as a quinone reductase (Elmore et al., 2007, Kim et al., 2008) but it has not been experimentally established how it receives the catabolic electrons: c_M552 could either accept electrons directly from HAO or through mediation by cytochrome *c*554 (*cyc*A, 3,771,986...3,772,645), which has been experimentally verified to re-oxidize HAO by serving as a dedicated electron acceptor (Hooper et al., 2005, Klotz et al., 2008, Kern et al., 2011, Simon and Klotz 2013). There is experimental evidence for redox interaction between homologues of c_M552 and HAO in reverse direction in *Epsilonmicrobia* (Hanson et al., 2013). Genes *hao*AB and *cyc*AB reside in one gene cluster in the genome of D1FHS; prior work identified that they comprise one

transcriptional unit in other Gamma-AOB (Stein et al., 2013). However, *amo*, *hao*AB and *cyc*AB are not exclusively encoded by AOB genomes (Arp et al., 2007), the only recognized gene identified as unique to AOB is *ncy*A, which has been exclusively identified in chemolithotrophic AOB (Campbell et al., 2011b, Klotz and Stein 2011, Stein et al., 2013). Gene *ncy*A encoding nitrosocyanin (3107183...3107608), a novel soluble red copper protein has been detected in equimolar quantities with HAO in the periplasm of AOB (Arciero et al., 2002b, Hooper et al., 2005). Since the protein sequence and structure provide no clues about a potential catalytic function, a possible role in directing the path of electrons extracted from HAO was proposed, and the functional co-residence of *ncyA* with the *amoCAB*, *haoAB*, and *cycAB* genes has been proposed to constitute the molecular underpinning of ammonia-dependent chemolithotrophy (Stein et al., 2013).

The genome of *N. wardiae* contains complete sets of gene inventories that support aerobic growing utilizing O_2 as the final electron acceptor, reduce NAD(P), establish a proton gradient used for the synthesis of ATP. The encoded electron transport chain consists of NADH-ubiquinone oxidoreductase (NUO) (a.k.a., complex I), cytochromes *bc*1 (a.k.a., complex III), and a Cu-aa3-type cytochrome c oxidase (a.k.a., complex IV). The proton motive force generated during the electron transport is utilized for ATP production by a F_0F_1-type ATP synthase (a.k.a., complex V). This electron transport pathway is coupled with a complete tricarboxylic acid cycle and gluconeogenesis pathways (see text below; Fig. 3.5).

Nitrosococcus wardiae encodes two type-1 NUOs, which are able to convert the electron flow across ubiquinone into proton translocation, one facilitates electron transfer from NADH to ubiquinone pool and generates NAD^+, while the other one is involved in a

reverse electron flow generating NADH. In addition, *N. wardiae* contains a sodium-NADH-ubiquinone oxidoreductase (Na^+-Nqr) encoded by *nqr*ABCDEF (Kerscher et al., 2008). Instead of pumping protons, it will translocate sodium ions during the oxidoreduction of ubiquinone. Na^+-Nqr has been reported to be involved in energy conservation during nitrogen fixation by *Rhodobacter capsulatus* (Schmehl et al., 1993, Kumagai et al., 1997).

The ubiquinone-cytochrome c reductase, cytochrome *bc*1 (complex III; 980203...982780), can accept or donate electrons from/to the quinone pool (functions like a so called Quinone-Reactive Protein complex) or exchange electrons directly with periplasmic cytochromes such as c_{552} or P460. This makes the redox function of cytochrome *bc*1 the gateway to branched electron transfer processes, its function is associated with proton pumping as is that of the oxygen-reducing heme-copper oxidase (complex IV), which in branched electron flow functions as a terminal cytochrome oxidase (Klotz et al., 2006, Simon and Klotz 2013). Two complete sets of genes encoding a Cu-aa_3-type complex IV were identified in the genome of *N. wardiae*.

In addition to the typical H^+-translocating F_0F_1 type ATP synthase, a bacterial Na^+-dependent V-type ATP synthase (subunits ABCDEKI) is also present in the genome of *N. wardiae*. *N. wardiae, N. halophilus* and *N. watsonii* all contain a subunit K but not subunit H in comparing with *N. oceani* (Klotz et al., 2006) . The presence of the Na^+-dependent V-type ATPase in addition to the NA^+-Nqr, Na^+/H^+ antiporters and Na^+/H^+ transporters indicates the possibility of a sodium-motive circuit independent of and in addition to the existing proton-motive circuit in *N. wardiae* similar to other known *Nitrosococcus* species, which enables AOB with intracellular membrane systems to

partition energy conservation as an adaption to life in a high-salt marine environment (Klotz et al., 2006, Arp et al., 2007).

b) Cyanase

D1FHS is the first AOB with identified genes encoding cyanate hydratase (EC 4.2.1.104) and a pertinent transporter for cyanate. The protein sequence of the D1FHS cyanase is 85% similar to the cyanase identified in *Nitrospina gracilis* (Palatinszky et al., 2015). While the *N. wardiae* genome does not encode a dedicated cyanate transporter functional in Cyanase-positive nitrite-oxidizing bacteria, a gene encoding a functionally similar transporter in the formate/nitrite exchanger family (FNT; 1257442...1258383) was identified. The hypothesized cyanate hydratase gene is 276 bp long encoding a protein with 91 amino acid residues.

Strains D1FHS and Nc4 both grew well on AMS media. As measured by nitrite production, both strains entered the exponential growth phase at around day 5, but neither strain was able to grow on ammonium-free MS medium with only minimum levels of nitrite detected in the spent medium (Fig. 3.3). The addition of 1 mM KOCN did not have an apparent inhibitory effect on growth or maximum nitrite production of strain D1FHS during its lag phase (day 0 to day 3 (Fig.3.3. A). However, after D1FHS entered exponential phase, the rate of nitrite production decreased dramatically and no measurable nitrite production was observed beyond day 7 (Fig.3.3. A). In contrast, growth of strain Nc4 was significantly inhibited in 1mM KOCN-amended MS medium starting from the initial inoculation (Fig.3.3. B).

The transcript abundance of all four tested genes, *amo*A, *cyn*, *hao*A and *hao*B, in ammonium-treated samples were significantly higher than in the other samples. In

comparison to nitrogen-free treatment, the transcript levels were about 4.2-fold (*amo*A), 19.1-fold (*cyn*), 12.3-fold (*hao*A) and 15.9-fold higher (*hao*B) (Fig. 3.4). Although, the expression levels of *hao*A steady-state mRNA in cyanate-only and in ammonium plus cyanate treatments were significantly elevated compared to the sample without treatment, there was no significant difference between these two treatments (p value > 0.05) indicating that the induction of transcription was potentially due to ammonium, which forms spontaneously in smallest quantities from cyanate degradation. It has been experimentally established that very small concentrations of ammonium have dramatic effects on transcription in AOB (El Sheikh and Klotz, 2008).

c) Hydrogenase

Both genomes of the *N. wardiae* and *N. halophilus* strains encode a [NiFe]-hydrogenase, which like in betaproteobacterial AOB such as *Nitrosospira multiformis* ATCC 25196 (Norton et al., 2008) or *Nitrosospira lacus* APG3 (Garcia et al., 2013), might contribute to catabolic activities as in the recently described nitrite-oxidizing bacterium *Nitrospira moscoviensis* (Koch et al., 2014). The [NiFe]-hydrogenase encoded in the genomes of *N. wardiae* and *N. halophilus* belong to group 3a encoded by a gene cluster *hyp*ABCDE (Vignais et al., 2001, Vignais and Billoud 2007). *Hyp* D is highly conserved and can be used as a marker to differentiate hydrogenase groups (Watanabe et al., 2012). HypB is a GTP/GDP-binding protein and binds Ni atoms (Barz et al., 2010). [NiFe]-Hydrogenase in group 3a is a F420-reducing trimeric enzyme and the tentative function is to uptake H_2 during methanogenesis (Vignais et al., 2001, Vignais and Billoud 2007). This indicates that D1FHS may be able to derive energy from external H_2 in addition to ammonia.

3.3.3 Nitrogen Uptake and Metabolism

a) Nitrogen Uptake and Assimilation

Both *N. wardiae* and *N. halophilus* have ammonium/ammonia transporters for ammonium uptake. The genome of D1FHS encodes a "Rhl-type" ammonium transporter sharing 72% protein sequence similarity with the ammonium transporter encoded in the genome of *Nitrosospira multiformis* ATCC 25196 (Norton, et al. 2008). In addition, the D1FHS genome also encodes a "marine_trans_1 type" ammonium/ammonia transporter. In contrast, the genomes of *N. oceani* and *N. watsonii* lack the genes encoding either type of ammonium transporter. While genomes of *N. oceani* and *N. watsonii* encode both urea transporter and urea hydrolase, these are absent from *N. wardiae* and *N. halophilus* thereby creating two eco-genotypes that uptake ammonium directly (*N. wardiae* and *N. halophilus*) or acquire ammonium intracellularly by hydrolysis of imported urea (*N. oceani* and *N. watsonii*). Neither *Nitrosococcus* genome encodes an ABC transporter dedicated to urea uptake,

Both *N. wardiae* and *N. halophilus* have the complete capacity for low and high affinity ammonium assimilation (Fig. 3.6. B). When the external ammonium concentration is high, assimilation of ammonium will occur primarily via the low-affinity system: NAD-specific glutamate dehydrogenase (EC1.4.1.2; 1279663...1284441; 2368260...2369552). In contrast, the GS-GOGAT system (glutamine synthase-glutamine oxoglutarate aminotransferase/glutamate synthase) is known for high-affinity ATP-dependent ammonium assimilation when the concentration of ammonium is low. A type I Glutamine synthetase (GSI, EC 6.3.1.2) is encoded by the gene *gln*A (289859...291268). Five genes encode the glutamate synthase (GltS; EC 1.4.1.13) including three large

chains (1648946...1650955, 2370979...2371467,2973313...2977743) and two small chains (2977743...2979185, 3487093...3490626). GSI activity is possibly regulated by acetylating under the function of the adenylyltransferase encoded by gene *gln*E (EC 2.7.7.42). Gene *gln*B encoding the regulatory protein P_{II} and gene *gln*D (1945019...1947706) encoding a P_{II} uridylyl-transferase (EC 2.7.7.59) were also identified in the genome of *N. wardiae* (Fig. 3. 6.B).

b) Amino Acid Metabolism

The *N. wardiae* genome encodes the biosynthesis of all 20 amino acids required for the synthesis of proteins. Unless otherwise indicated, all amino acids mentioned were the L form. 20 aminoacyl-tRNA synthetases (AARS) were identified in the genome of *N. wardiae*, including one form of lysyl-tRNA synthetase class II (EC 6.1.1.6) (1933061...1934563) which is the same with *N. halophilus*, but *N. oceani* and *N. watsonii* have two forms (lysinyl-tRNA ligase (EC:6.1.1.6); K04566 lysyl-tRNA synthetase, class I [EC:6.1.1.6] and lysyl-tRNA synthetase (EC:6.1.1.6); K04568 lysyl-tRNA synthetase, class II [EC:6.1.1.6]) and one form of GlxRS (EC 6.1.1.17 , 974473...975912) while *N. halophilus*, *N. oceani* and *N. watsonii* have two distinct forms. Neither genome of *N. wardiae, N. halophilus*, *N. oceani* and *N. watsonii* encodes AsnRS (EC 6.1.1.22) and GlnRS (EC 6.1.1.18), a situation typical for AOB that has been discussed in the literature (Norton et al., 2008). Five genes were identified encoding for the asparagine synthetase (EC 6.3.5.4).

c) Urea and Polyamine Cycling

The genomes of *N. wardiae* and *N. halophilus* lack the genes that encode either known enzyme for urea hydrolysis and they are thus non-urea-lytic, which fits the fact

that they lack genes encoding dedicated urea transporters (Fig. 3.6.B). In contrast, all strains of *N. oceani* and *N. watsonii* are ureolytic and urea uptake positive (Koper et al., 2004, Klotz et al., 2006, Campbell et al, 2011b), which likely reflects an adaptation to the increasingly urea-generating biogenic activities in the Holocene (see chapter 4). The capacity to hydrolyze urea was also considered an evolutionary niche adaption for AOB in acid soils (Norton et. al, 2008). Nevertheless, the D1FHS genome encodes a complete urea cycle including: carbamoyl phosphate synthetase (EC:6.3.5.5; large chain 607815...611045; small chain 611126...612187), ornithine carbamoyl transferase (EC:2.1.3.3; 1786624...1787538), argininosuccinate synthase (EC:6.3.4.5; 2323010...2324227), argininosuccinate lyase (EC:4.3.2.1; 870667...872067) and arginase (EC:3.5.3.1; 3411209...3412126); whereby the arginase gene is clustered with the gene encoding ornithine aminotransferase (EC:2.6.1.13; 3410014...3411252) (Fig. 3.1). This complete set of urea cycle genes are also detected in all other sequenced *Nitrosococcus* genomes with the exception of the *N. halophilus* genome, which lacks the arginase gene, thereby rendering this strain incapable of urea production. The arginase gene was likely lost from the *N. halophilus* genome and it can thus be postulated that *Nitrosococcus* was ancestrally urea-metabolic. The identified non-specific "urea transporter" (2938178...2939398) in the *N. wardiae* genome is a relative of the binding protein that corresponds to the active transport system for short-chain amides and urea (FmdDEF) identified in several bacteria and archaea and that often cooperates with an outer membrane protein, FmdC, which is found also in some Beta-AOB such as *Nitrosospira multiformis* and *N. lacus* (Koper et al. 2004, Urakawa et al., 2016). Because the genome

of *N. wardiae* lacks the *fmdCDEF* genes, it can be concluded that *N. wardiae* does not have a urea transport system.

d) Denitrification

N. wardiae is capable of reducing nitrite through classical denitrification with dissimilatory nitrite reductase (NirK; EC 1.7.2.1; 119391...1120842) and nitric oxide reductase (NorCBQD; EC 1.7.99.7). Although the D1FHS genome contains a gene homologous to *nos*F encoding a maturation protein in nitrous oxide reductase-positive microbes, it lacks the genetic capacity to produce a functional nitrous oxide reductase (NosZ) complex. However, a gene encoding a flavo-hemoglobin nitric oxide dioxygenase protein (EC 1.14.12.17; 1276322...1277530) was identified, which is not encoded in the genomes of the other three *Nitrosococcus* species. A functional NO dioxygenase would enable the bacterium to oxidize nitric oxide to nitrate.

A nitrite transporter classified in the formate/nitrite family is also present. The genome contains also several clusters of genes that encode a variety of MCO, which could play a role in NOx transformations such as the *pan*I-type MCO, an alternative NOx reductase (Lawton et al., 2009).

Cytochrome P460 (*cyt*L, 3778385…3778981) can co-oxidize nitric oxide and hydroxylamine oxidation to nitrite whereas the gene *cyc*S encoding a beta sheet-structured cytochrome *c*' (*c'*-beta, CytS) has been implicated in microaerobic NO sequestration by oxidization dehydrogenation (Bergmann and Hooper 2003a, Elmore et al., 2007). Both enzymes, c'-beta and P460, are also likely redox partners of the soluble periplasmic cytochrome c552, which also serve electrons to several periplasmic di-heme cytochrome c peroxidases (Elmore et al., 2007, Simon and Klotz 2013).

3.3.4 Autotrophy

a) CO_2 Uptake and Fixation

Nitrosococcus wardiae is an autotroph utilizing the Calvin-Benson-Bassham cycle for carbon fixation. The rate-limiting step in the Calvin-Benson-Bassham cycle is the carboxylation reaction, which is catalyzed by Ribulose bisphosphonate carboxylase/ oxygenase (RuBisCO), which consists of a large and a small subunit (Cavanaugh and Robinson 1996). The small subunit is encoded by gene *cbb*S (1418695 ... 1419081). The gene *cbb*L encoding the large subunit was initially not assembled but later discovered in one of the smaller contigs that escaped assembly. While the large subunit contains the active site, the small subunit is more in involved in specialized functions (Spreitzer 2003). Two RubisCO operon transcriptional regulators CbbR (1419897... 1420808) and (2624690 ... 2625589) identified found in the genome assembly.

Like ammonia monooxygenase, the substrate of RubisCO must be non-protonated; therefore, cells need to actively convert bicarbonate to carbon dioxide, which is accomplished by carbonic anhydrase (EC 4.2.1.1). The D1FHS genome contains 5 genes encoding for carbonic anhydrase (930845 ... 931600) (1626287 ... 1627042) (1822123 ... 1822722). As expected, the genes encoding all enzymes requitred for a complete Calvin-Benson-Bassham cycle were present: Transketolase (1411196 ... 1413193; 3196500 ... 3199445), NAD-dependent glyeraldehyde-3-phosphate dehydrogenase (1413293 ... 1414306), Phosphoglycerate kinase (1414447 ... 1415625), Pyruvate kinase (1415704 ...1417155), and Fructose-1,6-biphosphate aldolase (1417316 ... 1418380) are encoded by one operon. Likewise, Ribulose-5-phosphate 3-epimerase (176690 ... 177361) and phosphoglycolate phosphatase (177592 ... 178269) form one

transcriptional unit. In contrast, Malate dehydrogenase (1232757...1233734), Fructose-1,6-biphosphatase, type I (829101...830108), Ribose 5-phosphate isomerase (310017 ... 310682) and Phosphoribulokinase (1396913 ... 1397785) are encoded by non-clustered genes in the genomes of all four *Nitrosococcus* species.

b) Central Carbon Metabolism

The genome of *N. wardiae* does not encode a formaldehyde assimilation pathway known to operate in methanotrophic bacteria because it lacks the ability to synthesize hexulose 6-phosphate synthase, hexulose 6-phosphate isomerase (needed fro the Ribulose monophosphate Pathway (RuMP)) or malyl-coenzyme A synthetase and Malyl-coenzyme A lyase (needed for the serine pathway). However, like the genome of *N. halophilus* Nc4, the D1FHS genome encodes a complete oxidation pipeline of methane to CO_2 (Ward et al., 2004b) that is based on a XoxF-type methanol dehydrogenase (Wu et al., 2015) and the methyl-tetrahydromethanopterin (THMPT) pathway (Chistoserdova et al., 2003, 2016). This key inventory is not encoded in genomes of the *N. oceani* (Poret-Peterson et al., 2008) and *N. watsonii* (Campbell et al., 2011a) (Table 3.3).

Further analysis of the *N. wardiae* genome revealed inventories for complete pathways for glycolysis and gluconeogenesis (Table 3.4). In addition, all genes necessary for complete pentose phosphate pathway and TCA cycle were identified. The alpha-ketoglutatrate dehydrogenase in the TCA cycle of D1FSH consists of three components: E1 component (EC 1.2.4.2; 747944...750775), E2 component (EC 2.3.1.61; 746633...747928) and E3 component (1.8.1.4; 2669773...2671182), a rare situation that is found as well in *N. oceani* (Klotz et al., 2006). Genes encoding for dihydrolipoamide

succinyltransferase (EC:2.3.1.61; 746633...747928), dihydrolipoamide dehydrogenase (EC:1.8.1.4; 745166...746611) and pyruvate dehydrogenase (EC 1.2.4.1; 2318250...2319278; 2319271 ... 2320251) were also identified (Fig. 3.5).

3.3.5 Sulfur Metabolism

Like *N. oceani, N. wardiae* has the capacity of acquisition and metabolism sulfur. The reduction of sulfate to sulfide pathway under the function of enzymes sulfate adenylyltransferase (cysND; EC 2.7.7.4), adenylylsulfate kinase (cysC; EC 2.7.1.25), and phosphoadenosine phosphosulfate reductase (cysH; EC:1.8.4.8) were identified in the genome of *N. wardiae* (Table 3.2). Produced sulfide may be further reduced to H_2S, which is the most reduced form of sulfur required for assimilation via biosynthesis of cysteine facilitated by an NADPH-dependent sulfite reductase (CysI; EC 1.8.1.2; Table 3.2). Similar to *N. oceani*, the gene cluster that encodes a putative polysulfide reductase (psr), a monoheme cytochrome (cccA), a cytochrome *c* oxidase, a transporter and the five genes encoding the ubiquinone complex are also present in the genome of *N. wardiae*. Because of the absence of a molybdopterin guanine dinucleotide-binding protein subunit-encoding gene in the vicinity of the cluster, the functionality of the putative polysulfide reductase is awaiting experimental confirmation. Likely as in *N. oceani*, the genetic basis for the formation of internal or external granules of sulfur compounds were also absent from the D1FHS genome (Klotz et al., 2006).

3.3.6 Use of Organic Substrates

Based on the genome sequence, *N. wardiae* is capable of utilizing sucrose, glycogen and polyphosphate as energy reserves. The genome of *N. wardiae* contains genes encoding sucrose synthase (SuSy; EC 2.4.1.13; 865717 … 868101) and sucrose

phosphate synthase (SPS; 863558 … 865717). Fructose-6-phosphate produced by fructokinase (862644…863558) is a required substrate of sucrose phosphate synthase. Sucrose likely serves *Nitrosococcus* cells as an osmotic stress-mitigating compound (Klotz et al., 2006). Two genes each encoding Glucose-1-phosphate adenylyltransferase (3347133… 3348401; 3759251 ... 3760.522) and Glucose-1-phosphate thymidylyltransferase (1746366 … 1747244; 1785878 ... 1786549) as well as the gene encoding UTP glucose-1-phoashate uridylytransferase (1925003...1925890; Noc_2280) have been identified in the genome of *N. wardiae*. The *N. wardiae* genome also revealed the presence of genes encoding the enzymes necessary for the formation of glycogen from fructose-6-phosphate including: two contiguous genes encoding glycogen branching enzyme (EC 2.4.1.18; 751862 … 752179; 752258 ... 753496) of the GH-57-type, which is found also in *N. halophilus* and *N. watsonii* but not in *N. oceani* (GH-13-type), as well as ADP-glucose pyrophosphorylase, phosphoglucomutase (EC 5.4.2.2; 3078235...3079869), Glycogen synthase and Glycogen phosphorylase (EC 2.4.1.1; 2311742 … 2313967; 2513323 ... 2515017; 2820810…2821025; 2821027...2821173) (Fig.3.5). However, genes encoding for glycoside hydrolase were not found. A gene encoding for polyphosphate kinase (EC 2.7.4.1; 1813436 ... 1815508) was identified (Fig. 3.5), indicating that *N. wardiae* can use polyphosphate as a source of energy.

3.3.7 Cellular Growth and Motility

The genome of *N. wardiae* contains almost all genes that have been identified as participation in the cell cycle and in cell division of other Gammaproteobacteria. The genes encoding for cell division protein FtsH (2196176...2198023; 2926706...2927065; 3191460…3193385; 604457...605971) were identified. The genes encoding FtsEX, FstN,

and SulA were not identified, which agrees with the analysis of the genomes of the other three *Nitrosococcus* species. Two large clusters of genes encoding flagellation and motility were found in the genome. The master switch operon *flh*CD was not identified in any of the genomes of the four *Nitrosococcus* species, which is however present in the genomes of other AOB such as *N. multiformis*, indicating that this is a trait unrelated of the catabolic lifestyle of the Gamma-AOB. Similar to *N. oceani*, only one methyl-accepting chemotaxis protein (MCP, 3395919...3397562) was identified. Likely as an adaptation to the marine lifestyle, all four *Nitrosococcus* species contain genes encoding a sodium-driven polar flagellar motor protein MotA (507407…508147 and 2146791...2147561) in their genome in addition to the typical pmf-dependent flagella rotation mechanism.

3.3.8 Communication and Interaction with the Environment.

a) Sensory and Response Regulator Systems

The genome of *N. wardiae* harbors a large number of signal transduction and sensory response systems, including two-component systems, sensors and signal transduction proteins (Table 3.5). The two-component systems, including transcriptional regulators, hybrid sensor and regulators, histidine kinases, and response regulators, are utilized by bacteria to transduce environmental signals and regulate cellular functions. Interestingly, a gene *kai*C (1855401...1856804; 2020137...2021666) encoding the circadian central clock protein KaiC was encoded in the genome of *N. wardiae* but not in any of the other *Nitrosococcus* species. Organisms possessing the circadian clock protein exhibit an accurate time constant and temperature compensation, which allows the system to run at a steady rate independent of temperature fluctuations (Johnson et al., 1996). At

this point it is not clear whether the *kaiC* gene was lost from the other *Nitrosococcus* genomes or whether it was acquired by *N. wardiae*.

b) Stress Tolerance

The *N. wardiae* genome contains inventory (Table 3.5) that contributes to osmotic stress, oxidative stress, periplasmic stress, carbon starvation, cold shock and heat shock tolerance. It contains genes encoding an iron-containing superoxide dismutase (Fe-SOD; EC 1.15.1.1; 1789139...1789720; 3204604…3205191) and catalase (EC 1.11.1.6; 3918985…3921072) with a bacterioferritin (2637334…2637798) for supplying iron and heme (Ma et al., 1999). Like *N. oceani*, the genome encodes for an incomplete glutathione system. It harbors glutathione synthetase (GSS; EC 6.3.2.3; 1135819...1136781), glutaredoxin 2 & 3 (1326192...1326362; 812706...812963), thioredoxin (656288…657151; 621096...621422), cytochrome c_{551} peroxidase (EC 1.11.1.5; 51558...52637) and thiol peroxidase, but it lacks genes for bacterioferritin-comigratory protein, NADH-peroxiredoxin reductase (AhpE), glutathione oxidoreductase, and other isozymes of hydroperoxidase. The gene encoding the redox autoregulatory OxyR protein was absent as from other *Nitrosococcus* genomes. A large number of genes encoding RNA polymerase sigma factors known to regulate stress tolerance were identified in the genome in addition to the minimal growth sigma factor RpoD (sigma 70, 801655...803463; 2431385…2433121). For example, RpoS (sigma 38, 2136771…2137703; 3398340…3399272) is known to regulate the expression of stationary phase inventory including hydroperoxidase and the cell division protein BolA (1815538…1815819). RNA polymerase sigma factors RpoE (sigma 24, 149716...150306; 1473237...1473857) and RpoH (sigma 32, 1727917…1728780) are

involved in heat stress. RpoN (sigma 54, 1241474...1242922) is responsive to nitrogen starvation and the flagella assembly protein FliH (1849342...1850076) is needed for motility. Genes encoding ferric uptake regulation protein (Fur, 3952605...3953066) and Zinc uptake regulation protein (Zur, 1793301...1793792) were also found in the genome of *N. wardiae*. Like all other three sequenced *Nitrosococcus* type strain genomes, a gene encoding peptidyl-tRNA hydrolase (EC 3.1.1.29) regulating dormancy and sporulation-associated proteins with broader functions, was also identified.

c) Transport and Protein Secretion

The genome of *N. wardiae* harbors at least 254 CDs (Table 3.6) involved in active transport, passive transport (diffusion and facilitated diffusion) and protein secretion, including ATP-binding cassette (Andersson et al., 1991)-type transporters, phosphoenolpyruvate-dependent phosphotransferase (PTS) systems, channels/porins, and siderophores (Table 3.6).

Iron transport is essential to *N. wardiae* because it uses cytochromes c such as HAO and other heme-binding enzymes in ammonia oxidation and electron transport. At least 25 genes involved in iron uptake were identified in the genome of *N. wardiae.* No homologues of FecI or FecR were identified, but the encoded ferric uptake regulator (Fur, 3952605...3953066) likely regulates the expression of iron siderophores. Sixteen TonB-dependent iron siderophore receptors were identified.

Transport systems for metal cations other than iron were also identified in the genome of *N. wardiae*. Genes encoding ABC transporters for Zn^{2+}/ Mg^{2+} transport, ion channels for Mg^{2+}/Co^{2+}/Ni^{2+}transport and heavy-metal cation transport were found in the genome. Genes encoding a Zinc ABC transporter and a Zinc transporter in the ZIP family

were also identified. Efflux pumps were found for copper (CopCD), multidrug (SMR), heme, dipeptides, polar amino acids, antimicrobial peptide, bacteriocin/lantibiotic, polysaccharide, and tungsten. Genes encoding four facilitator-type cation efflux proteins ($Co^{2+}/Zn^{2+}/Cd^{2+}$) and Mg^{2+}/Co^{2+} efflux protein CorC were also present. Likelwise, the D1FHS genome encodes a P-type cation-transporting ATPase, KefB-type K^+ efflux system and a Trk-type K^+ uptake system. A major facilitator superfamily (MFS)-type cyanate transporter was found in the genome of *N. watsonii* and *N. oceani*, but not in *N. wardiae* and *N. halophilus*. As an adaptation to marine lifestyle, a variety of Na^+/H^+ antiporter systems were found, including four NhaA types, one NhaD type, one NhaP type, one CPA1 family and five other types.

A formate/nitrite (FNT family) transporter was identified for inorganic N transport in *N. wardiae* and *N. halophilus* but not found in *N. watsonii* and *N. oceani*. *N. wardiae* contains both ABC-transporter and phosphate-selective porins for phosphate transport (Fig. 3.5). Sulfate may be transported through a sulfate permease (SulP family). A chloride channel protein EriC was present for chloride transport.

In the genome of *N. wardiae*, 28 genes were related to type II secretion systems as a general protein secretion system and 21 genes encoding the type IV secretion system for conjugal DNA-protein transfer. Genes encoding the Sec-dependent pathway, a preprotein translocase, two clusters of Sec-independent TatABC systems were also identified.

3.4 Conclusions

Analysis of genome content of *N. wardiae* strain D1FSH confirmed that the assignment of this new strain to a new species was justified (Wang et al, 2016).

Comparative analysis of the D1FHS genome with the genomes of the three other type strains of the genus, *N. halophilus* Nc4, *N. oceani* C-107 and *N. watsonii* C-113, provided the basis for the identification of two eco-genotypes of the genus, the urea uptake and hydrolysis-positive oceani/watsonii type and the ammonium uptake-positive *wardiae*/*halophilus* type. Both eco-genotypes are urea-genic by function of the urea cycle even though the *N. halophilus* Nc4 genome has lost the arginase gene required to close the urea cycle. The comparative analysis also confirmed that the archetypal genus encoded [NiFe] hydrogenase, which potentially serves as a supplemental mechanism to access external reductant. The *N. wardiae* strain D1FHS is also potentially capable of accessing its primary source of energy and reductant from cyanate; however, our analysis revealed that this putative catabolic function was imported by lateral transfer and does not constitute a property of the ancestor of the genus *Nitrosococcus.* Phylogenetic analysis based on housekeeping genes and genome analysis supports that the *N. wardiae* species represented by its type strain D1FHS is the archetype of the *Nitrosococcus* genus.

Table 3.1: General characteristics and differentiation of the genus *Nitrosococcus*

Characteristics	*Nitrosococcus wardiae* D1FHS	*Nitrosococcus halophilus* Nc4	*Nitrosococcus watsonii* C-113	*Nitrosococcus oceani* C-107
16S rRNA genes	2	2	2	2
tRNA genes	45	44	45	46
Plasmids	0	1	2	1
Genome size	4,022,640 bp	4,079,427 bp	3,328,579 bp	3,481,691bp
G+C content	50.70%	51.60%	50.10%	50.40%
Genes in total	4179	3895	3106	3219
hypothetical proteins	4128	3664	2891	3121
Cell shape and Cell arrangement	Coccus or short Rod	Coccus or short Rod	Coccus or short Rod	Coccus or short Rod
	Singles & Pairs	Singles & Pairs	Singles & Pairs	Singles & Pairs
Temperature optimum	37 °C	28 - 32 °C	28 - 32 °C	28 - 32 °C
Salinity optimum	700 mM	800 mM	600 mM	500 mM
pH optimum	7.6 - 8.0	7.6 - 8.0	7.6 - 8.0	7.6 - 8.0
Ammonium tolerance	< 300 mM	< 600 mM	< 1600 mM	< 1200 mM
Ammonium optimum	50 mM	100 mM	100 - 200 mM	100 mM
Use of Urea Use of Cyanate	- +	- -	+ -	+ -
Reference	This study, Wang et al., 2016	Chapter 4; Koops et al., 1990	Chapter 4; Campbell et al., 2011a	Klotz et al., 2006

Table 3.2: Sulfur cycle

sulfate adenylytransferase 2.7.7.4 subunit1 subunit2	 1913164..1914426 1912290..1913138
adenylysulfate kinase 2.7.1.25	167149..167778
phosphoadenosine phosphosulfate reductase 1.8.4.8	1911464..1912165
NADPH-dependent sulfide reductase 1.8.1.2	3588800..3590626
thioredoxindisulfide reductase 1.8.1.9	1434019..1434978
thiol:disulfide interchange protein	1491697..1492353
thiosulfate sulfurtransferase 2.8.1.1	667842..668699
putative polysulfide reductase 1.2.7.- (quinol oxidase)	2101074..2101616 2101627..2104614 2104614..2105999
monoheme cytochrome *ccc*A	yes
cytochrome c oxidase	yes
ubiquinone complex	yes

Table 3.3: C1 carbon metabolism

methane oxidation to methanol	AMO (EC 1.14.13.25)	
methanol oxidation to formaldehyde (FA)	methanol dehydrogenase cluster (EC 1.1.99.8)	large subunit protein (3 genes 976,190...976,726; 976,777...977,034; 977,044 ... 977,910)
FA oxidation to formate		
Glutathione (GSH)-dependent pathway		
S-hydroxy netgl-GSH to S-formyl GSH	GSH-dependent FA DH	Not present
S-formyl GSH to formate	GSH S-transferase	Not present
Tetrahydrofolate (THF)-dependent pathway		
Condensation of THF with formaldehyde	5,10-Methylene THF reductase (EC1.5.1.20)	323,494...324,387 3884160...3885014
Methylene THF to methenyl THF	Methylene THF DH (EC 1.5.1.5)	Not present
Methenyl THF to formyl THF	methenyl THF cyclohydrolase (EC 3.5.4.9)	Not present
Formyl THF to formate	Formyl THF hydrolase (EC3.5.1.10)	962408...962971
Formyl THF to formate	5-formyl THF cycloligase (EC 6.3.3.2)	969086...969949
Tetrahydromethanopterin (THMPT)-dependent pathway		
Condensation of THMP with FA	formaldehyde activating enzyme	826944...827840
Methylene THMPT to methenyl THMPT	Methylene THMPT DH (EC1.5.99.9)	964850...965839
Methenyl THMPT to formyl THMPT		
Formyl THMPT to formate		

Table 3.4: Glycolysis

1. glucose->G6P	hexokinase 2.7.1.1	ketohexokinase 2.7.1.3
2. G6P->F6P	phosphoglycose isomerase 5.3.1.9	3762871...3764532
3.F6P→F 1,6PP	phosphofructokinase 2.7.1.11	Tagotose-6-phosphate kinase/1-phosphofructokinase (2.7.1.56) 3205191...3206129 Fructose-1,6-bisphosphatase, type EC3.1.3.11 829101...830108
4.F1,6PP->DHAP+G3P	Fructose-bisphosphate aldolase 4.1.2.13	Fructose-bisphosphate aldolase class I 3507999 ... 3509048 Class II 1417316... 1418380 2625734...2626969
5. DHAP->G3P	Triosephoephate isomerase 5.3.1.1	574949...575668
	proton-translocating pyrophosphatase 3.6.1.1 soluble pyrophosphatase	3096154...3098247 2234331...2234945
	NAD(P) transhydrogenase 1.6.1.2	378349...379797 (beta) 379810...381381 (alpha)

Table 3.5: Stress tolerance

class		No. of genes
Stress response		**87**
	Heat shock	16
	Cold shock	1
	Osmotic stress	19
	Oxidative	33
	Periplasmic stress	5
	Non-classified	13
Virulence, disease and defense		**43**

Table 3.6: Transporters in the genome of *N. wardiae* D1FHS

Transporter type	Number of genes	Function
ABC transporter	88	ABC transporters are transmembrane proteins that utilize the energy of adenosine triphosphate (ATP) binding and hydrolysis to carry out certain biological processes including translocation of various substrates across membranes and non-transport-related processes such as translation of RNA and DNA repair
Na+/H+ antiporter	12	Membrane proteins that exchange Na^+ for H^+
Ion Channels	11	Energy-independent facilitated diffusion
Iron uptake TonB/ExbB/ExbD Siderophore	25 16 9	 TonB-dependent Fe-siderophore/receptors
PTS system	5	Phosphoenolpyruvate-dependent phosphotransferase
Type II secretion	28	General secretory pathway
Type IV secretion	21	Conjugal DNA-protein transfer
Porin	1	Passive-mediated transport/facilitated diffusion
Efflux pump	37	Iron, antimicrobial peptide, multidrug, lysophospholipase L1 biosynthesis, polysaccharide/polyol phosphate, protease/lipase, bacteriocin/antibiotic export
Others	26	

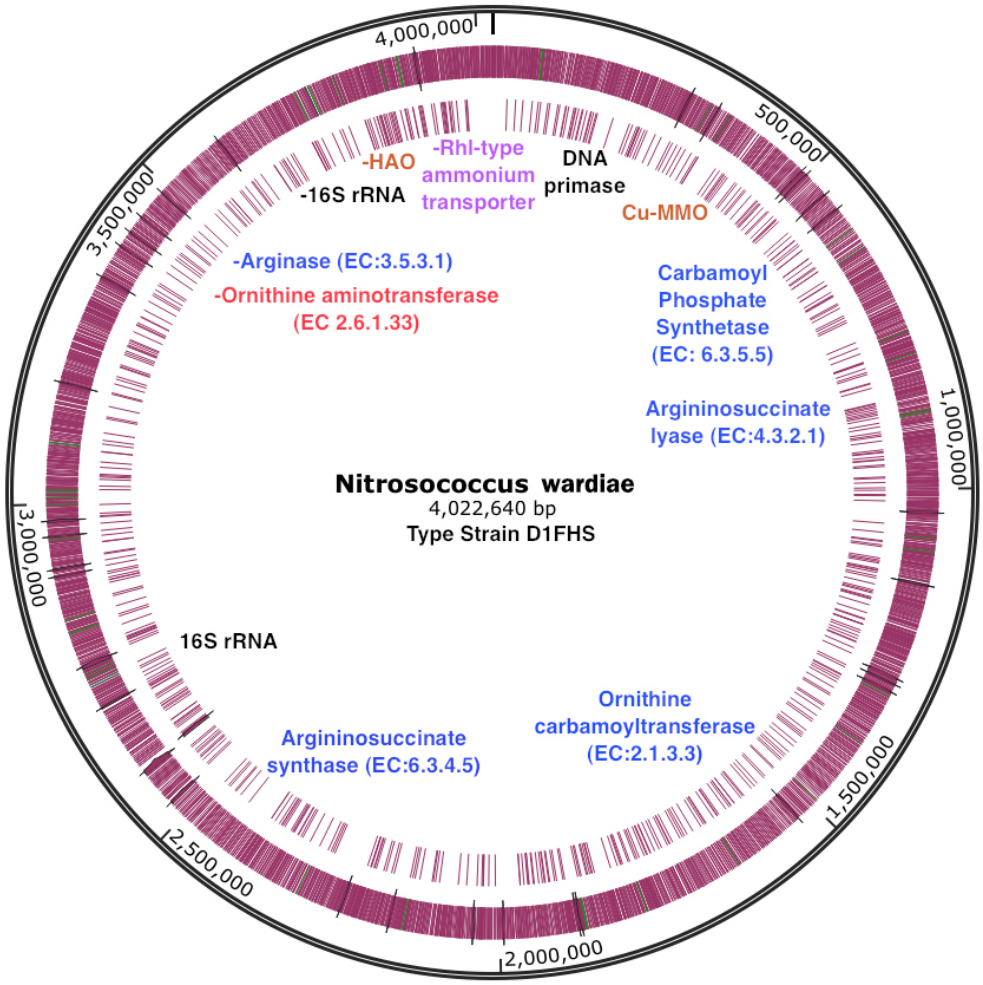
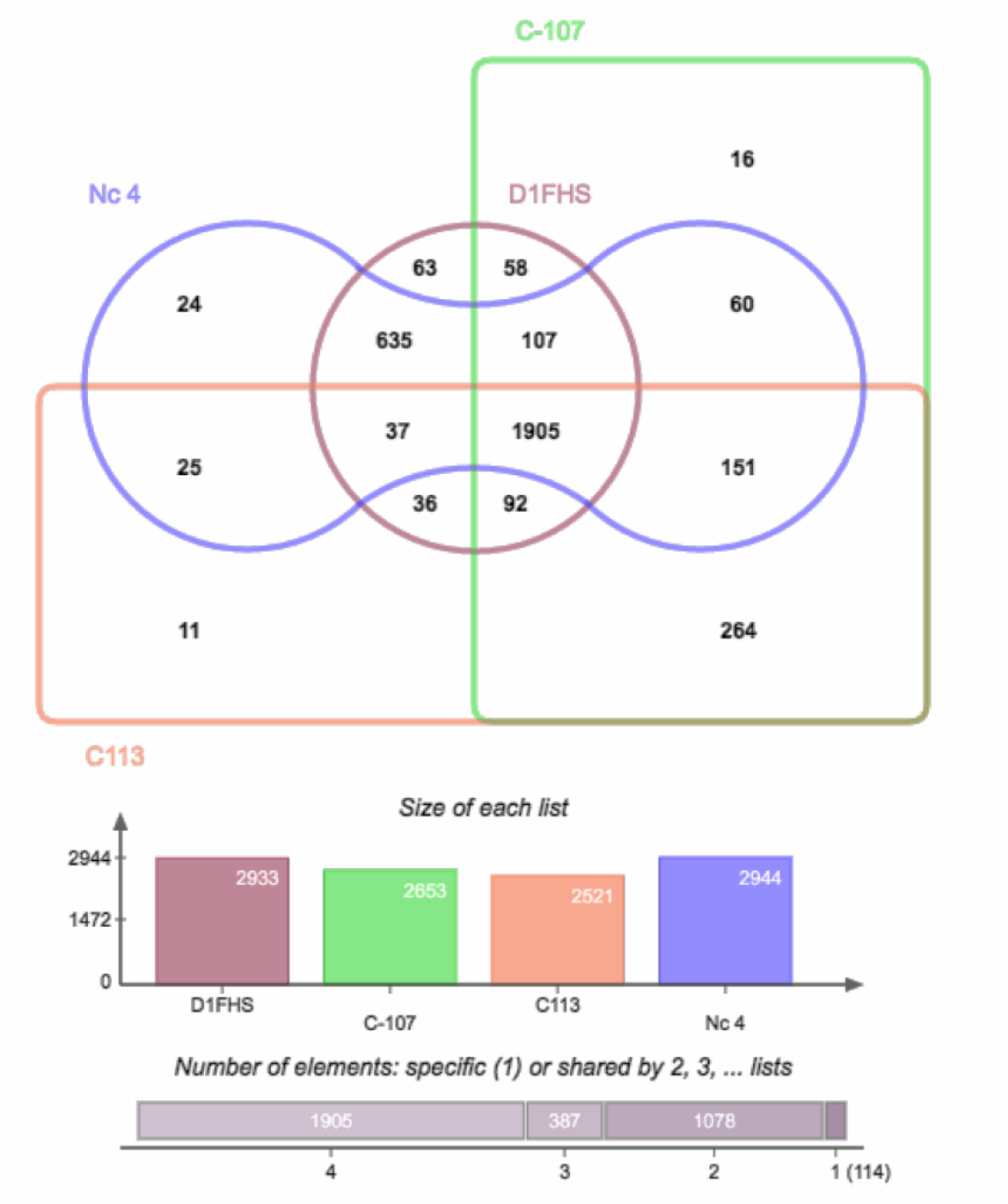


Figure 3.1: Circular genome of *Nitrosococcus wardiae* strain D1FHS. The outermost circle (black color) depicts the locations of genes. The inner two circles (purple color) indicated forward and reverse strand genes. Specific genes were manually annotated as indicated at the center.


C-107
16
Nc 4
D1FHS
63
58
24
60
635
107
37
1905
25
151
36
92
11
264
C113
Size of each list
2944
1472
0
2933
2653
2521
2944
D1FHS
C-107
C113
Nc 4
Number of elements: specific (1) or shared by 2, 3, ... lists
1905
387
1078
4
3
2
1 (114)

Summary

Species	Proteins	Clusters	Singletons
D1FHS	4128	2933	999
Nc 4	3664	2944	628
C113	2891	2521	341
C-107	3121	2653	419

Figure 3.2: VENN Diagram drawn using the OrthoVenn web server (Wang et al., 2015) showing the genome comparison of four *Nitrosococcus* species. The core genome of the four species of genus *Nitrosococcus* consists of 1905 CDs. The species form 3484 clusters, 3370 orthologous clusters (at least contains two species) and 1861 single-copy gene clusters.

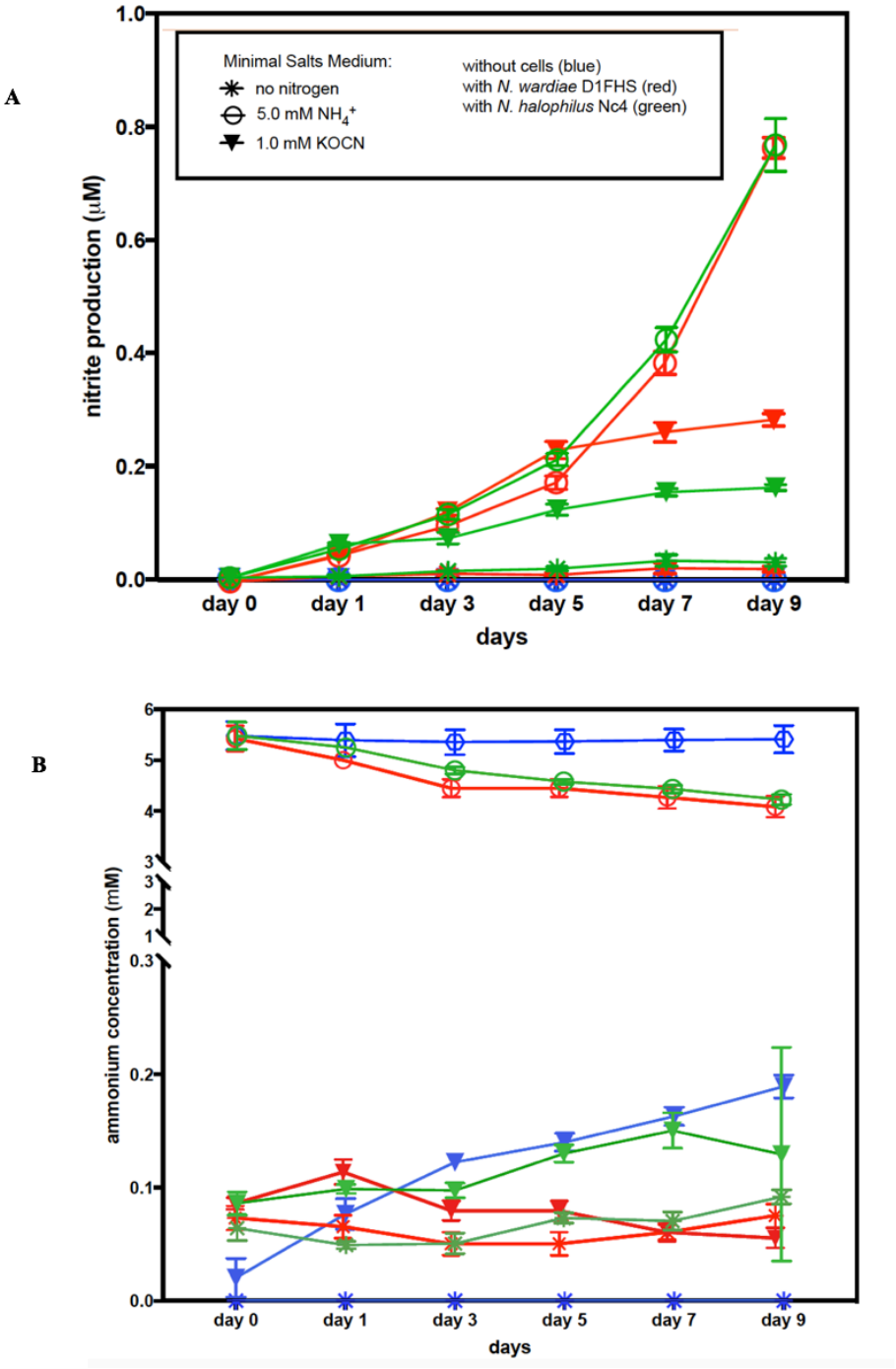


Figure 3.3: Growth curves of *Nitrosococcus wardiae* D1FHS *and Nitrosococcus halophilus* Nc4 in the presence of 5.0 mM ammonium or 1.0 mM cyanate.

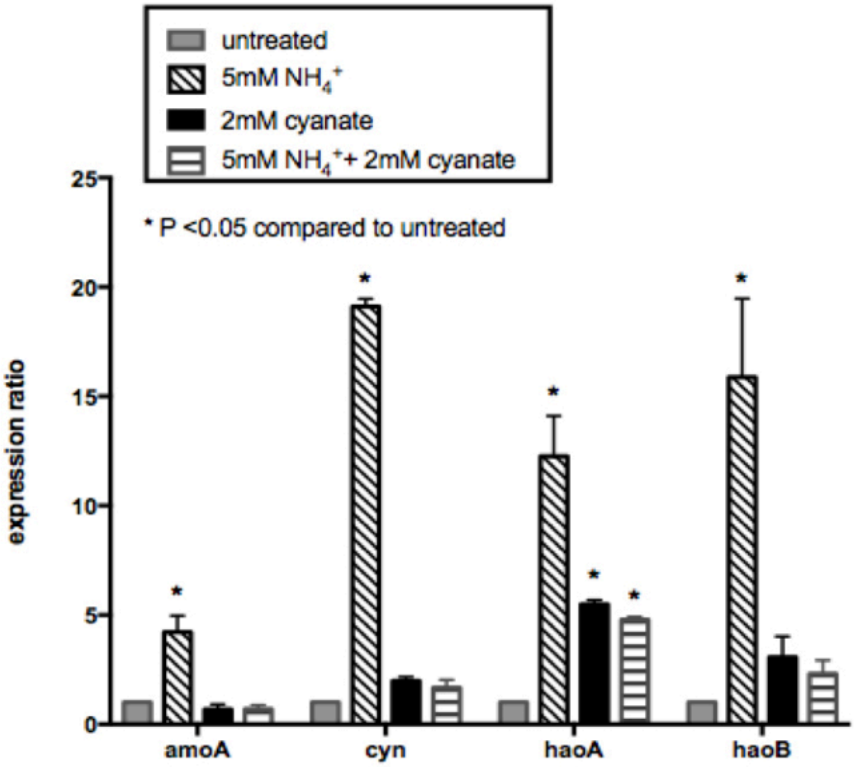


Figure 3.4: Relative abundance of genes *amo*A, *cyn*, *hao*A and *hao*B (normalized to 16S RNA gene transcripts) in *N. wardiae* strain D1FHS under different growth conditions determined using quantitative PCR. Steady-state levels of mRNAs from cells treated with 5 mM NH_4^+ (diagonal bar), 2 mM cyanate (black bar) or 5 mM NH_4^+ plus 2 mM cyanate (strip bar) were compared to the levels of the controls (no nitrogen sourcse; grey bar). Each qPCR reaction was performed in triplicate. Data bars are mean values of two independent trails with standard deviation error bars. Significance of difference was calculated by two-way ANNOVA and multiple *t*-test.

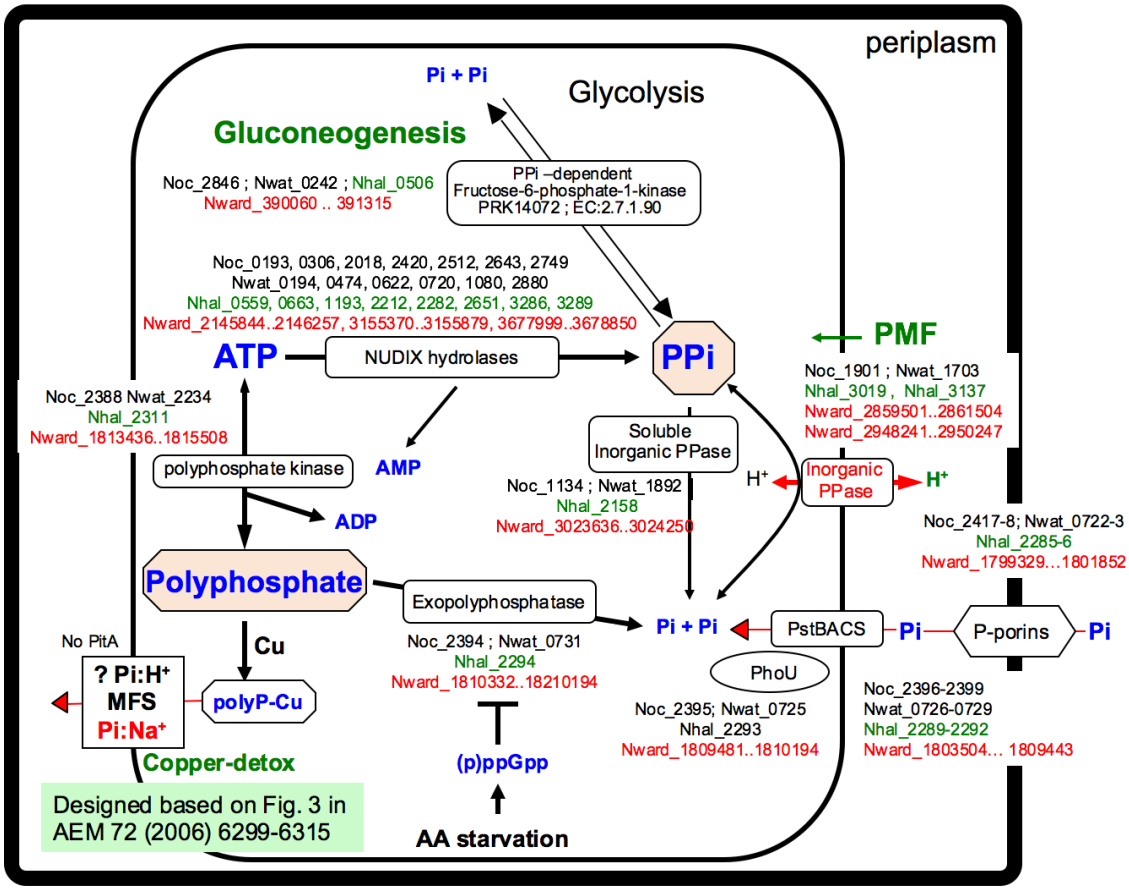


Figure 3.5: Comparison of proposed cycling of phosphate, polyphosphate and pyrophosphate and their involvement in energy metabolism, central pathways and stress tolerance in cells of four species of genus *Nitrosococcus*. Loci of genes encoding enzymes or transporter were listed below it (black: *N. oceani*; green: *N. halophilus* and red: *N. wardiae*).

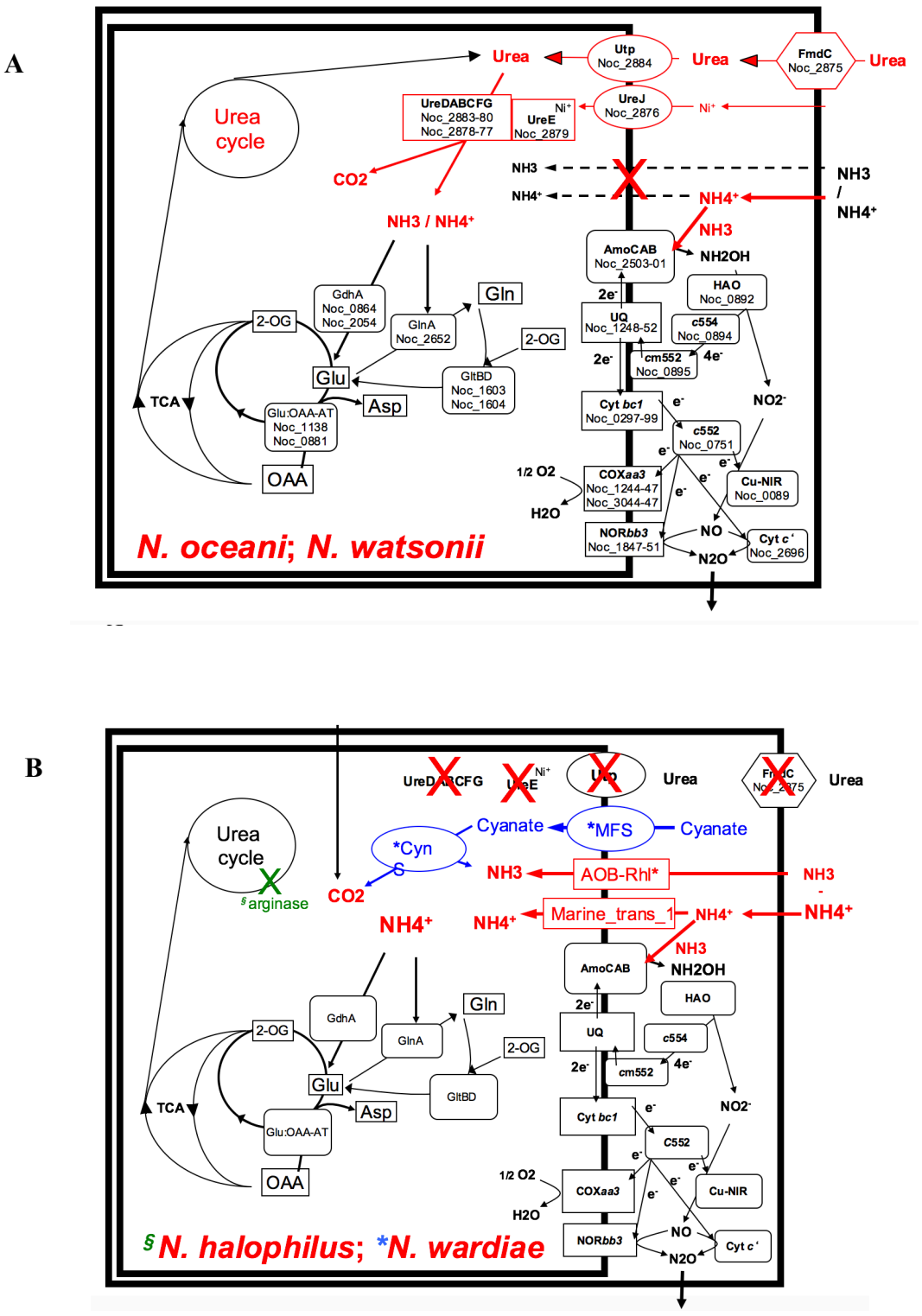


Figure 3.6: Comparison of proposed nitrogen uptake and assimilation in four *Nitrosococcus* species.

CHAPTER 4: COMPARATIVE GENOMIC ANALYSIS OF STRAINS OF THE AMMONIA-OXIDIZING BACTERIUM *NITROSOCOCCUS OCEANI* FROM VARIOUS OCEANIC LOCATIONS

4.1 Abstract

Strains of *Nitrosococcus oceani* are ammonia-dependent obligate aerobic chemolithotrophic bacteria that gain energy and reductant needed for assimilation and growth solely from the oxidation of ammonia. They are omnipresent in the world's oceans and as such important to the global nitrogen cycle. We compared high quality draft genome sequences of *Nitrosococcus oceani* strains isolated from the North (AFC27) and South (AFC132) Pacific Oceans and the coastal waters near Barbados at the interface between the Caribbean Sea and the North Atlantic (C-27) Ocean with the complete genome sequence of *N. oceani* C-107 (=ATCC 19707; JMC30415), the type strain isolated from the open North Atlantic, with the goal to identify indicators for autochthonous or allochthonous origins of the species.

Intriguingly, the chromosomes of strains C-107, C-27 and AFC27 were highly conserved in content and synteny, and all genomes contained one nearly sequence-identical plasmid despite being isolated from different oceanic gyres. In contrast, the genome of strain AFC132 revealed the presence of inventory unknown from other marine ammonia-oxidizing bacteria such as complete gene complements encoding NiFe-hydrogenase and a NRPS-like siderophore biosynthesis module. Comparative analysis of content and organization of the genomes of these closely related strains in context with

the literature suggests that AFC132 represents a metabolically more diverse ancestral lineage to the other strains with C-107 potentially being the youngest. The results suggest that the *N. oceani* species evolved by genome economization through the loss of many genes, which could have been facilitated by their rich complements of CRISPR and Restriction Modification systems.

4.2 Introduction

Strains of the purple sulfur bacterium *Nitrosococcus oceani* are Gram-negative marine Gammaproteobacteria in the family *Chromatiaceae* (Campbell et al., 2011a). The first sequenced genome from *Nitrosococcus oceani* strain C-107 (=ATCC 19707; JMC30415) provided a first snapshot of the genetic makeup of the species (Klotz et al., 2006). *Nitrosococcus oceani* strain C-107 is omnipresent in the world's oceans (Ward and O'Mullan 2002) and the type strain of the species in a genus of exclusively obligate aerobic and ammonia-dependent chemolithoautotrophs belonging to the ecophysiological cohort of "ammonia-oxidizing bacteria" (AOB). This ammonia-catabolic lifestyle provides the energy and reductant required for the assimilation of carbon from carbon dioxide as the sole carbon source (Arp et al., 2007, Stein et al., 2013). Oxidation of ammonia to nitrite ("nitritation") occurs via a two-step pathway in the periplasm: Initially, ammonia is oxidized to the toxic intermediate hydroxylamine ($NH_3 + 2e^- + O_2 + 2H^+ \rightarrow NH_2OH + H_2O$) by ammonia monooxygenase, an integral membrane protein complex encoded by the *amo*RCABD operon (El Sheikh and Klotz 2008). Hydroxylamine is then converted to nitrite by the membrane-associated periplasmic hydroxylamine dehydrogenase (HAO), encoded by the *haoA* gene in the *haoA-haoB-cycA-cycB* operon. While the 5'-end of the *haoB* transcript forms a stemloop that leads to

two pools of steady-state *haoA* and *haoAB* mRNAs (Stein et al., 2013), redox-active proteins that have been implicated in the channeling of the 4 electrons extracted from hydroxylamine by HAO to the ubiquinone pool (Hooper et al., 1997); however, participation of cytochrome *c*554 in this process has not been experimentally demonstrated (Simon and Klotz 2013, Stein et al., 2013). Two of the 4 electrons are recycled to ammonia monoxygenase via a still unknown mechanism for the oxidation of ammonia whereas the other two electrons are fueling the oxidative branch of the electron transport chain (Arp et al., 2007, El Sheikh and Klotz 2008, Stein et al., 2013). Nitrite is either released into the environment or subsequently reduced to nitrous oxide via nitric oxide by a pathway known as "nitrifier denitrification" (Stein and Klotz 2016) facilitated by a copper-dependent nitrite reductase (NirK) and a complement of nitric oxide reductases (Klotz et al., 2006, Arp et al., 2007, Simon and Klotz 2013, Stein et al., 2013).

Because all known strains of *N. oceani* are distributed exclusively in marine environments albeit at an overall low abundance (Ward and O'Mullan 2002, Hozuki et al., 2010, Campbell et al., 2011a), produce nitrogen oxides with N_2O being a potent greenhouse gas (Ravishankara et al., 2009) and because the ocean is a major source of atmospheric N_2O (Voss et al., 2013), *Nitrosococcus* strains contribute potentially to the global climate crisis. This and the omnipresence of *Nitrosococcus oceani* in the world's oceans generated the question whether different strains of the species isolated from different oceanic gyres differed in their genomic contents; in particular, genes encoding the catabolic inventory that facilitates transformations of nitrogen oxides. In this work, the genomes form three strains in pure culture (C-27, AFC27 and AFC132) isolated from different oceanic locations were sequenced and comparatively analyzed together with the

sequence of the previously analyzed type strain C-107 (ATCC 19707; JMC30415) (Scott et al., 2006).

4.3 Materials and Methods

4.3.1 Growth Conditions

All *N. oceani* strains were maintained at 30°C in the dark without shaking in liquid ammonia mineral salts media (Alzerreca et al., 1999) amended with hydroxylamine at 100 μM final concentration to eliminate contamination (Campbell et al., 2011a)., The pH of cultures was adjusted daily to 8.0 using K_2CO_3 (Alzerreca et al., 1999, Klotz et al., 2006). Cultures were propagated monthly by transferring 5-ml aliquots of culture into fresh media. Cells were checked under a light microscope at 1000x magnification routinely for uniform cell morphology (Wang et al., 2016).

4.3.2 DNA Extraction and Genome Sequencing

500 ml of mid-exponential phase culture were centrifuged (6000 x g, 15 min, room temperature) to harvest cells for genomic DNA extraction (Campbell et al., 2011a, Wang et al., 2016). Resulting gDNA samples were quantified using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, USA) and checked for integrity by (1%) agarose gel electrophoresis. About 1 μg of high quality gDNA was obtained from each culture for genome sequencing.

Draft genomes of strains C-27 and AFC132 were generated using Illumina MiSeq Sequencing platform using TruSeq techniques for library construction with size selection default value of 300-400 bp, which produced 27,907,041 paired ends reads for strain C-27 and 28,295,664 reads for strain AFC132 after quality check and trimming. The whole genome of AFC27 was shotgun sequenced at the J. Craig Venter Institute (MD, USA)

using 454 FLX Titanium technology funded by The Gordon and Betty Moore Foundation marine microbial genome sequencing project (http://www.jcvi.org/cms/research/past-projects/microgenome/overview/). After quality trimming of the draft sequence, 30,135 reads were generated and assembled into 49 contigs and 12 scaffolds with no gaps. The total assembly gap length was 8,252 with an estimated total genome sequence length of 3,480,059 bp (https://www.ncbi.nlm.nih.gov/assembly/GCF_000155655.1).

4.3.3 Genome Assembly, Annotation, and Bioinformatics Analyses

After quality checks, adaptor trimmed raw reads from C-27, AFC27 and AFC132 genome templates were used for *de novo* assemblies performed with the CLC Genomics Workbench v6.5 (CLC Bio/Qiagen, Denmark), annotation Pipeline 3.0 (PGAP) whereas Mauve 2.3.1 (Darling et al., 2004) and the RAST SEED server (v4.0; Aziz et al., 2008) were utilized for comparative annotation. To facilitate our comparative study, the IslandViewer3 software package was utilized to predict and analyze genomic islands (Dhillon et al., 2015); antiSmash 2.0 (Blin et al., 2013) was employed to analyze secondary metabolite biosynthesis gene clusters. The web server PHAge Search Tool (Zhou et al., 2011) was used to search for prophage sequences inside the bacterial genomes. Genetic relatedness among *N. oceani* species were comparatively assessed by pairwise calculating the average nucleotide identity (ANI) between the analyzed genomes (Konstantinidis and Tiedje 2005). The genome browser software UGENE was utilized for further analysis of the genomes. The sequence synteny web server SyntTax was used to assess the orthology of genomic regions and to predict functional relationships between genes. CRISPR genes were identified in the *N. oceani* genomes by employing the web tool CRISPRFinder (Grissa et al., 2007). The construction of phylogenetic trees after

sequence alignment was facilitated with MEGA6 (Tamura et al., 2013). For identification of Toxin-antitoxin systems, RASTA-Bacteria (Sevin and Barloy-Hubler 2007) and TADB were utilized (Dang et al., 2011).

4.3.4 Genome Sequence Deposits

The Whole Genome Projects for *N. oceani* strains C-27, AFC27 and AFC132 have been deposited at DDBJ/EMBL/GenBank with the accession numbers NZ_JPGN00000000, NZ_ABSG00000000 and NZ_JPFN00000000, respectively.

4.4 Results

4.4.1 General Genome Characteristics of Strains

The *Nitrosococcus oceani* strains compared in this study were isolated from different locations in the Atlantic and Pacific oceans (Ward 2002). The strains' chromosomal genome sizes are similar to each other with sizes of about 3.5 Mb (Table 4.1). Similarly, the GC content of all three genomes is about 50% (Table 4.1). The alignment of the genomes from different strains revealed a high degree of synteny. In the AFC132 genome, one likely rearrangement was observed in a contig with 69,741 bp in length (AFC132x-1908-1961) with the rearrangement being at position 61289 bp of the contig (Figure 4.3). Reads mapping back to this contig show high depth coverage at the point of concern (about 1096x) while mapping of reads to the chromosomal genome of C-107 showed no coverage at the homologous position (i.e. at position 1570000 bp) (Fig. 4.3) suggesting that the rearrangement is highly likely true.

Phylogenetic analyses (Fig. 4.4 A and 4B) demonstrated a very close relationship between the investigated strains. This is supported as well by the average nucleotide identity (ANI) analysis (Konstantinidis and Tiedje 2005), which revealed a near 100 %

identical ANI value in pairwise comparisons of the four *N. oceani* genomes (Table 4.2). The lowest ANI detected is 98.28%, which exists between the genomes of strains C-27 and AFC132 (Table 4.2). Even though the genomes of AFC27 and AFC132 are high-quality drafts and not closed genomes, their inclusion does not compromise the power of ANI analysis (Konstantinidis and Tiedje 2005, Richter and Rossello-Mora 2009). The ANI values obtained for the pairwise comparisons are well above the cut-off point of 95 % for delineating a species (Konstantinidis and Tiedje 2005) thereby confirming the assignment of all four strains to the same species (*N. oceani*).

4.4.2 Genetic Diversity Between the *Nitrosococcus Oceani* Genomes

Analysis of orthologous gene clusters using OrthoVenn showed that a core of 2619 orthologous genes are shared among the genomes of *N. oceani* strains while five unique genes were found in AFC132 but none in the others (Fig. 4.2). This suggests a closer evolutionary relationship between C-107, C-27 and AFC27 than either might have with AFC132.

Pairwise calculations of the Average Nucleotide Identity (ANI) and alignments of the genome sequences from strains C-27, AFC27 and C-107 demonstrated a high level of sequence identity and conservation between these strains, including the arrangement of genes in the genomes (synteny, Table 4.2). Essentially, the main difference between the C-107 genome and the C-27 and AFC27 genomes is the presence of a large genomic island in the C-107 type strain, which is predicted by IslandViewer 3 (as discussed below). Another genetic difference between the strains was detected in the region that aligns with the noc_2361 and noc_2362 genes in C-107, which both encode Fis transcriptional regulators but with moderate sequence identity to each other. The

noc_2361 and noc_2362 genes in C-107 are combined to form a single gene in the AFC132 genome that is predicted to encode a sigma 54 transcriptional regulator in the same class of these regulators found in the C-107 genome. Interestingly, the C-27 and AFC27 genomes lack the DNA between the noc_2361 and noc_2368 homologs present in C-107 resulting in the absence of 5 genes and the fusion of noc_2361 and noc_2368 homologs to form a single gene coding for a flagellin protein in the two strains. This leaves the genomes of C-27 and AFC27 as the most closely related genomes of the 4 compared *N. oceani* strains even though both strains originated from different oceanic gyres. The main feature distinguishing the AFC27 and C-27 genomes from one another is a gene encoding a DNA polymerase III beta subunit domain that is inserted within the noc_0394 homologous hypothetical gene.

4.4.3 Nitrogen Metabolism Inventory

Given the importance of *N. oceani* in the nitrogen cycle, an in-depth comparative analysis for nitrogen metabolism-related genes was performed in this study. Generally, genes involved in the nitritation and denitrification pathways such as *amoCAB* and *haoA* as well as *nir*K and norCB are conserved in all of the *N. oceani* strains. The genomes also encode alternative nitric oxide reductases CytS (cytochrome c'-beta) and sNOR (Simon and Klotz). The expression products of these genes account for the abilities of all the strains to release nitrous oxide when the cultures are maintained in ammonium minimal media. Because also urease-encoding genes are found in the genome of all *N. oceani* strains, all four strains can grow on (assimilated N from) urea.

Prior analyses of the *N. oceani* C-107 genome (Klotz et al., 2006) reported that this species has only single copies of *amo* and *hao* operons while genomes of

betaproteobacterial AOB species such as *Nitrosomonas eutropha* C91 (Stein et al., 2007), *Nitrosomonas* sp. strain AL212 (Suwa et al., 2011), *Nitrosospira multiformis* (Norton et al., 2008) and *Nitrosomonas europaea* (Chain et al., 2003) contain multiple operon copies with nearly identical nucleic acid sequences in their genomes. As with C-107, our data revealed that the genomes of the three sequenced *N. oceani* strains in this study also have single copies of *amo* and *hao* gene clusters in their genomes. Thus, the notion that *N. oceani* genomes contain only single copies of *amo* and *hao* gene clusters can be generalized. In addition, all three *N. oceani* genomes encode the red copper protein nitrosocyanin (Arciero et al., 2002a, Nakamura et al., 2003), which might play a role in ammonia catabolism by serving as an electron carrier protein (Arciero et al., 2002a) or by affecting electron flow indirectly by interaction with HAO or the implicated cytochrome proteins *c*554 or c_M552 (Stein et al., 2013). The nitrosocyanin-encoding gene (*nyc*A) has, so far, been discovered only in AOB genomes except for the genome of *Nitrosomonas* sp. strain Is79 (Bollmann et al., 2013).

All the *N. oceani* genomes also contain the genes encoding cytochromes P460 (*cytL*) and *c*'-beta (*cytS*), two members of high-spin cytochromes *c* (Elmore et al., 2007). The gene encoding cytochrome P460 was found to be transcriptionally active when grown in the presence of ammonium (Stein et al., 2013). The gene product was shown to oxidize hydroxylamine, although at a much lower rate than *hao* (Numata et al., 1990) with the possible pathway of the end production of nitrite by co-oxidation of NO, possibly in conjunction with Noc_0889 homolog pan1-type MCO (Klotz et al., 2006), which is present in other strains as well. This protein was originally hypothesized to be involved in hydroxylamine detoxification (Bergmann and Hooper 2003b); however, the

recently proposed model of hydroxylamine oxidation in ammonia-oxidizing Thaumarchaeota (Kozlowski et al., 2016) implicates this NO and NH_2OH co-oxidation chemistry as equally important for NO detoxification. Therefore, the role of the beta-sheeted cytochrome *c*'-beta is likely to mitigate toxicity of NO that escapes from incomplete hydroxylamine oxidation to nitrite by HAO under fully oxic conditions (Hooper and Terry 1979, Klotz and Stein 2011).

4.4.4 CRISPR/Cas System

Comparative analyses employing CRISPRFinder (Grissa et al., 2007) revealed that the genomes of strains C-107, C-27 and AFC27 contained the same CRISPR/Cas signature. In contrast, the genome of strain AFC132 exhibited a different CRISPR signature. The cluster of CRISPR-associated genes in the genome of AFC132 exhibited only low to moderate sequence identity as well as genetic organization in comparison to the other strains although the CRISPR gene clusters in all four genomes appear to belong to same subtype I-F (Siebers et al., 2011). The alignment of all 4 genomes indicated that the CRISPR gene clusters are located at the same position in the genome; suggesting that the diverging CRISPR clusters are the outcome of a replacement by heterologous recombination.

A detailed comparison of the CRISPR clusters identified that the direct repeat lengths were identical (28 bp) in the four *N. oceani* genomes (Table 4.3) albeit with a consensus sequence that differed between the genomes from AFC132 in the other strains. In addition, the CRISPR region in the genome of AFC132 was larger with a length of 2248 bp and the insertion of 37 spacers compared to the CRISPR length in the genomes of C-107, C-27 and AFC27, which was only 387 bp in length with only 6 spacers (Table

4.3). Furthermore, the genetic organization of the CRISPR/Cas gene cluster in the AFC132 genome differed from that of the rest of the strains. While the CRISPR/Cas gene cluster in AFC132 is organized in a unidirectional order, the CRISPR/Cas gene clusters in the other genomes are divided into two parts in a bidirectional manner with the *cas* genes and *cys* genes being juxtaposed. The significant reduction in CRISPR cluster length in C-107, C-27 and AFC27 compared to AFC-132 could be the result of functional pressure exercised by extensive editing and concomitant loss to a point where further loss would be lethal.

4.4.5 Restriction Modification Systems

Currently, four types of restriction modification (RM) systems are recognized (Roberts et al., 2003). The genome of the *N. oceani* type strain C-107 was reported to encode a large number of RM systems (Klotz et al., 2006). According to the REBASE database (Roberts et al., 2015) the chromosome of C-107 carries 20 methyltransferase gene clusters with another one on its plasmid. Analysis of the C-27 and AFC27 genomes revealed that both strains contain the same RM inventory as on the chromosome of C-107. In contrast, the total number and types of RM systems encoded in the AFC132 genome is smaller and some systems are incomplete. For instance, an M subunit-encoding gene (Noc_0443) of type I RM system (ORF439) is absent from the AFC132 genome. Also, AFC132 lacks the homologs of Noc_1157-64 that are found in the genomes of the other strains and encode a type I RM system. More dramatically, genes encoding two Type I RM systems (noc_1158-60 and noc_1201-05) are observed in the genomes of all strains except AFC132. For the latter strain, the locus of noc_1201-05 included a unique gene cluster encoding a Type III RM. In the location of noc_1158-60,

only a remnant of noc_1159 is found in the AFC132 genome. In contrast, the AFC132 genome contains two unique gene clusters encoding Type II RM systems that are located at unique locations compared to the 3 other *N. oceani* genomes. Their residence in the AFC132 genome is most likely the result of horizontal transfer. One of these Type II gene clusters contains also the genes encoding a Type IV system. Interestingly, the R (noc_2927 homolog) and M (noc_2928 homolog) genes in another system in the AFC132 genome are separated by a gene encoding a protein with a "phage abortive infection" (Abi) system domain (see below for discussion). Observed coding differences in some of the genes encoding S and M proteins could affect the sequence specificity or the methylation activity of the restriction modification systems.

4.4.6 Polysaccharide and Glycosyl Transferases

All *Nitrosococcus oceani* genomes encode a large complement of glysosyl transferases and inventory involved in polysaccharide biosynthesis; however, there are differences between the strains with regard to extent and location. For example, the gene encoding a capsular polysaccharide (noc_1971) in C-107 is replaced by a longer gene cluster containing genes related sulfotransferase, methyltransferase, glysosyl transferase and polysaccharide biosynthesis in the homologous position in AFC132. Noc_1971 is predicted by the DOOR2 database (http://csbl.bmb.uga.edu/DOOR/) to be part of an operon containing glysosyl transferase genes upstream of the former gene. The rest of the genes in the operon are found in AFC132 as well. In contrast, the C-27 and AFC27 genomes retained the operon found in C-107 at the same location. The noc_1942 encoding a polysaccharide biosynthesis protein in the C-107 genome is missing at the respective position in C-27 and AFC27 genomes whereas the AFC132 genome contains

the entire gene cluster noc_1938 to noc_1946. Also, noc_1988 encoding glysosyl transferase found in C-107, C-27 and AFC27 is substituted by a unique segment in the AFC132 genome containing small hypothetical genes as well as genes possibly encoding a glysosyl transferase, glucose-methanol-choline oxidoreductase and a transposase. Furthermore, a large gene cluster (AFC132x_2704-16) containing various glysosyl transferase genes as well as genes encoding polysaccharide biosynthesis and aminotransferase reside in the AFC132 genome in place of a small hypothetical gene (noc_2541) in the other strains.

4.4.7 Toxin-Antitoxin Systems

Differences in the molecular inventory encoding toxin-antitoxin systems between the AFC132 and the C-107, C-27 and AFC27 genomes were observed. In the AFC132 genome, a gene cluster (AFC132_2275-78) containing genes encoding an ABC permease as well as a putative toxin-antitoxin system that are uniquely inserted between the noc_2143 and noc_2144 genes in the other 3 genomes. On the other hand, the noc_0437-0438 Toxin-antitoxin system is absent from the AFC132 genome.

4.4.8 Mobile Genetic Elements

Analyses on the genomes revealed that transposase genes are abundant and distributed widely across the genomes of all *N. oceani* strains. While there are conserved transposon insertions in the chromosomes of all *N. oceani* strains, we identified several unique transposon insertions at varying loci. Some of the transposon insertions in one strain are replaced by a cluster of multiple coding genes in other genomes suggesting the occurrence of horizontal gene transfer events.

Although the high quality draft status of the C-27, AFC27 and AFC132 genomes precludes exclusive statements about the absence of transposon elements present in C-107, we nevertheless made some interesting observation regarding genes that are present in the genomes. For example, the AFC132_1452 single transposase (OrfB_IS605 family), also found in C107, is not observed in the corresponding position in the genomes of C-27 and AFC27. Likewise, the noc_1199 transposase gene of the same family is not observed in the corresponding position of AFC132 genome but are resident in the C-27 and AFC27 genomes.

In addition to single transposase insertions in the genomes, we observed large aggregations of mobile elements within small regions. Examples of these are homologous regions of noc_0046-54 and noc_0067-74, which are conserved in the genomes of C-107, C-27 and AFC27 but not in AFC132. These conserved insertions constitute either horizontally acquired genomic islands in the ancestor of the C-107, C-27 and AFC27 genomes after delineating from AFC132 or the islands were lost from the AFC132 genome after its delineation from the common ancestor genome. The noc_0046-54 gene cluster includes several hypothetical coding genes as well as a gene encoding a TOPRIM domain-containing protein.

A unique cluster of insertion elements is observed in the AFC132 genome at a position between gene homologs of noc_2929 and noc_2930 in the other three genomes. This unique region contains a few putative coding genes including one that encoding a DEAD/DEAH box helicase-like protein and another encoding an excisionase family DNA-binding domain. This region is marked as a possible genomic island using IslandViewer3 analysis.

4.4.9 The Genomic Island in C-107

Bioinformatics analyses using the IslandViewer3 software revealed several putative genomic islands in the genome of C-107. The most conspicuous genomic island found only in the C-107 genome encompasses a 41-kb DNA region (nt 713786-754986) and contains 38 genes. Most of these genes are annotated as hypotheticals; however, the island encodes also a copy of the molecular chaperone DnaK and a pair of genes encoding a two-component regulatory system. Unsurprisingly, this genomic island contains several genes encoding mobile element functions such as transposase, phage integrase and resolvase-like protein. The genes flanking this genomic island in the C-107 genome are conserved in the other three *N. oceani* genomes where they sandwich a mobile element thereby identifying this location as an insertion hot spot.

Analysis employing IslandViewer3 identified gene cluster noc_0431-49 as another genomic island in the C-107 genome. Examination of this region identified closely clustered genes encoding two Toxin-antitoxin systems and one RM system. While this genomic island is conserved in the C-27 and AFC27 genomes, the homologous region in AFC132 lacked the genes encoding the Toxin-antitoxin systems and the sequence of the gene encoding the DNA-specificity subunit of the RM system was modified. In light of the pertinent literature, this genomic region could be designated a defense island (Makarova et al., 2013). The genomes of C-107, C-27 and AFC27 but not AFC132 contained another possible genomic island predicted by IslandViewer3 (noc_0581-0585) that encodes putative membrane proteins that could contribute to the formation of a phosphate permease.

4.4.10 Prophage DNA

The genome of AFC132 includes one intact prophage gene cluster as identified by the PHAST web server. In contrast, only one remnant prophage gene cluster (noc_1855-87) was discovered in the closed C-107 genome (Klotz et al., 2006), which we have identified in the same location in the C-27 and AFC27 genomes but not in AFC132. Interestingly, a DNA region containing phage integrase (AFC132_2122-2126) is inserted into the arginyl-tRNA gene (homolog of Noc_R0034), resulting in the duplication of the arginyl-tRNA gene in the AFC132 genome.

4.4.11 Extra-Chromosomal DNA (Plasmid)

The genomes of strains C-27 and AFC27 appear to include a plasmid that is identical or near identical in sequence to plasmid A in the C-107 strain. In detail, a contig of 40-kb in the assembly of C-27 genome data corresponded to plasmid A of C-107 with 100 % nucleic acid identity. When using the CLC Genomic Benchwork 6.5 software and the sequence of C-107 plasmid A as the template for mapping raw reads from the other *N. oceani* genomes, we were able to thread the reads belonging to C-27 and AFC27 along the entire length of the C-107 plasmid A sequence with successful circularization towards the ends. Thus, this suggests that plasmid A is present as an independent replicon in strains C-27 and AFC27. For AFC27, the plasmid is found as one of the scaffolds in the AFC27 whole genome sequencing data with overlapping ends.

After joining the overlapping ends in the scaffold of the AFC27 whole genome sequencing data that was sequence-identical to C-107 plasmid A, the sequence resulted a size of 40,421 bp, which is one nucleotide more than in plasmid A of C-107 and C-27 (40,420 bp). We were unable to find evidence for extra-chromosomal DNA in our sequencing data of the AFC132 genome.

4.5. Unique Inventory Encoded in the AFC132 Genome

4.5.1 Nonribosomal Peptide Synthetase (AFC132 contig_23)

All four *N. oceani* strains are equipped to produce siderophores that serve to obtain iron as a key nutrient from the environment because the genomes of all strains contain the hydroxamate-type siderophore biosynthesis gene cluster (Noc_1811 to Noc_1814) as reported for the genome of the C-107 type strain (Klotz et al., 2006). Based on antiSmash 2.0 analysis, the AFC132 genome contains a unique gene cluster encoding nonribosomal peptide synthetase (NRPS) that may enable AFC132 to facilitate the production of additional siderophores. In addition, all four genomes contain a number of genes that are homologous to some of the genes in the NRPS gene cluster of AFC132. One of these genes conserved at the same locus in all strains is homologous to ferrichrome iron receptor Noc_0321, Another gene, Noc_0322, is predicted to encode a PepSY-associated trans-membrane helix, which is putatively involved in reduction of ferric siderophores (Eijkelkamp et al., 2011).

4.5.2 NiFe-Hydrogenase (AFC132 contig_18)

Based on their main redox metal, hydrogenases are generally categorized in three classes, which are the NiFe-hydrogenase, FeFe-hydrogenase and Fe-hydrogenases (Vignais and Billoud 2007). Of the four genome-sequenced *N. oceani* strains, only the AFC132 genome was found to encode a putative large NiFe-type hydrogenase in group 3b (Vignais and Billoud 2007), which is encoded by a unique gene cluster at a locus aligning with the Noc_2716 gene. A search employing the SyntTax web server (Oberto 2013) identified the presence of NiFe-hydrogenase in a few other species with identical genetic arrangement as in the pertinent AFC132 gene cluster in that the *hyp*AFCDE

genes are succeeded downstream by the structural genes for the hydrogenase subunits. These include *Nitrospina gracilis* strain 3/211 (Lucker et al., 2013) as well as *Marinobacter lipolyticus* SM190, *Marinobacter santoriniensis* NKSG1, *Rhodothermus marinus* SG0.5JP17-172, *Rhodothermus marinus* DSM 4252, *Sorangium cellulosum* So0157-2, *Sorangium cellulosum* So-ce56 and *Luminiphilus syltensis* Nor5-1B (Spring et al., 2013).

4.5.3 Terpene Synthesis Genes

The four *N. oceani* strains share two large gene clusters (Noc_1022-55 and Noc_1304-41) whose expression products might be involved in terpene metabolism. Using antiSmash 2.0 analysis, the proteins encoded by these two gene clusters are predicted to produce ectoine-terpene and phosphonate-terpene. In the AFC132 genome, a gene (noc_2203) conserved in the C-107, C-27 and AFC27 genomes, is replaced by a unique cluster of genes including one (AFC132x_2347) encoding a putative terpene biosynthesis protein, that is preceded upstream by gene encoding a possible Crp/Fnr family transcriptional regulator.

4.5.4 Other Uniquely Missing or Present Elements in the AFC132 Genome

A number of homologous genes or cluster of genes conserved in the C-107, C-27 and AFC27 genomes were not identified in the genome of AFC132. For example, the gene noc_1342 encoding peptidoglycan-binding protein LysM that interacts with peptidoglycan hydrolases in bacteria (Buist et al., 2008) was not identified in the AFC132 genome. Likewise, the gene homologous to noc_0057 and encoding a putative ATPase in the other strains was absent from the AFC132 genome. A cluster of genes (noc_0465-

0469) including a gene encoding glucose-methanol-choline oxidoreductase was remarkably absent from the AFC132 genome.

Another form of differences between the genomes of AFC132 and the other strains is the replacement of chromosomal sections conserved in C-107, C-27 and AFC27 with DNA unique to AFC132. One example with putative biological significance is the replacement of a three-gene cluster including noc_1135 that encodes DnaJ in C-107, C-27 and AFC27 with a cluster of putative chemotaxis genes (AFC132_1189-92). Likewise, the noc_2388-2390 homologous genes conserved in C-107, C-27 and AFC27 are substituted by two cytochrome *c*-encoding genes in the genome of AFC132. Further, two hypothetical genes (noc_1612-13) conserved in C-107, C-27 and AFC27 are replaced in the AFC132 genome by a gene encoding a putative DOPA 4,5-dioxygenase, an enzyme known to participate in betalain pigment biosynthesis in plants (Christinet et al., 2004). Despite being identified in some bacterial genomes, the function of this enzyme in bacteria is not yet known (Dalfard et al., 2006). Also, the homologous region containing the Noc_0563-0575-homologous cluster in the C-107, C-27 and AFC27 genomes is replaced in the APC132 genome by a smaller unique segment of several genes including a gene encoding a DNA polymerase III subunit epsilon-like protein and a putative transcriptional regulator.

In contrast, the AFC132 genome contained some genes that were absent from the C-107, C-27 and AFC27 genomes. For example, a gene (AFC132_2702) encoding a putative tellurium stress and resistance protein, TerD, is inserted between noc_2537 and noc_2538-homologous loci of C-107, C27 and AFC27. Another interesting gene uniquely

discovered in the AFC132 genome encodes an RNA-directed DNA polymerase (AFC132x_3183).

4.6. Discussion

4.6.1 Genomes Between *N. Oceani* Strains Are Generally Highly Conserved

Strains of widely distributed marine bacterial species such as *Prochlorococcus marinus* or *Photobacterium profundum* but not *Crocosphaera watsonii* were found to exhibit moderate to high genomic diversity (Gurevich et al., 2013, Thompson et al., 2013, Lauro et al., 2014). Being an omnipresent bacterial species in the world's oceans (Ward and O'Mullan 2002), we thus set out to examine the extent of genomic diversity within the *N. oceani* lineage by focusing on four strains in pure culture isolated from the Atlantic (C-107 and C-27) and Pacific (AFC27 and AFC132) Oceans. Phylogenetic analysis confirmed that these strains are closely related and the genome sequences obtained using a next-generation sequencing approach confirmed that the genomes of strains ATCC19707, C-27, AFC132 and AFC27 contained identical 16S rRNA sequences, two in each genome, which is in agreement with a previous study (Campbell et al., 2011a). An earlier study by Ward and O'Mullan (2002) showed a slightly different rRNA phylogenetic tree, which is likely due to the fact that these authors did not have pure cultures of all strains.

The most surprising finding of this study is the high level of genome sequence conservation between the *N. oceani* strains despite their origin from different oceanic gyres. We calculated almost identical ANI values close to 100 % for the genomes of C-107, C-27 and AFC27, and even the genome of AFC132 was ~98.6% identical to the three other genomes (Table 4.2). Such low sequence diversity at the strain level was also

reported for strains within other marine bacterial species such as the cyanobacterium *Crocosphaera watsonii* (Zehr et al., 2007, Gurevich et al., 2013), *Vibrio cyclitrophicus* (Shapiro et al., 2012) or *Alteromonas macleodii* (Lopez-Perez et al., 2013). For example, strains of *A. macleodii* exhibited high genome sequence conservation as demonstrated by two-way ANI values near 100% (i.e. 99.98 %) for any two *Alteromonas macleodii* isolates (i.e. the strains U4 and U7 from the same clonal frame); however, these were all isolated from different parts of the Mediterranean Sea (Lopez-Perez et al., 2013). In contrast, *N. oceani* strains C-27, C-107 and AFC27 were isolated from locations in the Atlantic and Pacific Oceans that have been physically separated during evolutionary times by the landmass of the Americas and their calculated two-way ANI values were also near 100% (Table 4.2), which suggests direct clonal history.

All four strains encode complete sets of nitrogen compound transformation inventory, which is important to their nitrogen compound-dependent chemolithotrophic lifestyle in the sea. Likewise, the complement of defense and stress-tolerance is well conserved despite living in different ecological settings. Furthermore, the marine environment is abundantly populated by bacteriophages with a high diversity in their specificity suggesting that marine bacteria including *N. oceani* are constantly exposed to attack by phages (Labrie et al., 2010, Suttle et al., 2010). Surprisingly, three of the four *N. oceani* strains (C-107, C-27 and AFC-27) evolved/maintained an identical genetic make-up suited for defense against foreign DNA such as their CRISPR and RM system complements. Even though the forth strain was slightly better equipped, the high level of genome identity (in sequence and gene order) suggests that their existing defense armory

was sufficient to provide effective resistance against horizontally transferred foreign DNA including phages in their respective environments.

4.6.2 Differences Between *N. Oceani* Genomes and Responsible Mechanisms

The genome of strain AFC132 differs from that of the other strains as expressed by a ~1.5% lower 2-way ANI value of 98.56% (Table 4.2). Although the AFC132 genome retained a high level of genomic synteny in comparison with the other genomes, our alignments suggest the occurrence of one genomic rearrangement within the magnesium/cobalt transporter *corA* homologous gene (Noc_1416) in AFC132 leading to the deletion of the N-terminal part of the gene.

4.6.3 Plasmids and Mobile Genetic Elements in The Genomes of *N. Oceani* Strains

The *N. oceani* strains C-107, C-27 and AFC27 but not AFC132 are postulated to carry a near-sequence identical plasmid A. This resembles the situation of *Alteromonas macleodii* whose isolates from the same clonal frame (isolated from the same clonal frame is presumed to be derived from common ancestor and have low diversity) do not necessarily contain a plasmid (Lopez-Perez et al., 2013).

Gene clusters encoding toxin-antitoxin systems were identified in all *N. oceani* genomes. Toxin-antitoxin systems are well characterized in their function to retain plasmids; however, the roles of toxin-antitoxin systems encoded in the chromosome are not well understood. There are reports suggesting that toxin-antitoxin systems in chromosomes can play various biological roles such as increasing the tolerance to antibiotics (Dorr et al., 2010), to protect against phage attack (Magnuson 2007) or they are implicated in post-transcriptional regulation (Bertram and Schuster 2014). Thus, the chromosomal toxin-antitoxin systems retained in the *N. oceani* strains may provide the

respective strain with beneficial properties under certain circumstances. As reported in the results above, C-107, C-27 and AFC27 genomes differ from the AFC132 genome also in their content of encoded toxin-antitoxin systems.

Initially, the publication of the genome of *N. oceani* C-107 genome revealed a high number of mobile elements, especially transposases (Klotz et al., 2006). In this study, the examination of other *N. oceani* strains confirmed the residence of a numerous of mobile elements in their genomes. Although, we could not assess the full complements of the mobile elements in all strains due to the incomplete closure of the genomes and the repeat nature of many of the mobile elements, we were able to examine the mobile elements identified, including unique insertions into genomes of the different strains. Genomic comparison revealed that the majority of the transposases identified in the C-107 genome were found in the other strains as well. Analysis of genome alignments also revealed that the distribution of transposase genes was most similar between C-27 and AFC27 genomes while many of their locations in the AFC132 genome were different. Nevertheless, the transposases encoded in the AFC132 genome belonged to the same families found in the genomes of the other strains. Some genes encoding transposases of the same family (e.g. IS605 OrfB) in all *N. oceani* genomes were distributed differently suggesting random insertion or jumping events. While all four genomes contained more than a few phage genes, only one intact prophage was discovered in the genome of AFC132. Given the high level of genome synteny, *N. oceani* may have been exposed to virus attack frequently but capable of preventing major genome re-arrangements and divergence. One major difference between the genomes of strains C-107 versus C-27 and AFC27 is the existence of a large genomic island in the genome of the C-107 strain.

Furthermore, we identified the existence of so called “flexible genomic islands”, which are defined by Gonzaga et al. (2012) as genomic regions that have homologous position in the genomes, have similar inferred function but contain different sets of genes. Examples of such flexible genomic islands in the studied *N. oceani* genomes are the in someone clusters encoding RM and CRISPR/Cas systems as well as flagella and polysaccharide biosynthesis genes.

4.6.4 Intracellular Mechanisms Responsible for The Stability of Diverging *N. Oceani* Genomes

The CRISPR/Cas system acts as a defense mechanism in archaea and bacteria functionally analogous to immunity systems, which give the organisms protection against invasion of previously encountered foreign genetic elements such as phages and plasmids (Barrangou et al., 2007, Wiedenheft et al., 2012). This system is widespread among archaea (about 90%) and bacteria (about 40%) (Horvath and Barrangou 2010). CRISPR loci consist of arrays of direct repeats separated by spacer DNA and are usually located close to CRISPR-associated (*cas*) genes that code for heterogeneous families of proteins (Haft et al., 2005). The length of the direct repeats can vary between 23 and 47 bp while the length of the spacer varies between 21 and 72 bp (Horvath and Barrangou 2010). The CRISPR loci are formed by the incorporation of new spacer DNA derived from the invading foreign DNA elements into the CRISPR thereby conferring adaptive immunity (Barrangou et al., 2007). Thus exposure to different foreign invading DNA agents can result in the build up of CRISPR loci that provide immunity even between closely related microbial strains (Horvath et al., 2008, Díez-Villaseñor et al., 2010, Rezzonico et al., 2011). Generally, the genomes of the three *N. oceani* strains C-107, C-27 and AFC27

harbor similar CRISPR/Cas systems. Although AFC132 contains CRISPR/Cas systems of the same subtype and exhibits the same direct repeat length, it has a slightly different direct repeat consensus sequence (Table 4.3), a much longer CRISPR length and about a six-fold higher number of spacers. Also, the genetic organization of CRISPR/Cas systems in the AFC132 genome compared with the other genomes is different.

Restriction enzymes (RE) are also employed by bacteria to defend themselves against invasive foreign DNA by recognizing particular sequence patterns of DNA and /or its methylation state and cutting the DNA into shorter segments. Host DNA with sequence patterns recognized by REs can be protected by methylation thus making them impervious to the actions of the restriction endonucleases and enable restriction enzymes to differentiate between host and foreign DNA. There are four types of restriction systems which are designated I to IV (Roberts et al., 2003). In general, the inventory of restriction modification (RM) systems encoded in the genomes of C-107, C-27 and AFC27 are identical. The majority of these RM systems are conserved in AFC132 as well but AFC132 also encodes unique RM systems: For instance, a gene cluster encoding a type I RM system in the genomes of C-107, C-27 and AFC27 (Noc_1199-1205) is replaced in the genome of AFC132 by a gene cluster encoding a type III RM system. This replacement was likely the outcome of a horizontal transfer event as these RM systems mobilize independent of accessory mobile elements (Furuta et al., 2010).

As reported above, two unique genes are found to wedge between the R (noc_2927 homolog) and M (noc_2928 homolog) genes in the AFC132 genome, which constitute a hypothetical operon in conjunction with the R gene. One of these inserted genes encodes a protein with a phage abortive infection (Abi) system domain. Abi

systems are involved in defense response against phage attack and they operate similar to a TA system, by which the toxin is neutralized by a cognate antitoxin (Fineran et al., 2009, Dy et al., 2014). In AFC132, the gene upstream of the Abi-like encoding gene is annotated as a hypothetical gene, which might code for the antitoxin part of the system.

In addition to these well-characterized mechanisms involved in defensive nucleic acid manipulations, a gene encoding a putative reverse transcriptase was identified exclusively in the AFC132 genome. The putative gene product encoded in the AFC132 genome is most sequence similar to the retron-type reverse transcriptases belonging to the group II intron family, which are considered diversity-generating retro-elements (Simon and Zimmerly 2008). This RNA-directed DNA polymerase is known to produce multi-copy single-stranded DNA (msDNA), which is a hybrid DNA-RNA molecule with extensive secondary structures (Lampson et al., 2005). While msDNA has been hypothetically implicated in epigenetic regulatory events such as social behavior of bacteria, its function in general remains elusive.

4.6.5 Extracellular Mechanisms Responsible for The Stability Of Diverging *N. Oceani* Genomes

As reported above, our genome analysis identified evidence for the replacement, either partly or entirely, of polysaccharide synthesis-related genes/ gene clusters in the genomes of AFC132 and the other strains. In *Alteromonas macleodii*, complete replacement of gene clusters implicated in synthesis of lipopolysaccharide O-chain, EPS and flagellation-related genes in different isolates was reported to create dramatic differences in the cell surfaces of different strains and thus likely provided the ability to produce target-differences for attachment by viruses (Lopez-Perez et al., 2013). In

addition to two gene clusters (Noc_1022-55 and Noc_1304-41) present in all 4 genomes, the AFC132 genome contains an additional unique gene cluster with a predicted role in terpene metabolism. While a biochemical or physiological function of terpenes in bacteria is elusive, various terpenes are reportedly produced by a number of marine bacteria (De Carvalho and Fernandes 2010).

4.6.6 Differences in the Metabolic Capacity Between the *N. oceani* Genomes

One of the prominent features with physiological significance and unique to AFC132 is the existence of a putative NRPS gene cluster encoding the capacity for siderophore production. The other *N. oceani* genomes only contained pertinent genes required for uptake and processing of the cognate siderophore but not its synthesis. This could be indication that they have lost the capacity for synthesis and evolved to be cheaters in order to streamline the genome and conserve resources. This trait has been suggested for other marine bacteria such as marine *Vibrio* species where selective loss of siderophore biosynthetic genes in conjunction with keeping of the cognate receptor genes seems to have occurred (Shapiro et al., 2012). In addition, this has been suggested for and studied in betaproteobacterial ammonia-oxidizing bacteria from fresh- and wastewater (Chain et al., 2003, Wei et al., 2006).

A second prominent feature unique to the AFC132 genome is the presence of a putative gene cluster encoding NiFe-hydrogenase. As reported in the results, the NiFe-hydrogenase gene cluster of AFC132 has the same genetic organization as the NiFe-hydrogenase gene cluster found in some bacterial species, with one of them being the nitrite oxidizing bacterium (NOB) *Nitrospina gracilis* strain 3/211, a representative of the dominant genus of nitrite-oxidizing bacteria in the marine environment (Lucker et al.,

2013). The group 3b of bidirectional (NADP)-dependent hydrogenases (i.e. group 3b) (Vignais and Billoud 2007) of *N. gracilis* is postulated to have the capacity to perform a reversible oxidation of dihydrogen gas using NAD(P)+ as the oxidant (Silva et al., 2000). This type of hydrogenase is also reported to reduce elemental polysulfide to H_2S (Ma et al., 2000). Although the function of this hydrogenase encoded in the AFC132 genome remains to be tested, there is the possibility that it provides the strain with increased catabolic flexibility by providing the ability to utilize polysulfide reduction as an alternative electron termination process.

Other AOB such as *Nitrosococcus halophilus* Nc4 (Koops et al., 1990), *Nitrosospira multiformis* ATCC 25196 (Norton et al., 2008) and *Nitrosomonas* sp. Is79 (Bollmann et al., 2013) were discovered to have genes required for NiFe-hydrogenase biosynthesis as well. Like *N. oceani*, *N. halophilus* Nc4 is also marine bacterium while the betaproteobacteria *N. multiformis* ATCC 25196 and *Nitrosomonas* sp. Is79 have been isolated from soil and freshwater sediments, respectively. Their putative NiFe-hydrogenases are somewhat different from the one in proposed for AFC132 with the *N. halophilus* Nc4 genome encoding a large hydrogenase subunit containing the F420-reducing domain, while both *N. multiformis* ATCC 25196 and *Nitrosomonas* sp. Is79 genomes encode group 3d hydrogenases. The diversity in hydrogenases and their overall scarcity in AOB suggest that the NiFe-hydrogenase gene cluster in AFC132 was acquired horizontally.

The arrangement of genes in the NiFe-hydrogenase gene cluster of AFC132 is found in bacteria predominantly inhabiting alkaline or saline environments including *Nitrospina gracilis* strain 3/211. Strains in the nitrite oxidizer *Nitrospina gracilis* are the

most likely partners to *N. oceani* in the marine nitrification process and, like with *N. oceani*, not all strains are necessarily hydrogenase-positive (Lucker et al., 2013). It is thus possible that hydrogenase-positive *Nitrospina* strains cooperate with hydrogenase-negative *Nitrosococcus* strains or vice versa, thereby establishing "cross-feeding" scenarios analogous to the ones proposed for pairs of ammonia- and nitrite-oxidizing microbes complementing each other for the catabolic capacity to utilize hydrogen, urea or cyanate (Koch et al., 2014, Koch et al., 2015, Palatinszky et al., 2015). On the other hand, there is a report for the requirement of hydrogenase in *Synechocysti* ssp. PCC 6803 to thrive in mixotrophic, nitrate-limiting conditions (Gutekunst et al., 2014).

Interestingly, not all bacterial genomes with the AFC132-type NiFe-hydrogenase genetic arrangement carry a *hypB* gene. The HypB protein, in concert with HypA, facilitates the maturation process of NiFe-hydrogenase by participation in nickel insertion (Xia et al., 2012). This protein has GTPase activity and possesses a CHxnC motif, which binds to either nickel or zinc (Leach et al., 2005, Chan et al., 2012). Of the various species discovered to have this NiFe-hydrogenase genetic arrangement in our study, only the two *Rhodothermus marinus* strains have *hyp*B genes that are located further upstream of the NiFe-hydrogenase-encoding gene cluster. Analysis software Operon DB (Pertea et al., 2009) predicted that the *hyp*B gene and the gene cluster encoding NiFe-hydrogenase are located in independent operons in *Rhodothermus marinus* DSM 4252. Given the absence of *hyp*B gene in some of these species including *N. oceani* AFC132, it is perhaps that HypB is not critical for the maturation of NiFe-hydrogenase. Several studies have shown that hydrogenase activities of *hyp*B mutants can be partially restored by increased nickel concentration supplementation (Olson et al., 2001, Mehta et al., 2003). Perhaps,

nickel insertion into NiFe-hydrogenase can be facilitated by some other enzyme already found in the HypB-negative organism. Homologs of HypB involved in the maturation of other Nickel enzymes are known: for instance, UreG participates in the maturation of urease in a similar way as does the accessory protein CooC for the incorporation of nickel into CO dehydrogenase (Jeon et al., 2001, Hube et al., 2002, Boer et al., 2010). In Helicobacter pylori, HypB contributes to nickel incorporation into urease and a *hyp*B mutant exhibited lower urease activity due to deficiency of nickel in the urease (Olson et al., 2001). Another possibility is that an unknown gene product performs the role of HypB in species lacking the gene encoding typical HypB. Because *N. oceani* AFC132 is urease positive, it is most likely that its UreG protein plays a role in the maturation of the putative NiFe-hydrogenase.

When compared to the other strains, the AFC132 also lacks coding sequences that are present in the other three *N. oceani* genomes. While all 4 genomes contain a gene cluster encoding a phosphate transport system (noc_2394-2401 homologs), only the AFC132 genome misses a gene cluster encoding a phosphate transport function. This finding might mirror the reported genetic difference between *Prochlorococcus* and *Pelagibacter* strains isolated from the P-richer Atlantic and the Pacific Oceans, demonstrating that the genomes of Atlantic isolates were better equipped with phosphate acquisition systems compared to their Pacific relatives (Coleman and Chisholm 2010).

4.6.7 Implications of Genome Differences in the Evolution of *N. oceani*

There is likely no single one mechanism that determines the holophyletic evolution of bacteria. For example, recombination has been implicated as the main driver of genome divergence in *Vibrio cyclitrophicus* (Shapiro et al., 2012) and *Alteromonas*

macleodii (Lopez-Perez et al., 2013) whereas the observed high diversity among *Photobacterium profundum* strains of different bathytypes was attributed, to a large part, to horizontal gene transfers (Lauro et al., 2014). On the other hand, *Crocosphaera watsonii* likely diversified by genome degradation, which resulted in the delineation into small-cell and large-cell phenotypes (Gurevich et al., 2013).

The high level of content identity and synteny of the *N. oceani* genomes suggests that they have undergone only minimal divergence either after recent distribution to their current geographical locations from the same point of origin or, because this former scenario seems unlikely, the genomes encode effective mechanisms that ensured recalcitrance through evolutionary times. AOB in pure culture never grow beyond threshold densities and their abundance in the natural marine environment is very low (Ward and O'Mullan 2002, Campbell et al., 2011a). Zehr et al. (2007) suggested that the low genetic diversity observed in bacteria that maintain low abundance such as *Crocosphaera. watsonii*, maybe a necessary evolutionary strategy adopted by these bacteria to maintain their species identity while more abundant bacteria may derive succession benefits from a higher genomic diversity between the populations in the species (Dufresne et al., 2005). Both scenarios, increasing diversity or high conservation, can be the outcome genomic rearrangement (Gurevich et al., 2013) if coupled with pertinent epigenetic mechanisms. Former studies of the catabolic inventory in betaproteobacterial ammonia-oxidizers postulated the operation of rectification mechanisms that ensured near sequence identity between multiple copies of large segments within given AOB cells and a significantly high sequence identity between different populations of strains from different species (Arp et al., 2007); however,

genome divergence between different AOB populations of the same species as reported in tis study, have never been attempted. Such rectification mechanisms are likely also at work in the genomes of *Nitrosococcus* genomes, all of which have two copies of ribosomal RNA clusters but only one copy of gene clusters encoding key catabolic functions (Klotz et al., 2006, Campbell et al., 2011b, Simon and Klotz 2013, Wang et al., 2016) The genomes of *N. oceani* strains C-107, C-27 and AFC27 as well as AFC132 have an approximate size of 3.5 Mb and a GC content of approximately 50% without displaying a skew for a slightly larger size, a higher number of genes and a lower GC value (C-27, AFC132) or a smaller size, a lower number of genes and a higher GC value (C-107, AFC27) that correlates with the location of their isolation (Atlantic: C-107, C-27; Pacific: AFC27, AFC132) (Table 4.1). The draft genomes are larger in size than the closed genome of C-107 except for AFC27, which is potentially due to the applied sequencing technology; however and interestingly, there appears to be a slight skew towards a lower number of CDS in the strains isolated from Pacific Ocean locations with the genome of AFC132 encoding the lowest number of CDS while being the largest genome in size. While the genomes of C-27 and AFC27 differ in size, GC content and the number of genes, they appear to encode an identical number of CDS, smaller than the number of CDS in C-107, which correlates with the finding that both genomes are high in synteny and identity (Fig. 4. 2, Table 4.2).

The greater deviation of the AFC132 genome from the other 3 genomes correlates with a markedly lower number of CDS, the lack of a plasmid and unique deletions and insertions. Many of the locations that indicate deletions and insertions in the AFC132 genome are associated with the residence of transposases, which facilitate genome

rearrangements but are also involved in horizontal transfer of free foreign or phage DNA. The plasmids A of C-107, C-27 and AFC27 contain mostly hypothetical genes, mobile elements and genes encoding restriction modification systems. Such largely cryptic plasmids may have also played a role in the process of genome evolution, adaptation and speciation by shuttling genes among the strains (Arp et al., 2007). When compared to AFC132, the other *N. oceani* genomes appear to have undergone a process of genome streamlining by getting smaller in size, shedding genes with the capacity to provide greater metabolic capacity (siderophores, NiFe-hydrogenase) and by increasing the coding density of the genome (Table 4.1). Most interesting in this context is that this putative genome economization correlates with a leaner CRISPR/Cas complement when compared with the AFC132 genome. This trajectory of genome economization in marine bacterioplankton has been pointed out before by Swan et al. (2013).

The phylogenetic tree includes all *N. oceani* strains in one monophyletic clade that shows them most closely related to *Nitrosococcus watsonii* C-113 (Fig. 4.4). In the context of genome content and its economization, it appears most parsimonious that the AFC132 genome closest related to the ancestral lineage of the species and that C-107 represents the most delineated genome with the C-27 and AFC27 genomes in between. Former studies, that delegated the evolutionary roots of molecular inventory required for nitrogen-based chemolithotrophy to sulfur-dependent catabolism (Simon and Klotz 2013; and references therein) and the fact that the genus *Nitrosococcus* belongs to the family purple sulfur bacteria suggests that the genomes of ancestral *Nitrosococcus* species likely encoded a variety of inventory conferring metabolic potential that is now by and large absent from these obligate ammonia-catabolic chemolithotrophs. Hence the fact that the

AFC132 genome still retains the capacity to express NiFe-hydrogenase and an NRPS-encoding gene cluster likely capable of siderophore biosynthesis, and that it contains a larger complement of defense inventory lost from the other genomes supports the notion of its ancestral status. Genome economization in context with reduced metabolic versatility such as the loss of siderophore production capacity while maintaining the ability for uptake and processing of the cognate siderophore produced other microorganisms (the evolution of cheaters) have been also reported for betaproteobacterial AOB (Chain et al., 2003). A search in existing databases did not reveal the origin of plasmid A in the genomes of C-107, C-27 and AFC27; nevertheless, its stacking with two putative Toxin-antitoxin systems (a HicB-HicA TAsystem at Noc_A0004-Noc_A0003, and TA system at Noc_A0034-Noc_A0033) indicate the stable incorporation of the plasmid in these genomes, which makes it more likely that the plasmid was acquired by the genome ancestral to these three strains than the possibility that the AFC132 genome lost it. Further, the incorporation of the large genomic island in the C-107 genome a conserved location encoding a transposase in genomes of all strains suggests that the acquisition of the island occurred after delineation of the C-107 genome from the AFC27 and C-27 genomes.

4.7 Summary

In conclusion, analysis of 4 genomes representing strains isolated from different oceanic gyres (AFC27: North Pacific; AFC132: South Pacific; C-27: coastal waters of the Southern North Atlantic; C-107=ATCC 19707=JMC30415: open North Atlantic) revealed extremely low diversity with regard to sequence identity, content and synteny. Detected differences in genome properties by and large did not disclose a correlation with

the location of isolation; nevertheless, strain AFC132 isolated in the South Pacific, which is reported to be less concentrated in oxidized fixed nitrogen, is proposed to be closest related to the ancestor of the species whereas C-107 appears to be the youngest evolutionarily. Given that there are many isolates with near identical sequences of select genes to strain C-107 (Ward and O'Mullan 2002, Campbell et al., 2011a), their distribution in the oceans worldwide makes sense in light of the economization of their genomes. In contrast, strains of further delineated species in the genus *Nitrosococcus* have been isolated in distinct locations characterized by high salt concentration (*N. halophilus*; Campbell et al., 2011a) and organics-polluted waters (*N. wardiae*; Wang et al., 2016). This distinction also reflects itself in the fact that all strains of *N. oceani* (but not *N. halophilus* and *N. wardiae*) are ureolytic and urea uptake positive (Koper et al., 2004, Klotz et al., 2006), which likely reflects an adaptation to the increasingly urea-generating biogenic activities in the Holocene. While it cannot be excluded that one of the strains was transplanted to the different oceanic location, either naturally or inadvertently by human activities, the characteristics illustrated in Table 4.2 and Fig. 4.2 suggest that the results reflect a general process of genome adaptation in bacteria whose ancestor evolved within a genomic framework that mandates low abundance (basically autochthonous). While this hypothesis has been reported numerous times as cited above, additional research is needed to find causal links between specific genomic inventory and trajectories of genome evolution.

Table 4.1: General genome characteristics of *Nitrosococcus oceani* strains

Strains	Chromosomal genome size	Chr. GC %	No. of plasmid	rRNA	tRNA	Total genes	CDS
C-107	3481691 (complete)	50.32	1	6	46	3265	3149
C-27	3539918 (541 contigs)	49.94	1	6	47	3485	3076
AFC27	3471807 (49 contigs)	50.30	1	5	44	3224	3076
AFC132	3545101 (630 contigs)	49.83	Not detected	3	46	3427	2937

Table 4.2: Two-way average nucleotide identity (ANI) analysis

	C107	C27	AFC27	AFC132
C107	--	99.99%	99.98%	98.56%
C27	99.99%	--	99.99%	98.56%
AFC27	99.98%	99.99%	--	98.58%
AFC132	98.56%	98.56%	98.58%	--

Table 4.3: Comparison of the CRISPR between the *Nitrosococcus oceani* strains

	CRISPR length	Direct repeat	DR consensus sequence*	Number of spacers
C-107	387	28	GTTCAC**C**GCCGCACAGGC**G**G**T**TTAGAAA	6
C-27	387	28	GTTCAC**C**GCCGCACAGGC**G**G**T**TTAGAAA	6
AFC27	387	28	GTTCAC**C**GCCGCACAGGC**G**G**T**TTAGAAA	6
AFC132	2248	28	GTTCAC**T**GCCGCACAGGC**A**G**C**TTAGAAA	37

*letters in bold denote difference in nucleotide base

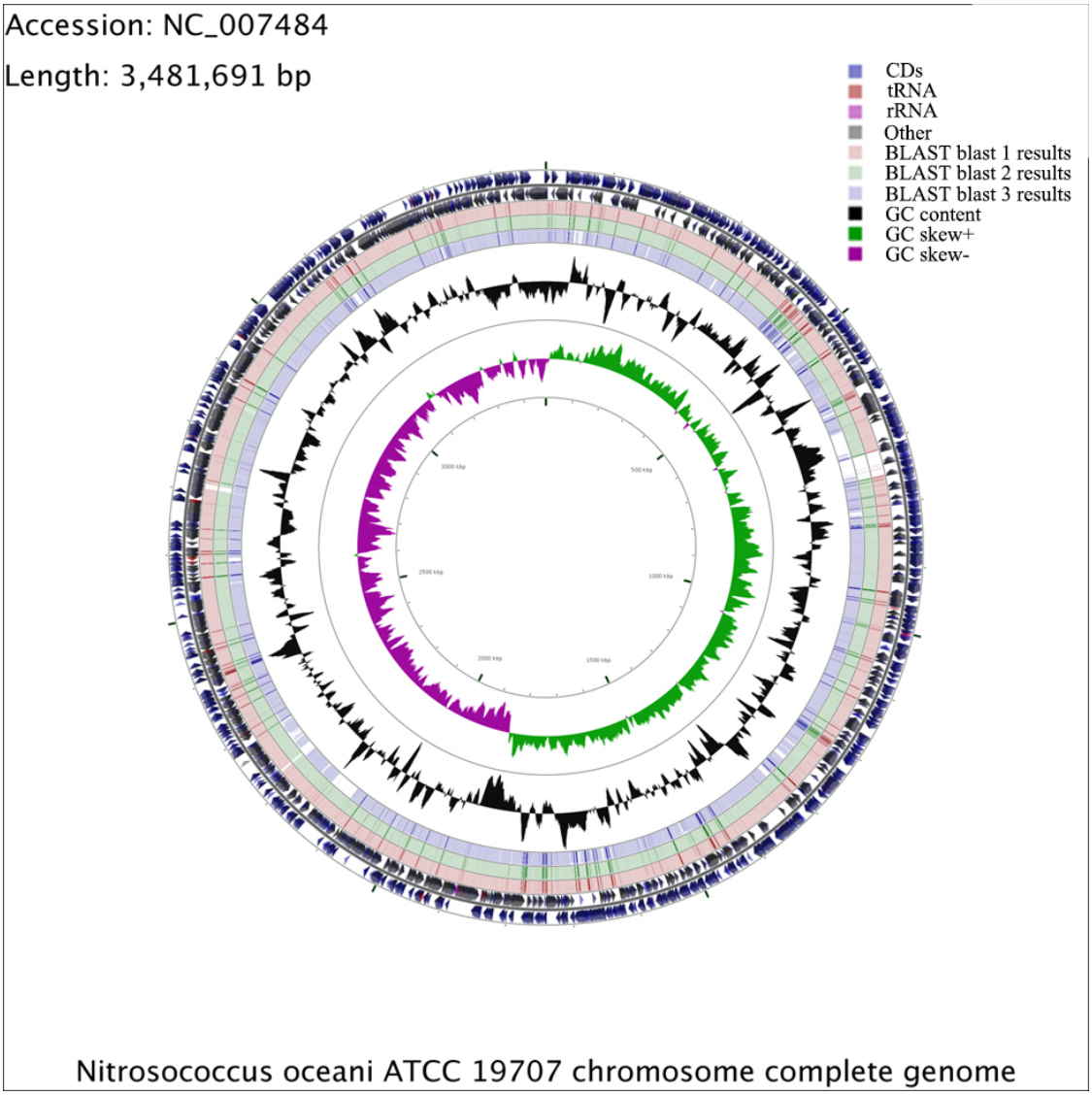
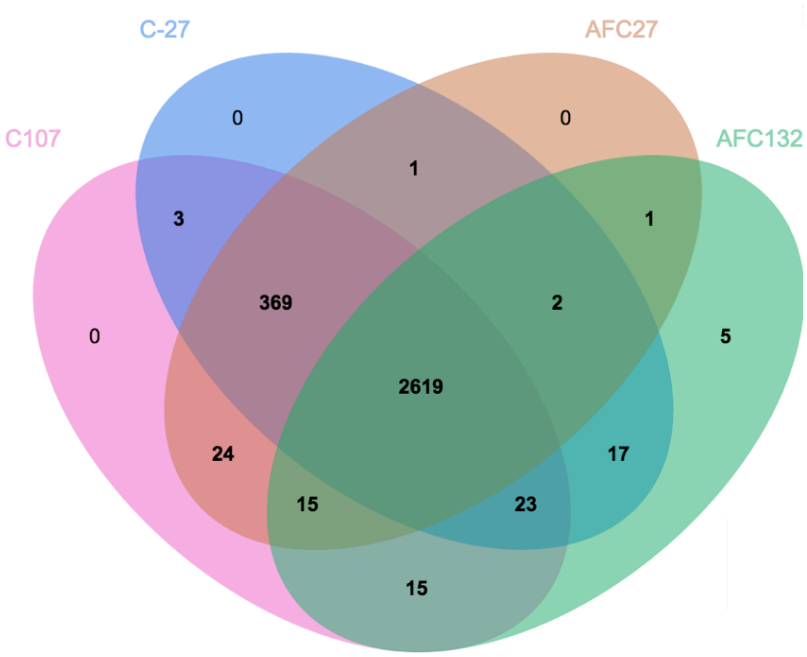
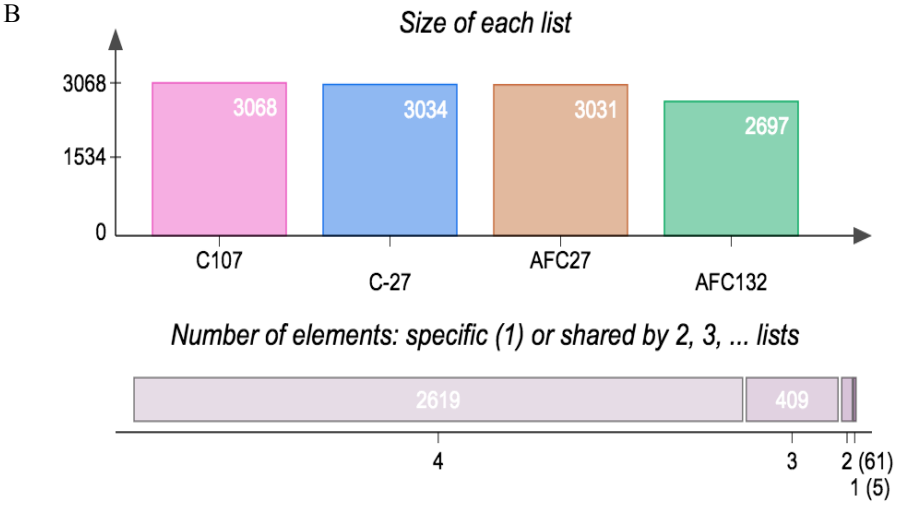


Figure 4.1: Circular view of the *N. oceani* C-107 genome genome generated using BLAST alignment using CGViewer software. Circles correspond to the following features, starting with the outermost circle: (1) forward strand genes (2) reverse strand genes (3) BLAST 1 result (Chapman et al., 2002) BLAST 2 result (mint green) (5) BLAST 3 result (blue) (6) GC content (7) GC skew+ (green) (8) GC skew- (magenta).

A


C-27
AFC27
C107
AFC132
0
0
1
3
1
369
2
0
5
2619
24
17
15
23
15

B


Size of each list
3068
1534
0
3068
3034
3031
2697
C107
C-27
AFC27
AFC132
Number of elements: specific (1) or shared by 2, 3, ... lists
2619
409
4
3
2 (61)
1 (5)

C

Summary

Species	Proteins	Clusters	Singletons
C107	3121	3068	43
C-27	3076	3034	40
AFC27	3052	3031	16
AFC132	2937	2697	233

Figure 4.2: VENN Diagram drawn using the OrthoVenn web server (Wang et al., 2015) showing the genome comparison of four *Nitrosococcus oceani* strains. The core genome of the *Nitrosococcus oceani* species consists of 2619 CDS. The species form 3094 clusters, 3089 orthologous clusters (at least contains two species) and 2612 single-copy gene clusters.

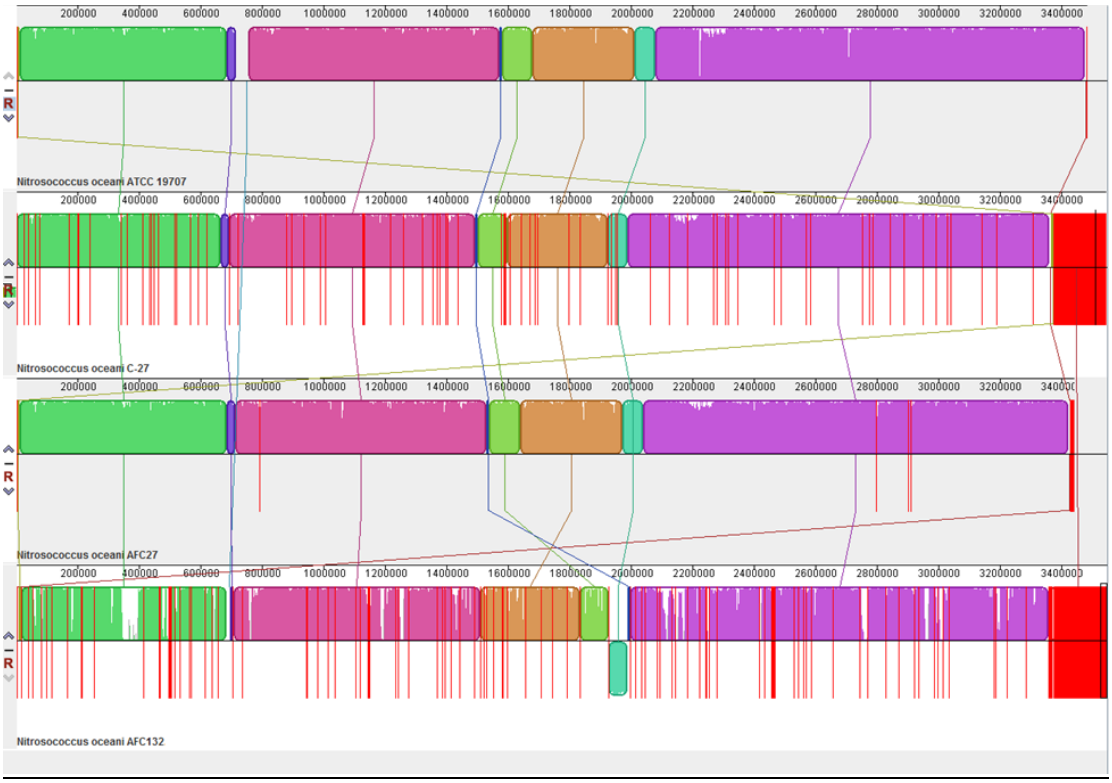


Figure 4.3: Genomic alignment of the four compared *Nitrosococcus oceani* strains using Mauve 2.3.1 software (Darling et al., 2004).

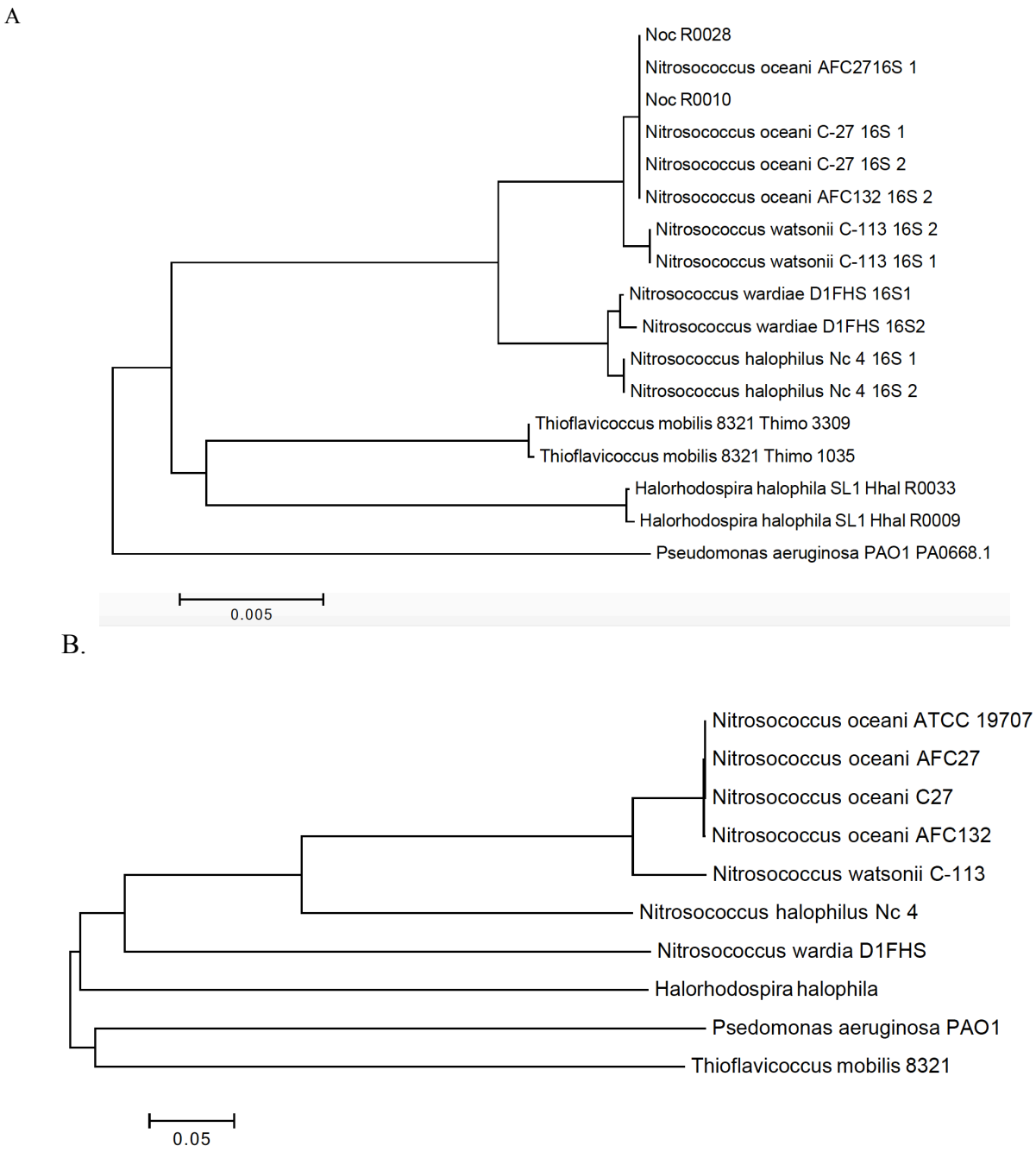


Figure 4.4: (A) 16S rRNA gene and (B) gene gyrB-recA concatenated gene phylogeny: Trees were constructed using the MEGA6 software package after subjecting the alignments to phylogenetic inference with the Neighour-joning algorhithm. Phylogenetic relationships between the *Nitrosococcus oceani* strains were determined based on 16S rRNA gene sequence alignments as well as alignments of concatenated sequences of the housekeeping genes of *gyrB* and *recA*.

CHAPTER 5: CHARACTERIZATION OF MOLECULAR INVENTORY INVOLVED IN NITROGEN COMPOUND TRANSFORMATIONS IN OTHER BACTERIA

5.1 Molecular Basis of Obligate Chemolithotrophy in AOB

5.1.1 Introduction of Hypothesis and Aims

The main goal for this project is to study the molecular basis of obligate chemolithotrophy in ammonia-oxidizing bacteria. Specifically, it is to determine how electrons extracted from HAO enter the quinone pool from the periplasm and how involved inventory interacts with each other in the quinone-reducing branch by functional complementation experiments *in vivo* in methane-oxidizing bacteria (MOB). MOB facilitate a wide array of N transformations including those that serve its assimilation such as the reduction of dinitrogen or NOx to ammonium. When supplied with ammonium under fully oxic conditions, none of transformations of less reduced N compounds are essential to the survival and growth of MOB and it is therefore possible to modify these pathways *in vivo*. This includes also the reactions that enable some methanotrophs to nitrify. As described above, nitrifying MOB cannot make a living on the oxidation of ammonia due to the inability to transfer electrons extracted from hydroxylamine by HAO in the periplasm into the quinone pool. The genomes from nitrifying methanotrophs (i.e., *M. capsulatus* strain Bath and *M. infernorum* strain V4), reveal that they do not include genes encoding homologues of cytochrome c554 or a member in the c_M552/NrfH/NapC/NirT protein superfamily that exchange electrons with

quinone pool (Ward et al., 2004, Simon and Klotz 2013). Methanotroph catabolism provides for reduction of the quinone pool from the cytoplasm; therefore, introduction of the ability to reduce quinone from the periplasm could create a facultative trait if the knock-in of pertinent candidate genes provides for a functional HURM (see Chapter 1). The two organisms used in this experiment are *Nitrosococcus oceani* strain C107 and *Methylococcus capsulatus* strain Bath. *Nitrosococcus oceani* is the first obligate ammonia-oxidizer isolated from seawater. Its genome reveals that the genome of *Nitrosococcus oceani* encodes only one copy each of *amo*CAB and *hao*AB-*cyc*AB gene clusters (Klotz et al., 2006, Stein et al., 2013). *Nitrosococcus oceani* is one of the model systems to study obligate chemolithotrophic ammonia-oxidizers. The non-chemolithotrophic ammonia oxidizing bacterium, *M. capsulatus* strain Bath, also possesses a Cu-MMO (pMMO) and an active HAO to facilitate the oxidation of ammonia to nitrite via a hydroxylamine intermediate. The genome sequence of *Methylococcus capsulatus* strain Bath growing in our laboratory in pure culture has been published (Ward et al., 2004) and many of its physiological characteristics have been identified. Fortuitously, *M. capsulatus* is presently the best-studied methanotroph in pure culture and experiments described in the literature indicate that transformation of *M. capsulatus* is realistic.

The experimental strategy pursued in this chapter will utilize molecular genetic and physiological methods to test the hypothesis that the inability of *Methylococcus capsulatus* strain Bath to support growth by ammonia oxidation is simply due to their inability to reduce the quinone pool with electrons harvested during ammonia oxidation. This hypothesis (Stein and Klotz 2011) can be tested by providing *Methylococcus*

capsulatus strain Bath with pertinent molecular inventory from ammonia-oxidizing bacteria in a gene knock-in scenario. Such scenario will also allow to test whether the transfer of electrons from hydroxylamine to the quinone pool will require inventory encoded by one or two individual genes. Therefore, the experiments will set out to test whether knocking-in *cyc*A (c554), *cyc*B (c_M552) or *cyc*AB genes will generate new phenotypes of *M. capsulatus* capable of utilizing energy and reductant resulting from ammonia oxidation, which can be tested initially physiologically and later biochemically (if needed).

Based on the finding that NapC, a quinone-reactive protein homologous to c_M552 (Bergmann et al., 2005, Simon and Klotz 2013), redox-interacts directly with reversely operating hydroxylamine dehydrogenase (*haoA*) to enable the reduction of nitrite to hydroxylamine and ammonium in *Nautilia profundicula* (Hanson et al., 2013), we hypothesize that (A) cytochrome c_M552 (*cycB*) can directly redox-interact with HAO and is thus necessary and sufficient for the transfer of electrons from hydroxylamine (extracted by HAO) to the quinone pool. If so, then cytochrome *c*554 (*cycA*), which is known to interact with HAO (Arciero et al., 1991b), likely transfers electrons to other electron acceptors in the periplasm not interacting with the quinone pool. Before the finding of reversely operating HAO, it was a paradigm (Hooper 1997) that both cytochromes, *c*554 (*cycA*) and c_M552 (*cycB*) are required to transfer electrons from hydroxylamine (extracted by HAO) to the quinone pool, which is our alternative hypothesis (B). This situation has been recently discussed (Simon and Klotz 2013). The objectives to test the hypotheses are: a) build constructs allowing expression of *cycA*, *cycB* and *cycAB* genes in *Methylococcus capsulatus*; b) transform *M. capsulatus* with the

built constructs; c) express recombinant genes in *M. capsulatus*; d) physiologically characterize the wild-type and recombinant *M. capsulatus* strains.

5.1.2 Material and Methods

5.1.2.1 Cloning and Expression Constructs

Genomic DNA was extracted from *Nitrosococcus oceani* strain C-107 and used as template for generating *cyc*AB, *cyc*B and *cyc*A amplicons using the polymerase chain reaction and primers (Appendix A) with RE sites useful for directional cloning: MfeI (RE1) and ClaI (RE2) (New England Biolab, USA) (Fig.5.1.1). In addition, the *haoA* promoter from *M. capsulatus* Bath WT was amplified with primers containing MfeI (RE1) and XbaI (RE3) (New England Biolab, USA) cut sites at their 5'-ends (Fig.5.1.1). The amplified genes and *hao* promoter were ligated to pCR2.1-TOPO TA vector (Invitrogen, USA) (Fig.5.1.1) thereby yielding cloning constructs pCAB1, 2, and 3, which were used to transform TOP10 *E. coli* competent cells (Invitrogen, USA) for propagation and maintenance following the manufacture's instruction. Enough cloning constructs were extracted from overnight transformed TOP10 cells to prepare for building expression constructs. The remainder transformed TOP10 cells were collected and stored at -80 °C with 25% glycerol. Collected cloning constructs, pCAB1, 2, 3 were digested with ClaI (RE2) and XbaI (RE3) following the manufacture's protocol. Fragments "*hao* promoter-*cyc*AB", "*hao* promoter- *cyc*A" and "*hao* promoter- *cyc*B" were collected after purification on 2% low-melt agarose gel. Plasmid vector, pBBR1MCS-2 (Kovach et al., 1995, Lynch and Gill 2006), digested with the same REs creating compatible overhangs were gel-purified (1.5% low-melt agarose gel) and ligated together with the above fragments. The ligation products were independently used to

transform *E. coli* TOP10 competent cells to generate maintainable expression vectors pBAB1, 2 and 3, respectively (Fig.5.1.1).

5.1.2.2 Tri-parental Mating and Construction of *M. capsulatus* Transformants

Expression constructs collected from TOP10 *E. coli* cells (donor strain) were used to transform conjugative strain *E. coli* S17-1 (helper strain) (Phornphisutthimas et al., 2007) with *Methylococcus capsulatus* Bath as the recipient strain. *Escherichia coli* S17-1 carrying the expression constructs and wild-type *M. capsulatus* were incubated individually on a shaker (Excella E25, Eppendorf, Germany) in LB or NMS media at 37°C, 225 rpm to generate sufficient biomass. Cell pellets were harvested when the cell densities reached $5x10^8$ cells ml^{-1}. Two ml of S17-1 culture and 30ml of recipient culture were mixed and centrifuged at 13,000 g, 4°C for 10 mins. Cell pellets were re-suspended with 2 ml NMS media and centrifuged again at 13,000 x *g*, 4°C for 3mins. The supernatant was discarded and the pellet resuspend in 200µl NMS media. The mixed culture was spotted on NMS agar plates (50ul/spot) and incubated at room temperature for 30 mins to allow the liquid to diffuse into the agar. Plates were then collected and incubated at 37°C upside down overnight. Bacterial lawns from individual spots were scraped off the agar with a sterile inoculation loop, resuspended in 1 mL of NMS and plated on NMS agar containing 25µg/ml kanamycin (to select for pBBR1MSC-2). Plates were incubated at 37°C for 10 days in sealed jars with the addition of 30% CH_4 (CH_4: air ratio is 3:7). Colonies able to grow on the plate will be visible after 10 days incubation (Welander and Summons 2012).

5.1.2.3 Screening and Confirmation of the Recombinant *M. capsulatus* Transformant Genotype

The kanamycin-resistant colonies were picked up and streaked onto fresh NMS plus 25μg/ml kanamycin agar plates for cultivation. Regrown colonies were checked for the presence of target genes using colony PCR (Fig. 5.1.3) with primers designed to target the plasmid and the insert gene. For propagation, target-positive colonies were transferred individually into liquid NMS medium plus 25μg/ml kanamycin supplemented with 30% CH_4 initially in a 250 ml Erlenmeyer flasks sealed with rubber stoppers and incubated at 37°C on a rotary shaker (225rpm). Growth and nitrite production by transformant cultures were monitored routinely by sampling using a needle fitted syringe (3 mL). To confirm that transformants carried the desired constructs, plasmids extracted from the liquid culture were used as a template for PCR and resulting amplicons were investigated using restriction analysis (Fig. 5.4), and sequencing with flanking primer sets.

5.1.2.4 Physiological Experiments with Wild-type and Transformant Strains of *M. capsulatus*

One ml of *M. capsulatus* Bath WT or transformant strains were inoculated into 50 ml of the following media plus 25μg/ml kanamycin in a 125ml Erlenmeyer flask sealed with rubber topper if needed: AMS containing 12.5 mM $(NH_4)_2SO_4$ (NH_4^+-only); AMS containing 30% CH_4 (AMS+CH_4); NMS containing 30% CH_4 (CH_4-only); NMS containing 12.5 mM $(NH_4)_2SO_4$ (NMS+ NH_4^+); NMS containing both 30% CH_4 and 12.5 mM $(NH_4)_2SO_4$ (NMS+CH_4+ NH_4^+); NMS containing 0.1 mM NH_2OH (NMS+NH_2OH) and NMS containing 0.1 mM NH_2OH plus 30% CH_4 (NMS+NH_2OH+ CH_4).

Nitrosococcus oceani C-107 was inoculated into the above six media (without kanamycin) as the control. All inoculations were performed in triplicate. Nitrite concentration and cell density of each culture were periodically measured using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific, USA).

5.1.2.5 Analyze the *in vivo* mRNA Levels of *cyc*AB, *cyc*B and *cyc*A, in *M. capsulatus*

To examine the expression of genes *cyc*AB, *cyc*B and *cyc*A in transformant strains of *M. capsulatus* grown in media containing NH_4^+ and/or CH_4, steady-state mRNAs were determined using real-time quantitative PCR (RT-qPCR). Total RNA was extracted from WT and transformant strains of *M. capsulatus* grown on the above NH_4^+-only and CH_4-only media using the Qiagen RNeasy mini Kit (Qiagen, Germany) following the manufacturer's recommended protocol (see details in Chapter 3). RNase-free DNase treated RNA samples were purified and reverse transcribed into single-strand cDNA with QuantiTect Reverse Transcription Kit (Qiagen, Germany) (see details in Chapter 3). Quantitative PCR was performed with cDNA samples (WT- NH_4^+; WT-CH_4; Transformant- NH_4^+ and Transformant-CH_4) in 10 µl reactions in triplicate. Each reaction consists of 5 µl SYBR Green Supermix (Bio-Rad, USA), 500ng forward and reverse primers, 40 ng cDNA template and sterile dH_2O. The reactions were monitored using a real-time thermal cycler (Bio-Rad, USA) with the following parameters: holding at 95 °C for 5 min, 40 cycles of denaturing at 95 °C for 3 s, annealing at 55 °C for 30 s and extending at 72 °C for 15 s. A melting curve was generated over a temperature range from 70 °C to 92 °C following PCR. Primers were specifically designed for hybridization with the *cycA* (q*cyc*A) and *cycB* (q*cyc*B) genes, and the 16S rRNA gene was targeted as

the internal reference. RT-qPCR data were then analyzed using the 2(-delta delta *Ct*) method (Livak 2001).

5.1.3 Results and Discussion

5.1.3.1 Confirmation of Cloning and Expression Constructs

Cloning and expression constructs were successfully built (Table 5.1.1) and confirmed by Sanger sequencing, PCR and gel-electrophoresis (Fig. 5.1.2). Plasmids pBAB1, pBAB2 and pBAB3 were extracted from TOP10 *E. coli* transformants and dissolved in 40 μl dH_2O for confirmation with PCR and for DNA sequencing. TOP10 *E. coli* cultures carrying these constructs were maintained and transferred regularly in liquid LB media and on LB agar at 4°C. Stock cultures were stored at -80°C in 25% glycerol for future use.

5.1.3.2 *Methylococcus capsulatus* Transformant Strain Mcap-pBAB1

Construct pBAB1 was successfully transformed into wild-type *M. capsulatus* Bath creating transformant strain Mcap-pBAB1, confirmed by colony PCR (Fig. 5.1.3), restriction analysis (Fig. 5.1.4) and Sanger sequencing (Appendix B) of the plasmid extracted from Mcap-pBAB1. Plasmid vector pBBR1MCS-2 does not have MfeI cutting site but the pBAB1 plasmid isolated from Mcap-pBAB1 was cut by MfeI (Fig. 5.1.1). Sanger sequencing revealed that the insert DNA sequence of the plasmid extracted from Mcap-pBAB1 was identical to the insert in pBAB1.

5.1.3.3 Physiological Assays of Mcap-pBAB1

Nitrite production and cell growth were observed for strain Mcap-pBAB1 inoculated into CH_4-only and AMS+CH_4 media (Fig. 5.1.5; however, Mcap-pBAB1 did not grow on any of the other media (data not shown) indicating that Mcap-pBAB1 could

not survive based on the oxidation of ammonia. In addition, that Mcap-pBAB1 could not tolerate high concentrations of hydroxylamine.

Transformant strain that Mcap-pBAB1 was expected to sustain its metabolism if not to grow on media without CH_4 but with NH_4^+ to support ammonium-based chemolithotrophy. However, physiological data demonstrated that the transformation of pBAB1 into *M. capsulatus* did not change its catabolic phenotype.

There are several explanations for the obtained result: 1) The *cycAB* genes are not transcribed in *M. capsulatus* host cells. This can be investigated using a reporter gene such as "*gfp*" encoding the green fluorescence protein, which would have to be placed downstream of the *cycAB* genes. In addition, RT-qPCR is a useful technique to measure steady-state transcript levels of target genes. 2) The pBAB1 insert is transcribed; however, the transcript(s) is (are) either not properly translated, the expression product(s) is (are) not properly folded of the properly folded products are not properly targeted and thus not functional in Mcap-pBAB1. The expression product of the *cycB* gene (c_M552) is a membrane protein that needs to incorporate in the membrane while exposing a significant portion of the protein to the periplasm (Kim et al., 2008). The expression product of the *cycA* gene (*c*554) is a soluble periplasmic redox-active protein (Iverson et al., 2001). Because neither gene product is a catalytic protein, enzyme activity assays are not feasible. Therefore, proteomics studies or at a minimum polyacrylamide gel electrophoresis of periplasmic fluids from Mcap-pBAB1 will be needed to verify the presence of both expression products. 3) Functional cytochromes *c*554 and c_M552 are both necessary but potentially not sufficient to direct electrons to quinone pool. If this is the case, we need to reexamine the list of catabolic inventory that is induced by

ammonium and/or hydroxylamine (Stein et al., 2013) to identify and test (by cloning from *N. oceani* and co-expression in *M. capsulatus*) if any of them is a candidate for participation in Hydroxylamine-Ubiquinone-redox module (HURM; Klotz and Stein 2008, Stein et al., 2013, Stein and Klotz 2016, Kozlowski et al., 2016).

Table 5.1.1: Constructs successfully built and sequenced

Cloning construct	Expression construct	Insertion
pCAB1: cycAB-pCR2.1	pBAB1: pBBR1MCS-2::cycAB	*M. capsulatus haoAB* promoter; *N. oceani cyc*AB
pCAB2: cycB-pCR2.1	pBAB2: pBBR1MCS-2::cycB	*M. capsulatus haoAB*; *N. oceani cyc*B
pCAB3: cycA-pCR2.1	pBAB3: pBBR1MCS-2::cycA	*M. capsulatus haoAB*; *N. oceani cyc*A

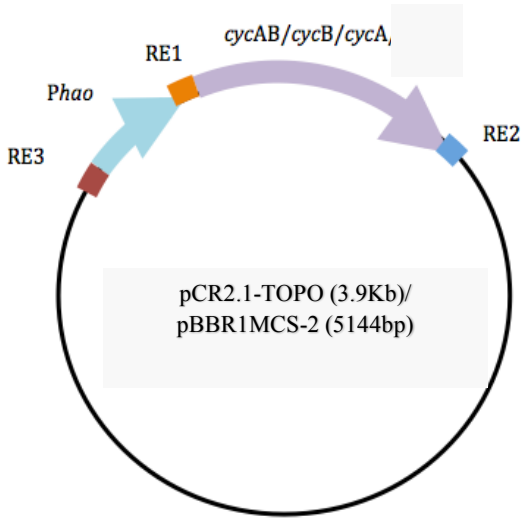


Figure 5.1.1: Map of cloning and expression constructs. The three restriction endonucleases are MfeI (RE1), ClaI (RE2) and XbaI (RE3). Insertion genes are cycAB, cycA or cycB. Selective marker genes for pCR2.1-TOPO are Amp^R and Kam^R. Slective marker genes for pBBR1MSC-2 is Kam^R.

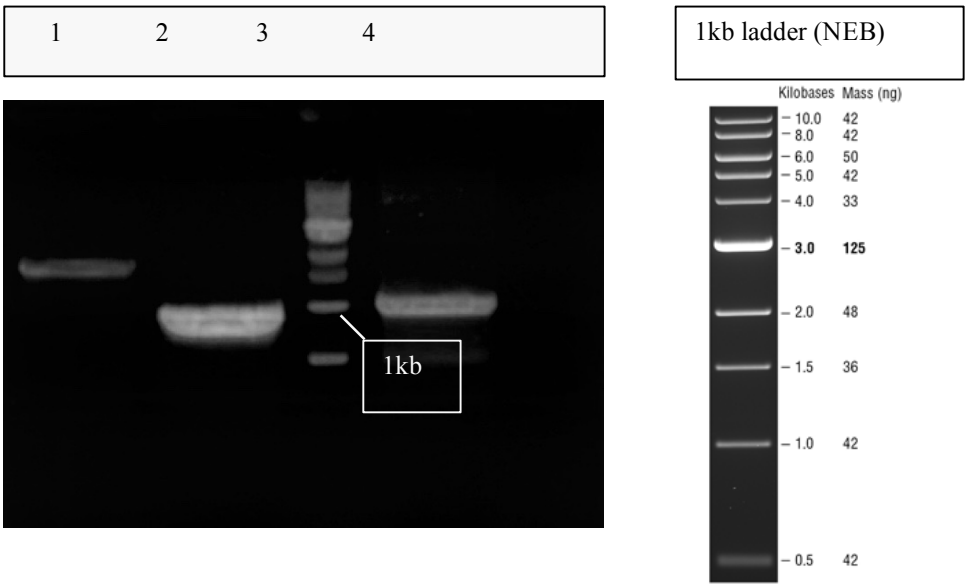


Figure 5.1.2: Gel electrophoresis of expression constructs PCR products
Length of inserts: P*hao* ~140bp; *cyc*A ~950bp; *cyc*B ~790bp; *cyc*AB ~1740bp. Lane1: P*hao*::*cyc*AB (~1880bp) from pBAB1; Lane2: P*hao*::*cyc*B (930bp) from pBAB2; Lane 3: 1kb ladder; Lane 4: P*hao*::*cyc*A (1090bp) from pBAB3.

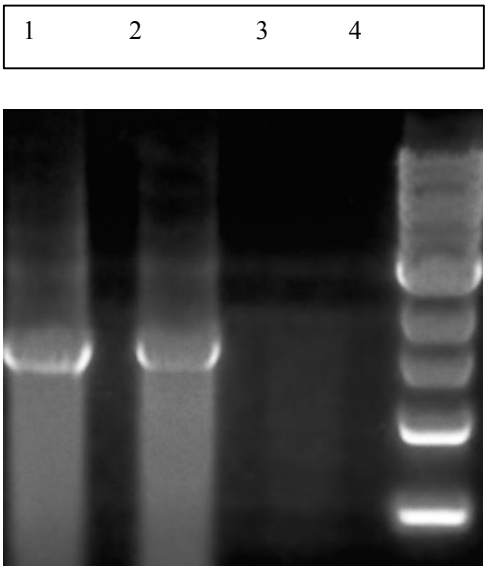


Figure 5.1.3: Gel electrophoresis of colony PCR of P*hao*-*cyc*AB.
Lane 1: positive control (plasmid pBAB1); Lane 2: *M.cap*-pBAB1; Lane 3: wild-type *M.cap*; Lane 4: 1kb ladder.

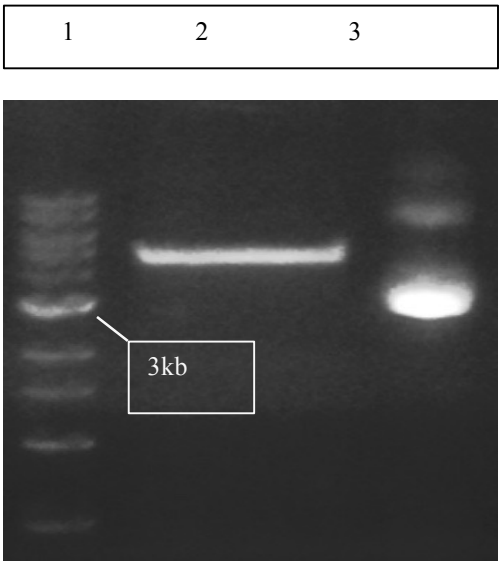


Figure 5.1.4: Restriction analysis of *M.cap*-pBAB1. Lane 1: 1kb ladder; Lane 2: plasmid extracted from T1 digested with restriction enzyme MfeI; Lane 3: uncut plasmid extracted from *M.cap*-pBAB1.

Figure 5.1.5: Growth of *M.cap*-pBAB1 on agar plates.
Cultures from flasks were transferred to agar plates for maintenance. Left agar: growth of *M.cap*-pBAB1 on AMS+CH_4 +NH_3 plus 25µg/ml kanamycin (ANK); Right agar: growth of *M.cap*-pBAB1 on NMS+CH_4 plus 25µg/ml kanamycin (NK) plat

5.2 Nitrogen Compound Transformations by *Methylococcus capsulatus* Bath Are Affected by Temperature and Expression of Genes in the *hao* Cluster

5.2.1 Introduction

Many aerobic obligate methanotrophic bacteria have the genetic capacity to reduce dinitrogen (N fixation) and oxidize nitrite to ammonium (ammonification), oxidize ammonia to nitrite (nitrification) and reduce nitrite to nitrous oxide via NO (denitrification). Therefore, dependent on their particular microenvironment, they can facilitate nearly all known transformations in the nitrogen cycle except for hydrazine synthesis, the signature reaction of anammox bacteria (Campbell et al., 2011a, Stein and Klotz 2011, 2016). Most of these transformations must occur in the absence of oxygen, which methanotrophs with an obligate aerobic catabolism achieve by reactive compartmentation. While the studies of carbon metabolism in methanotrophs are fairly advanced, much less is known about the molecular underpinnings and the regulation of the N-transformations processes they facilitate as well as their regulation, which responds to many environmental signals including substrate availability, pH and temperature. Due to the modular nature of the N cycle (Stein and Klotz 2011, 2016), directions and outcomes of these transformations are dependent on import and export of nitrogen compounds as well as pool sizes of intermediates.

Methylococcus capsulatus strain Bath is an obligate aerobic methane-oxidizing bacterium that can oxidize ammonia to nitrite (nitrification) via the intermediate hydroxylamine (Poret-Peterson et al., 2008). Its genome has been sequenced (Ward et al., 2004) and employed for genome-informed interpretation of assimilative and dissimilative physiological processes. *Methylococcus capsulatus* Bath assimilates ammonium by either

alanine dehydrogenase (ADH) or the glutamine synthetase – glutamate-oxoglutarate-amino transferase (GS-GOGAT) complex (Trotsenko and Murrell 2008). Ammonium can be obtained directly from the environment, produced by the reduction of N_2 by nitrogenase (*nif*) or by reduction of nitrate to nitrite (nitrate reductase, *nasA*) followed by reduction of nitrite to ammonium (siroheme nitrite reductase, *nirB*), processes that take place in the cytoplasm. Nitrate and nitrite pools in the cytoplasm and the periplasm are exchanged by the carrier proteins NarK and/or FosA, which functionally connect assimilatory (cytoplasm) and dissimilatory (periplasm) N-transformation reactions across the plasma membrane as well as between compartments of the Intracellular Membrane System (ICM). During nitrification, *Methylococcus capsulatus* Bath oxidizes hydroxylamine produced by monooxygenation of ammonia by particulate methane monooxygenase (pMMO, *pmoCAB*) primarily using a functional hydroxylamine dehydrogenase (HAO, *haoA*) (Poret-Peterson et al., 2008). In addition, it also has the capacity to utilize the tetraheme cytochrome *c* protein P460 (*cytL*) for the co-oxidation of hydroxylamine and nitric oxide to nitrite, which was assumed to be the basis of methanotroph nitrification before the discovery of functional HAO (Poret-Peterson et al., 2008). The latter authors also reported that *M. capsulatus* Bath produces nitrous oxide by reduction of nitric oxide, which can originate from nitrite or hydroxylamine (Campbell et al., 2011a, Klotz and Stein 2008, Stein and Klotz 2011, 2016). While most genomes of aerobic nitrifying methanotrophs encode *bona-fide* NO-forming nitrite reductases (*nirS* and *nirK*), which facilitate the first step of "nitrifier denitrification" yielding nitrous oxide, the genome of *M. capsulatus* Bath lacks *nirS* and *nirK* genes (Ward et al., 2004). To explain the measured nitrous oxide in *M. capsulatus* Bath cultures (Poret-Peterson et

al., 2008), it was proposed that NO could be the incomplete product of hydroxylamine dehydrogenation by HAO (Hooper et al., 1990, Kostera et al., 2008) or that NO-forming reduction of nitrite could be facilitated by a reversely operating HAO (Poret-Peterson et al., 2008). Following this prediction, octaheme cytochrome *c* proteins in the HAO family (Kern et al., 2011, Simon and Klotz 2013) operating by default as nitrite reductase producing NOx and ammonium were, indeed, discovered in several bacteria (Campbell et al., 2009, Hanson et al., 2013, Maalcke et al., 2014) and another study presented evidence for the reversibility of "regular" hydroxylamine dehydrogenase isolated from the ammonia-oxidizing bacterium, *Nitrosomonas europaea* (Kostera et al., 2008, Kostera et al., 2010).

In addition, a prior study on the expression of N-transformation genes in response of *M. capsulatus* Bath to exposure to external ammonium, nitrite and NO-generating agents reported that *M. capsulatus* Bath differentially expresses genes encoding nitric oxide reductases (*norCB*, and *cytS*) and inventory employed for hydroxylamine oxidation (*cytL* and *haoA*) in response to individual or a combination of these stresses, discovered that *M. capsulatus* Bath grown in NMS with CH_4 as its sole source of energy, reductant and carbon produced measurable N_2O only when amended with ammonium and nitrite together (Campbell et al., 2011b). Because the expression of *haoAB* and *cytS* (and not of *norCB*) was induced by ammonium (Poret-Peterson et al., 2008) and nitrite induced expression of *norCB* (and not *haoAB* and *cytS*), Campbell et al. (2011b) concluded that these genes together encode the required inventory for the formation of N_2O in *M. capsulatus* Bath. The tandem co-expression of the *haoAB* genes was first reported for *M. capsulatus* Bath (Poret-Peterson et al., 2008) and after confirmation of the same in AOB

(Stein et al., 2013), the second open reading frame in the *haoA-orf2-cycAB* gene cluster conserved in AOB (Bergmann et al., 2005) was given the *haoB* gene designation albeit without a hypothesized protein function (Poret-Peterson et al., 2008). Poret-Peterson et al. (2008) also proposed that a stem loop in the 5'-end of the *haoB* gene could be the reason for differential lengths of *haoA* transcripts as an ammonium-induced process involving 3'-end RNA processing, which has been proposed to be responsible also for the processing of *amoCAB* transcripts in AOB (El Sheikh and Klotz 2008, El Sheikh et al., 2008). The presence of this stem loop in the 5'-end of the *haoB* gene has been confirmed for all *haoAB* gene tandems in nitrifying ammonia- and methane-oxidizing bacteria (Campbell et al., 2011b).

The putative translation product of the *haoB* transcript is predicted to be a periplasmic, membrane associated protein whose N-terminal transmembrane-spanning domain immediately succeeding the signal peptide is encoded by the palindromic stem loop-forming sequence in the the 5'-end of the *haoB* transcript and whose C-terminal protein sequence exhibits low identity between deduced sequences of all known putative HaoB proteins (Campbell et al., 2011b). By extending this work with the goal to discover a function for the putative HaoB protein, we succeeded in generating an *M. capsulatus* Bath mutant that lacks the stem loop in the 5'-end in the *haoB* gene. Experimentation with this 5'-end stem loop mutant in comparison with WT *M. capsulatus* Bath produced results indicating differential temperature-dependent regulation of NOx pools. Determined steady-state mRNA levels of above listed N-transformation inventory in *M. capsulatus* Bath WT and mutant under ammonium stress implicate the *haoB* gene

product in the modulation of nitrite levels by a yet unknown mechanism, which is subject to ongoing investigation.

5.2.2 Materials and Methods

5.2.2.1 Culture Conditions

5.2.2.1.1 *M. capsulatus* Bath Maintenance

Batch cultures of *M. capsulatus* Bath were maintained at 45°C on a rotary shaker (200 rpm) in 100 ml nitrate mineral salts medium (NMS) containing 5~10 μM $CuSO_4$ (Whittenbury et al., 1970) in 250 ml Wheaton bottles sealed with rubber stoppers under an initial headspace mixing ratio of 30:70 (CH_4 to air). NMS agar plates were prepared with addition of 1.5% (w/v) Bacto (Difco) agar. *Methylococcus capsulatus* Bath grown on NMS agar plates was incubated at 45°C in a gas-tight chamber with the addition of CH_4.

5.2.2.1.2 Temperature and Ammonium Effects on Batch Culture Activities

Stationary phase cultures grown at 45 °C in NMS were harvested by centrifugation (6000 x g, 10 min, 25 °C) and washed twice with phosphate buffer (5.4 g $Na_2HPO4.\ 7H_2O$ and 2.6 g KH_2PO_4 per liter distilled H_2O). Cell pellets were resuspended in 200 μl of the same phosphate buffer and transferred into 50 ml of NMS, NMS + ammonium (2.5 mM $(NH_4)_2SO_4$) or ammonium minimal salts medium (AMS, 2.5 mM $(NH_4)_2SO_4$)) supplemented with CH_4 in 250 ml Wheaton bottles sealed with butyl rubber stoppers. Cultures were grown at 30 °C, 37 °C or 45 °C respectively for overnight or a few days on a rotary shaker (200 rpm). Ammonium and nitrite concentration of cultures were measured (Nicholas and Nason 1957) after incubation of 24 hours and 48 hours.

5.2.2.1.3 Batch Culture for RNA Extraction

For RNA extraction, 100 ml of early-exponential phase *M. capsulatus* Bath WT and mutant cultures in NMS maintained at 45 °C were harvested by centrifugation (6000 x g, 10 min, 25 °C) and rinsed twice with phosphate buffer. Pellets were resuspended in 1 ml of phosphate buffer and inoculated to 100 ml of NMS+ CH_4 (nontreatment) or NMS+CH_4 with 2.5 mM $(NH_4)_2SO_4$ (NH_3 treatment) and additional 5 μM $CuSO_4$. Cultures were incubated for 4 hours under the above-described growth conditions. Cells were then harvested for RNA extraction.

5.2.2.2 Construction of Stem Loop Mutant (SLM)

In order to knock out the stem loop between *hao*A and *hao*B from the genome of *M. capsulatus* Bath WT, PCR product containing the *hao*A gene and 5'-terminial of *hao*B was firstly cloned into pCR2.1-TOPO vector created "pMC1a" construct (with primers RE1haoAR and MCA0956F; Appendix A). Another PCR product, which contains a modified *hao*B gene (*hao*B') with the stem-loop of interest disrupted was cloned to pCR2.1-TOPO vector and generated "pMC1b" construct (with primers RE1haoAF and MCA0957R; Appendix A). Fragment *hao*B' of pMC1b was spliced downstream of *hao*A operon in the pMC1a via restriction digestion and DNA ligation and produced "pMC1ab" construct (Fig.5.2.1). The modified *hao*AB operon containing the disrupted stem-loop region was then transferred to the pEX18Tc-based suicide plasmid with kanamycin resistance feature (Hoang et al., 1998). Using the constructed plasmid, allelic exchange was performed with *M.capsulatus* Bath , facilitated by *E. coli* S17-1 strain (Simon et al., 1983). The resulting SLM strain was verified with PCR and sequencing.

5.2.2.3 RNA extraction, cDNA Synthesis and RT-qPCR

RNA was extracted and purified with Qiagen RNeasy mini Kit (Qiagen, Germany) following the manufacturer's protocol (see details in Chapter 3). RNA samples were then treated with RQ1 RNase-free DNase (Promega, USA) (see details in Chapter 3). After checking integrity and quantity, RNA was converted to first-strand cDNA with QuantiTect Reverse Transcription Kit (Qiagen, Germany). Quantitative PCR was performed with synthesized cDNA samples (WT nontreatment; WT treated with NH_4^+; SLM nontreatment and SLM treated with NH_4^+) in 10 µl reactions in triplicate. Each reaction consists of 5 µl SYBR Green Supermix (Bio-Rad, USA), 500ng forward and reverse primers, 40 ng cDNA template and sterile dH_2O. All reactions were run on real-time thermal cycler (Bio-Rad, USA) with the following parameters: holding at 95 °C for 5 min, 40 cycles of denaturing at 95 °C for 3 s, annealing at 55 °C for 30 s and extending at 72 °C for 15 s. A melting curve was generated over a temperature range from 70 °C to 92 °C following PCR. Primer pairs (Appendix A) specifically designed to compare the relative expression of *hao*, *heme*8, *orf*2, *cyt*L, *cyt*S, *nor*C, *nor*B, *nas*A, *nir*B, FNR homologue (2120, 2130), ADH (alanine dehydrogenase) and GS (glutamine synthetase) with 16S rRNA gene as internal reference. QPCR Data were analyzed using 2(-delta delta *Ct*) method (Livak 2001).

5.2.3 Results and Discussion

5.2.3.1 Construction of SLM

Genomic DNA extracted from WT and SLM cells was used for PCR with qhaoterm2-F/R primers. Amplicons obtained from SLM genomic DNA exhibited the expected fragment size (Fig.5.2.2). The nucleic acid sequences of amplicons obtained

from SLM and WT cells were aligned using Clustal Omega (Sievers et al., 2011, McWilliam et al., 2013, Li et al., 2015). Sequence analysis revealed that the probability of stem loop formation in the spacer between the *haoA-B* genes in SLM compared to the stem loop in the 5'-end of the *hao*B gene in WT decreased from 57.90% to 0.56% (Fig.5.2.3). Consequently, an increase in *hao*A promoter activity would increase the abundance of the *hao*AB transcript with only a 0.6% chance for differential expression of *hao*A and *hao*B and subsequent 3'-end processing of *hao*A yielding a *hao*A' transcript. The translation product (HaoB) of the modified *hao*B in SLM cells was predicted by the TMHMM 2.0 server (http://www.cbs.dtu.dk/services/TMHMM/) to be a transmembrane protein (Fig.5.2.3).

5.2.3.2 Temperature Effect on NO_2^- Production in WT and SLM Cells of *M. capsulatus* Bath

In cultures of both *M. capsulatus* Bath WT and SLM cells, NO_2^- production was first observed at higher temperature as soon as one hour after initial inoculation (45 °C, Fig. 5.2.4 A and B). Observations also showed that higher temperature speeds up nitrite removal in both strains (Fig.5.2.4 A and B). For SLM *M. capsulatus* Bath, nitrite was completely removed in 24 hours at 45 °C compared to 48 hours were needed at 37 °C. Nitrite was not depleted at 72 hours at 30 °C. At a temperature of 45 °C, NO_2^- was still measurable in WT at 24 hour and 48 hour. NO_2^- production only slightly increased after incubation for 24 hours in WT *M. capsulatus* Bath (Fig. 5.2.4 A) at 30 °C; however, it was dramatically increased in SLM *M. capsulatus* Bath after 24 hours' incubation (Fig. 5.2.4 B). In parallel, cell density was monitored by measuring OD_{600} with a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific, USA) and no significant

difference between WT and SLM cultures was recorded during the experiment.

WT *hao*AB operon transcription at low basal level has a 42% probability of read-through thereby producing different template levels for translation into HaoA & HaoB. Zero ammonium in the medium means that no hydroxylamine and, subsequently, no nitrite can be produced as a result of nitrification activity. Therefore, measured nitrite should result solely from NasA activity, which reduces nitrate in the cytoplasm where NirB reduces nitrite subsequently to ammonium for assimilation. Theoretically, enzyme activities are generally higher and transport across the PM is faster at higher temperatures; however, manipulation activities of nucleic acids are usually lower due to their coilicity. Our observations suggest that higher temperature reduces the levels of detectable nitrite (Fig. 5.2.4). This could be the result of reduced nitrate reduction activity (due to lower NasA enzyme activity or lower NasA levels) and/or higher nitrite reduction activity (due to enhanced NirB enzyme activity or higher enzyme levels) whereby different enzyme levels are most likely tied to changes in steady state transcript levels. There are no reports in the current literature that describe temperature dependence of NasA and NirB enzyme activities or their synthesis processes; therefore, their roles in the dynamics of nitrite pool turnover as a function of temperature is likely due to substrate availability (which is determined by uptake) or the involvement of other protein functions.

5.2.3.3 Effect of Ammonium on Nitrite Levels and Gene Expression

Identical growth of WT and SLM *M. capsulatus* Bath cultures was observed (data not shown) when cultivated at 45 °C in NMS, [NMS + 5 mM NH_4^+] or AMS media supplemented with CH_4. The concentration of NH_4^+ in cultures decreased along with time

for both strains in [NMS + 5 mM NH_4^+] and AMS media; however, the SLM strain consumed NH_4^+ more effectively in AMS media at 48 hour as determined by spectroscopy (absorbance at 475 nm; Fig.5.2.5 B). Experimental detection of NO_2^- was more variable than measurements of NH_4^+ levels for different treatments, times and strains. For CH_4-only treatment (no ammonium present), nitrite levels in SLM cultures were significantly lower than in WT cultures at 48-hour whereas the levels were identical at 24-hour (Fig.5.2.5 B). When supplemented with ammonium, SLM cultures contained less nitrite than WT cultures after 24 hours (higher rates of NO_2^- removal by NirB or less NO_2^- production by NasA) whereas levels of NO_2^- were similar in both cultures at 48-hour (Fig.5.2.5 B).

Real-time quantitative PCR results revealed that the steady-state mRNA levels of *haoB* (*orf*2) were increased ~15 fold in WT and ~30 fold in SLM cultures of *M. capsulatus* Bath (Fig.5.2.6 C). Thus, the stem loop between *haoA* and *haoB* clearly reduced the transcription of *haoB* in the presence of ammonium. The steady-state mRNA levels of *nasA* (encoding nitrate reductase) and *nirB* (encoding nitrite reductase), which correspond to the concentrations of pertinent enzymes required for nitrate/nitrite reduction were down regulated under ammonium treatment (Fig.5.2.6 A & B). This is best explained by a regulation mechanism that saves reductant needed for the subsequent reduction of nitrate and nitrite when accessible ammonium levels satisfy assimilation requirements. We have determined mRNA levels for additional genes that encode proteins involved in N-oxide detoxification: cytochrome *c*′-beta (*cytS*) and cNOR (*norCB*, cytochrome *c*-dependent nitric oxide reductase), both implicated in NO reduction (Poret-Peterson et al., 2008) whereas *cytS* steady-state mRNA levels increased

~ 17-fold (both in WT and SLM) in response to ammonium, co-transcribed *norCB* mRNA levels only increased ~ 5-fold. This is in consistent with previous findings (Poret-Peterson et al., 2008).

In contrast to the previous study by Poret-Peterson et al. (2008), which reported that the steady-state level of *hao*A mRNA targeted at the 3′ end (q2hao2) of the transcript was lower than at the 5′ end (q1hao) suggesting that a truncated *haoA* transcript might lead to synthesis of a truncated HaoA' enzyme, here the detected steady-state *hao*A mRNA levels at both ends were identical (Fig.5.2.6 A & B). In addition, there was no difference between detected steady-state *hao*A mRNA levels targeted at the 3'-end in WT and SLM *M. capsulatus* Bath cultures, which are significantly elevated in cells of both cultures (Fig.5.2.6 A & B). In contrast, exposure to ammonium caused significantly higher steady-state *haoB* mRNA levels in cells of SLM *M. capsulatus* Bath compared to WT cells (Fig.5.2.6 C). Because this potential for higher HaoB protein levels correlates with lower nitrite levels in SLM cultures, we postulate that the putative HaoB protein, which is anchored in the PM towards the periplasm similarly to the HaoA protein complex, is facilitating "nitrite removal" by a hitherto unknown mechanism. Because SLM and WT cultures have comparable growth rates and thus comparable putative N-assimilation rates, the pools of fixed nitrogen compounds differ measurably only for nitrite and expression potential for cytoplasmic nitrite producing (NasA) and consuming (NirB) enzymes is equally reduced in the presence of external ammonium, the differing nitrite levels are best explained by a HaoB interaction with the hydroxylamine dehydrogenase complex (HAO) that lowers its effectiveness to disproportionate hydroxylamine to nitrite and, instead, releases the less oxidized intermediate nitric oxide.

5.2.4 Conclusions

Ammonium and hydroxylamine increase *haoA* transcription (Poret-Peterson et al., 2008, Campbell et al., 2011b) and, concomitantly, HAO produces nitrite and NO. WT cells should produce more nitrite at higher temperatures. Hence with time, more nitrite is being produced by nitrate reduction in the cytoplasm and by hydroxylamine oxidation in the periplasm than nitrite is being removed by nitrite reduction to ammonium in the cytoplasm. With consistent external supply of ammonium, more nitritation capacity will produce considerably more nitrite than can be removed. WT cells also synthetize less HaoB than HaoA due to the ammonium-induced termination of polymerase at the 5'-end *haoB*; however, if higher temperatures increase read-through the stem loop in the 5'-end of *haoB*, this would lead to more *haoB* transcript and more HaoB protein with time. Compared to WT, SLM cells have significantly less premature termination of *haoAB* transcription likely approaching zero at higher temperatures, which may lead to nearly equimolar translatable *haoA* and *haoB* transcript levels and thus significantly more HaoB protein compared to WT. Public databases to not include any proteins sequence-similar to HaoB and no classified conserved domains (CCD) have been identified. Without a link to homologous protein functions to be derived from public databases, we propose that HaoB interacts with the HAO complex leading to less nitrite formation and, instead, increased release of NOx from HAO into the periplasm. Continuing collaborative studies will focus on the proposed interaction between HaoB and the HAO complex, which is a circular-symmetric protein-ligated homotrimer of HaoA proteins (Simon and Klotz 2013) and references therein).

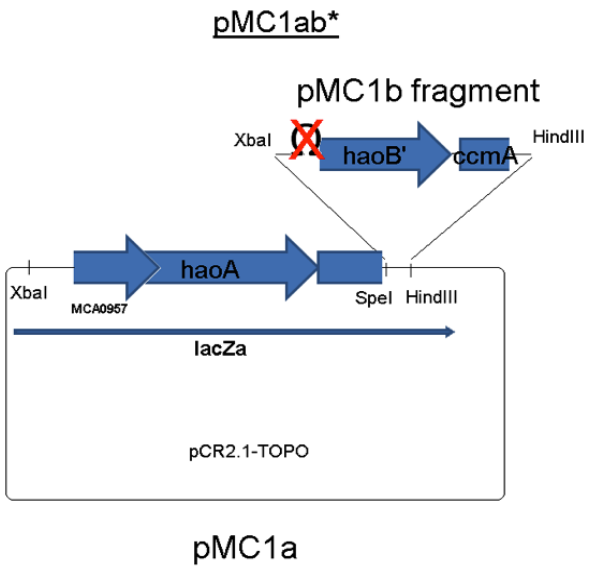


Figure 5.2.1: Scheme of construction of pMC1ab. *ccmA* (mac_0954 heme exporter protein CcmA). pMC1a was firstly digested with HindIII and SpeI to produce a 5699-bp fragment; pMC1b was digested with HindIII and XbaI to produce an 1849-bp fragment; HindIII-XbaI digested pMC1a fragment was ligated into pMC1a HindIII-SpeI site to produce pMC1ab (7548 bp).

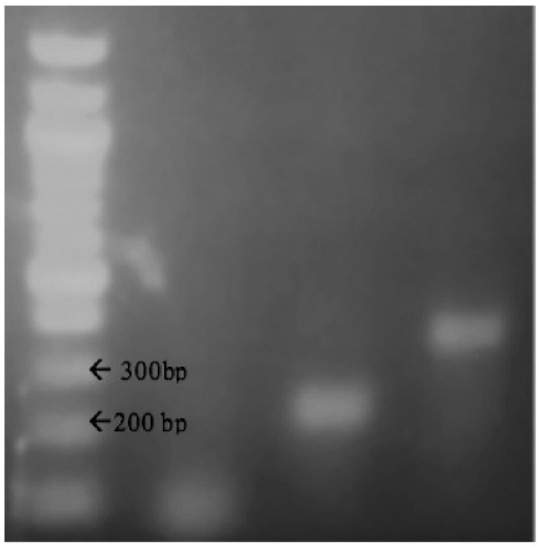


Figure 5.2.2: PCR with qhaoterm2-F/R primers. Expected size for amplicon derived from stem loop mutant is ~312 bp compared to ~192-bp from WT template. Lane 1: 100-bp ladder; Lane 2: Negative control; Lane 3: WT *M. capsulatus* Bath; Lane 4: SLM *M. capsulatus* Bath.

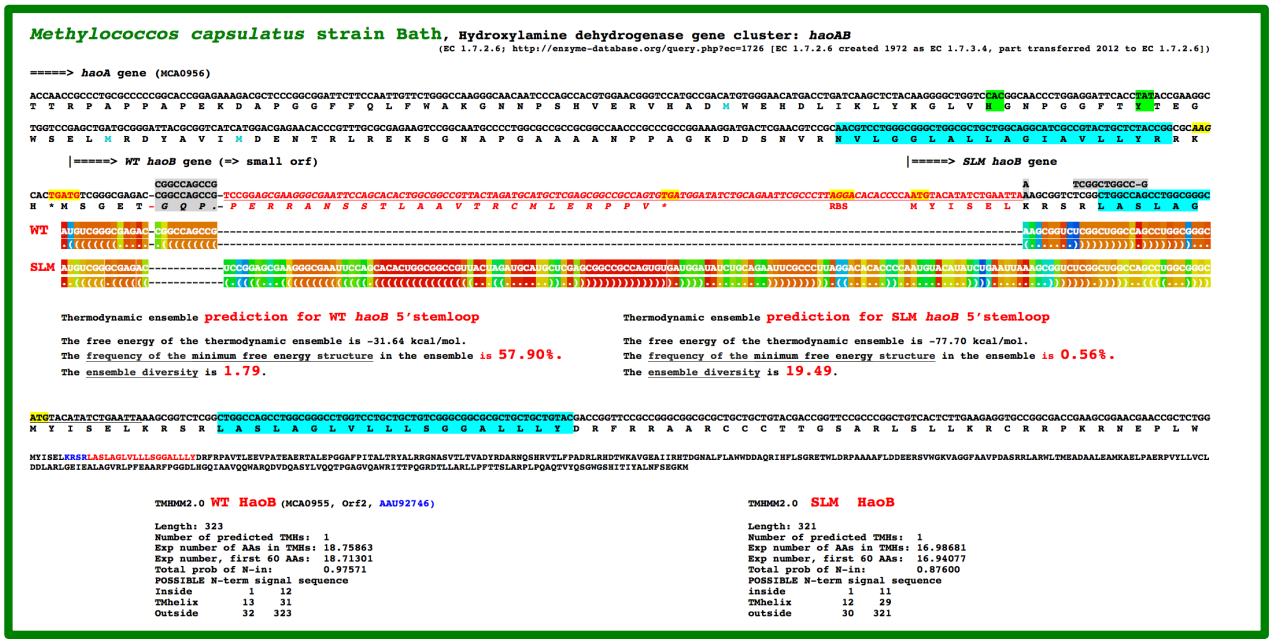


Figure 5.2.3: Sequence analysis of *M. capsulatus* Bath WT (Fitch and Upper) and SLM (lower) *hao*A operon.

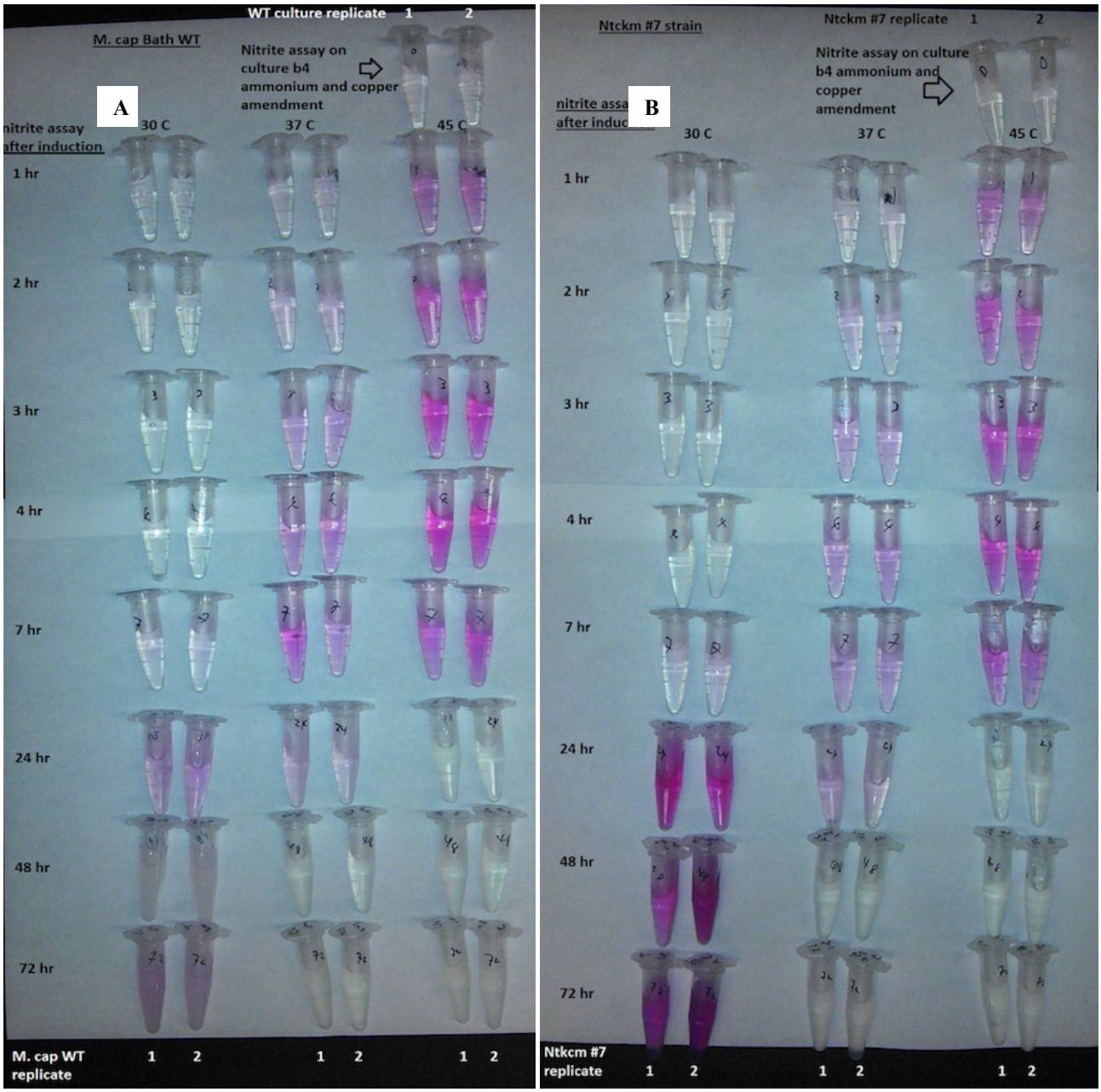


Figure 5.2.4: Hourly measurement of nitrite production by *M. capsulatus* Bath WT (A) and SLM (B) grown in NMS media at 30 °C, 37 °C and 45 °C. Number "1" and "2" indicate replicate experiments. Each tube contained 650 µl of culture sample and 250 µl of 1% sulfanilamide and 0.02% N- (1-napthyl) ethylenediamine dihydrochloride. Darker colors indicate a higher concentration of NO_2^- produced in cultures. NO_2^- was measured at time point of 1 hour, 2 hour, 3 hour, 4 hour, 7 hour, 24 hour, 48 hour and 72 hours after the initial inoculation.

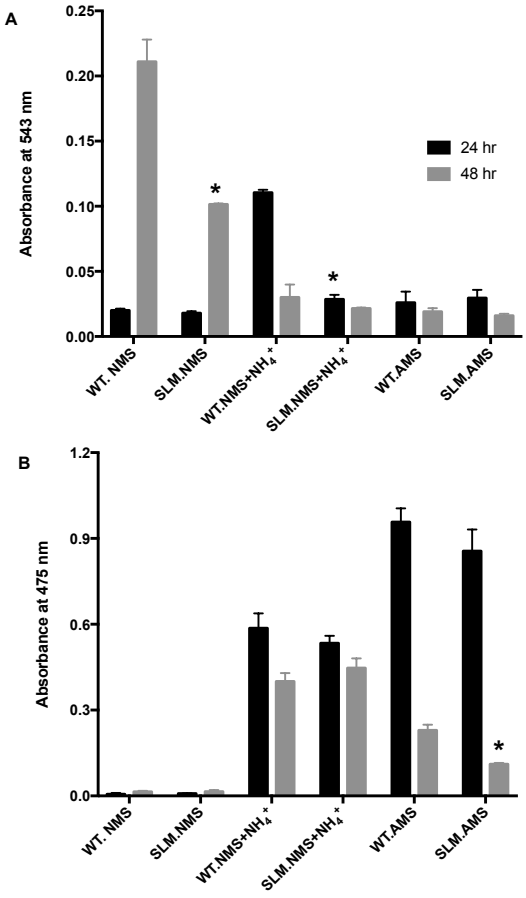


Figure 5.2.5: Effect of ammonium on nitrite production of *M. capsulatus* Bath WT and SLM cultures. Y-axis of figure A is the absorbance at 543 nm to measure nitrite levels at 24 hour (black bar) and 48 hour (grey bar) after initial inoculation. Figure B shows the ammonium concentration in samples at 24 hour and 48 hour by measuring absorbance at 475 nm. Data bars are mean values of two independent trials with standard deviation.

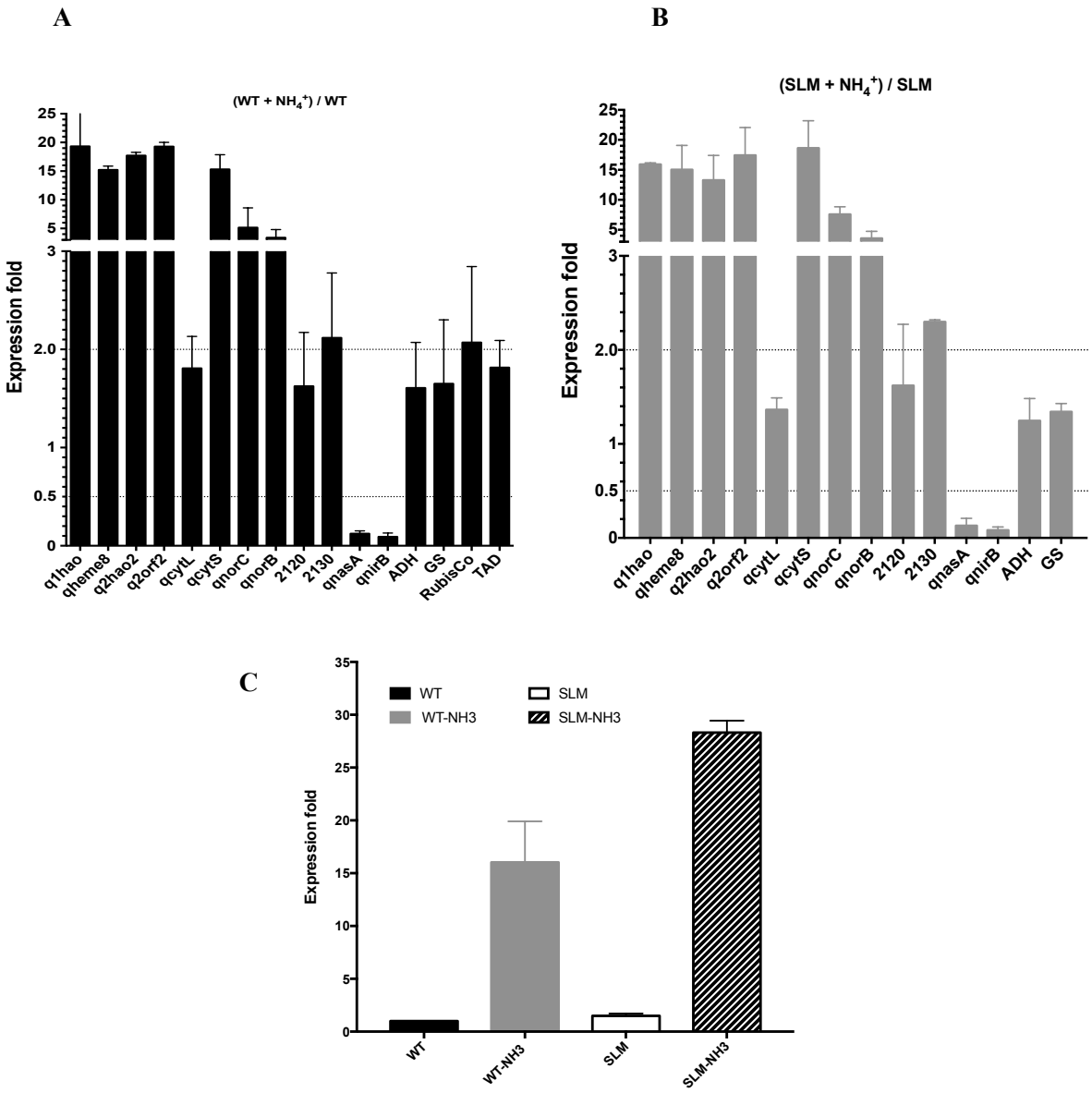


Figure 5.2.6: Ammonium effect on steady-state mRNAs levels of *hao*A, *hao*B (*orf*2), *cyt*S, *cyt*L, *nor*C, *nor*B, *nas*A, *nir*B and C1 metabolism related genes measured by quantitative PCR. Each qPCR reaction was performed in triplicate. Data bars are mean values of two independent trials of WT and SLM *M. capsulatus* Bath after 4-hour exposure to CH_4-only or 5 mM NH_4^+ with standard deviation error bars. Figures A and B show ammonium effect on expression levels of genes related to nitrogen transformation and C1 transformation in WT (A) and SLM (B) *M. capsulatus* Bath. Figure C shows the ammonium effect on expression level of *haoB* (*orf*2) for both WT and SLM *M. capsulatus* Bath.

5.3 Ammonification and Denitrification Pathways in *Nautilia profundicola*

5.3.1 Introduction and Background

Sulfur-dependent deep-sea vent *Epsilonproteobacteria Nautilia profundicola* grows chemolithotrophically on hydrogen with ammonium or nitrate (Smith et al., 2008). It is also able to use nitrate as a sole source of nitrogen and as a terminal electron acceptor (Hanson et al., 2013). However, it lacks any kind of known ammonium-forming assimilatory (NirB/NirA) or dissimilatory (Gunsalus and Wang 2000) nitrite reductases, or NO-forming (NirS/NirK) nitrite reductases, nor are homologs of nitrite transporters (NarK, FocA, etc.) present (Campbell et al., 2009). A reverse-HURM pathway has been proposed (see details in Chapter 1), which is consisting of HAO and cytochrome c_M552, together with nitrate reductase (NapA) and hydroxylamine reductase (Hcp) for both nitrogen assimilation and respiration in *Nautilia* generating hydroxylamine and ammonium (Campbell et al., 2009, Hanson et al., 2013, Simon and Klotz 2013; Chapter 1). Indeed, the pathway is thought to employ two catalytic redox modules operating in sequence: the reverse-HURM reducing nitrite to hydroxylamine followed by a hydroxylamine reductase, also known as the hybrid cluster protein (Har/Hcp), that converts hydroxylamine to ammonium. Genes hypothesized to be key in the pathways are (Fig. 5.3.1): 1) *nap*A (NAMH_0556), catalytic subunit of periplasmic nitrate reductase; 2) *cyc*B (napC, NAMH_0559), quinol oxidase; 3) *hao*A' (NAMH_1280), hydroxylamine oxydoreductase, lacking the protein-ligating tyrosine reside; 4) *har* (hcp, NAMH_1074) hydroxylamine reductase HCP; 5) *amt*B-1 (NAMH_0215) and *amt*B-2 (NAMH_0397) ammonia transporters; 6) [Fe-S] cluster protein (NAMH_1302) providing reducing power to HCP. Consistent with our hypothesis above, components of the proposed pathway

(Fig. 5.3.1) displayed a strong increase in transcript abundance in nitrate-grown cultures relative to ammonium as reported previously by Hanson et al. (2013). Transcripts of *nap*A (NAMH_0556), *hao*A (NAMH_1280), *cyc*B/*nap*C (NAMH_0559), were increased by 4.6-, 8.5- and 7.1-fold, respectively. The ammonium/methylammonium transporter AmtB were increased by 4.5-fold (*amt*B-1, NAMH_0397) and 10.3-fold (*amt*B-2, NAMH_0215). Transcripts of *har* (hcp, NAMH_1074) and the putative reductant-donating [Fe-S] cluster protein (NAMH_1302) were increased by 11.7- and 4.6-fold, respectively (Hanson et al., 2013).

To test the hypothesis of *haoA'* (NAMH_1280) being involved in nitrite reduction, heterologous hosts that are lacking of nitrite reduction capability are utilized for such test. *Pseudomonas protegens* Pf-5 and *Pseudomonas chlororaphis* O6, which were previously shown to be unable to reduce nitrite, were chosen as the heterologous hosts (Loper et al., 2012). The mobilizable pRK415 plasmid (Table 5.3.1), which can be maintained in *Pseudomonas* species, was used as the expression vector. The nitrite reduction cassette, consisting of the *haoA'* and *napC*, as well as their respective endogenous ribosomal binding site, was cloned into the pRK415 plasmid (Table 5.3.1). In the vector construct, the gene expression is to be driven by the *lac* promoter of the pRK415 plasmid, which is known to be constitutively active in *Pseudomonas* species (Hentzer et al., 2002). The plasmids are conjugated into *Pseudomonas* strains by *Escherichia coli* S17-1. Preliminary experiment on *P. chlororaphis* O6 indicates that *haoA'* might have some nitrite reduction activity (Fig.5.3.3). In this experiment, nitrite was not detected in the culture supernatant of *P. chlororaphis* O6 strain that did not carry the nitrite reduction cassette while they were observed in the wildtype as well as the

controls (Fig.5.3.2), where nitrite might be produced via reduction of nitrate by the host (Loper et al., 2012).

Hemerythrin has been also found and characterized in *Methylococcus capsulatus* strain Bath, a gammaproteobacterial aerobic methane-oxidizing bacterium that can oxidize ammonia to nitrite via the intermediate hydroxylamine. It lacks conventional dissimilatory NO- and ammonia-forming nitrite reductases (NirS/NirK) but possesses the assimilatory siroheme nitrite reductase (NirB) to detoxify nitrite to ammonium. *M. capsulatus* can also assimilate ammonium obtained directly from the environment by ADH or GS-GOGAT. The hemerythrin was expressed in *M. capsulatus* only when grown at high-copper concentrations when methane is oxidized by particulate instead of soluble methane monooxygenase (Burrows et al., 1984). Despite the different heme cytochrome c synthesis and maturation systems in *Epsilon*- and *Gammaproteobacteria*, the HaoA and HaoA' octaheme cytochrome c proteins have regular heme-binding motives (CxxCH) and conserved heme-ligating histidine residues not necessitating the operation of specific heme ligases (Klotz et al., 2008). *Nautilia hao*A' and *nap*C genes should therefore be expressible to functional Heme cytochrome c proteins in *Methylococcus capsulatus*. A mutant *Methylococcus capsulatus* (SLM; Table 5.3.1; Chapter 5.2), in which we have removed the stemloop in the 5.' region of the *hao*B gene of the *hao*AB tandem (Poret-Peterson et al., 2008) that is implicated in alternative processing of the *hao*A transcript and the production of HaoA', the predicted nitrite reductase in *M. capsulatus* wild-type (WT; Table 5.3.1). Therefore, mutant of *M. capsulatus* strain Bath will be the best suitable host for the functional study of [HaoA'-c_M552/NapC] facilitated nitrite reduction.

The following constructs were already built by Klotz lab (Table 5.3.1) prior to this study: pC0559, pC0559-1280, pC1280, pR0559, pR0559-1280, pR1280, pB0559, pB0559-1280, and pB1280. Therefore, the main goal of this study is to experimentally detect the functionality of "nitrite reductase-quinol oxidase complex" (HaoA'-c_M552/NapC complex) encoded in the *Nautilia profundicola* AmH chromosome in heterologous recombinant hosts WT and SLM *M. capsulatus*.

5.3.2 Methods and Materials

5.3.2.1 Bacterial Strains, Plasmids and Growth Conditions

Bacterial strains and plasmids used in this research are listed in Table 5.3.1. *E. coli* strains were maintained in Luria–Bertani (LB) broth at 37 °C. *M. capsulatus* Bath strains were grown at 45 °C with shaking at 200 rpm in NMS media (Chapter 5.2) and with supplementation of Kanamycin (25 mg/mL) for the selective growth of SLM strains.

5.3.2.2 Construction of pCR2.1-PpmoC1

Based on preliminary data previously collected by Klotz lab, cassette [NAMH0559-NAMH1280] (Fig.5.3.4 A) was not effectively expressed in M. capsulatus strain Bath. We theoretically postulate that it is because the lac promoter carried by plasmid pBBR1MCS-2 is not sufficiently effective in controlling M. capsulatus strain Bath. We decided to use a stronger promoter, which is the pmoC1 promoter (PpmoC1; Fig.5.3.4) in the chromosome of M. capsulatus strain Bath. Genomic DNA was extracted from M. capsulatus strain Bath and used as template for generating amplicon of PpmoC1 by PCR using specific designed primers (PpmoC F/R; Appendix A) containing RE site (XbaI or SalI; New England Biolab, USA) at their 5'-ends for directional cloning (Fig. 5.3.3). Fragment PpmoC1 (Fig.5.3.4) collected after purification on 2% low-melt agarose

gel cloned to pCR2.1-TOPO TA vector (Invitrogen, USA) and subcloned in XbaI/SalI sites of pBBR1MCS-2 (Kovach et al., 1995, Lynch and Gill 2006) to create construct pBBR- PpmoC1 (Table 5.3.1) verified by sequencing, PCR and RE analyses (Fig.5.3.4 B). Plasmid pBBR- PpmoC1 were used to transform TOP10 E. coli competent cells (Invitrogen, USA) for propagation and maintenance following the manufacture's protocol.

5.3.2.3 Generation of Expression Constructs pBBR-P-1280, pBBR-P-0559-1280.

Plasmid construct pBBR- P*pmo*C1 extracted from TOP10 *E. coli* were digested with XbaI and SalI in buffer CutSmart at 37 °C (New England Biolab, USA). Fragment "XbaI/P*pmo*C1/SalI" was collected after purification on 2% low-melt agarose gel (Fig.5.3.3). Previously created plasmid constructs pBBR-1280 and pBBR-0559-1280 were digested with the same REs creating compatible overhangs were gel-purified (1.5% low-melt agarose gel) and ligated together with the above fragments (Fig.5.3.3). The ligation products were independently used to transform *E. coli* TOP10 competent cells to generate maintainable expression vectors pBBR-P-1280 and pBBR-P-0559-1280, respectively (Fig.5.3.3). The collected plasmids using the alkaline mini-prep method (Wizard Plus SV Minipreps DNA Purification Systems; Promega, USA) (Chapter 2) were then confirmed by DNA sequencing, PCR and RE analyses (Fig.5.3.4 C to G).

5.3.2.4 Construction of *M. capsulatus* Transformants

Expression constructs collected from TOP10 *E. coli* cells (donor strain) were used to transform conjugative strain *E. coli* S17-1 (helper strain) (Phornphisutthimas et al., 2007) with *Methylococcus capsulatus* Bath WT and SLM strains as the recipient strain. The procedure of tri-parental mating was described in detail previously (Chapter 5.1).

Colonies able to grow on the antibiotic selective agar plate will be picked up after incubation (Welander and Summons 2012).

5.3.3 Results and Discussion

5.3.3.1 Confirmation of Built Constructs

Expression constructs pBBR-P-0559 and pBBR-P-0559-1280 were successfully created (Table 5.3.1) and confirmed by sequencing, PCR, gel-electrophoresis and RE analyses (Fig. 5.3.4). Plasmids were extracted from TOP10 *E. coli* transformants and dissolved in 40 μl dH_2O for PCR using specific designed primer sets and (Fig. 5.3.4 B,C,D,E) and DNA sequencing at the DNA Core Facility, University of Louisville. Consistent with the PCR and sequencing results, RE analyses confirmed that the correct size fragments were inserted at the right positions of the plasmid pBBR1MCS-2 backbone (Fig. 5.3.4 F & G). TOP10 *E. coli* cultures carrying these constructs were maintained and transferred regularly in liquid LB media and on LB agar at 4°C. Stock cultures were stored at -80°C in 25% glycerol for future use.

5.3.3.2 Analysis of Transformants

Unfortunately, the expected transformant strains (*M. cap* WT/SLM-pBBR-P-1280 and *M .cap* WT/SLM-pBBR-P-0559-1280) have not obtained yet. Therefore, the next step is to successfully create *M. cap* WT/SLM-pBBR-P-1280 and *M. cap* WT/SLM-pBBR-P-0559-1280 transformant strains. Next in time, it is necessary to analyze the *in vivo* expression levels of insert genes (*hao*A' and *nap*C) and phenotype of the transformant strains of *M. capsulatus*. The ultimate goal of this project is to fully study the functionality and regulation of the proposed "reverse HURM" pathway in the procedure of nitrate assimilation and respiration in certain *Epsilonproteobacteria*. To do

so, we need to build constructs carrying the key genes (see text above) and test in heterologous recombinant hosts including *E. coli* strain LMS0873 (hcp mutant– highly sensitive to hydroxylamine), *Pseudomonas stutzeri* strain MK202 (mrS::Tn5, SmR, KmR - highly sensitive to nitrite) beyond *M. capsulatus* and study their functionality individually.

Table 5.3.1: Strains and plasmids used in this study

Strain or plasmid	Description	Source or Reference
E. coli		
Top 10	Chemically competent cells	Invitrogen, CA, USA
S17-1	Tp^R, Sm^R, recA, thi, pro, hsdR-M+RP4: 2-Tc:Mu: Km^R, Tn7, λpir	Simon et al. (1983)
Methylococcus capsulatus		
Wild-type	gammaproteobacterial nitrifying and denitrifying methanotroph; HAO^+, $HaoA'^+$, $NirB^+$, $NirK/NirS^-$, $NrfA^-$	Laboratory stock
SLM	Stem loop between *hao*A and *hao*B was removed; HAO^+, $HaoA'^-$, $NirB^+$, $NirK/NirS^-$, $NrfA^-$	This study; chapter 5.2
Plasmids		
pCR2.1-TOPO	Cloning vector; Ap^R, Km^R	Invitrogen, CA, USA
pRK415	RK2-derived broad-host-range expression vector; low-copy number; Tc^R	Keen et al. (2008)
pBBR1MCS-2	Broad-host-range expression vector; multiple cloning site in LacZ; medium-copy nuber; Km^R	Kovach et al. (1995)
pCR-0559	pCR2.1-TOPO:: *nap*C	This study
pCR-0559-1280	pCR2.1-TOPO:: *nap*C-*hao*A'	This study
pCR-1280	pCR2.1-TOPO:: *hao*A'	This study
pCR-P	pCR2.1-TOPO::*pmo*C1 promoter	This study
pRK-0559-1280	pRK415:: *nap*C- *hao*A'	This study
pRK-1280	pRK415:: *hao*A'	This study
pBBR-P	pBBR1MCS-2::*pmo*C1 promoter	This study
pBBR-0559-1280	pBBR1MCS-2:: *nap*C- *hao*A'	This study
pBBR-1280	pBBR1MCS-2:: *hao*A'	This study
pBBR-P-0559-1280	pBBR-0559-1280 with *M. capsulatus* Bath *pmo*C1 promoter	This study
pBBR-P-1280	pBBR-1280 with *M. capsulatus* Bath *pmo*C1 promoter	This study

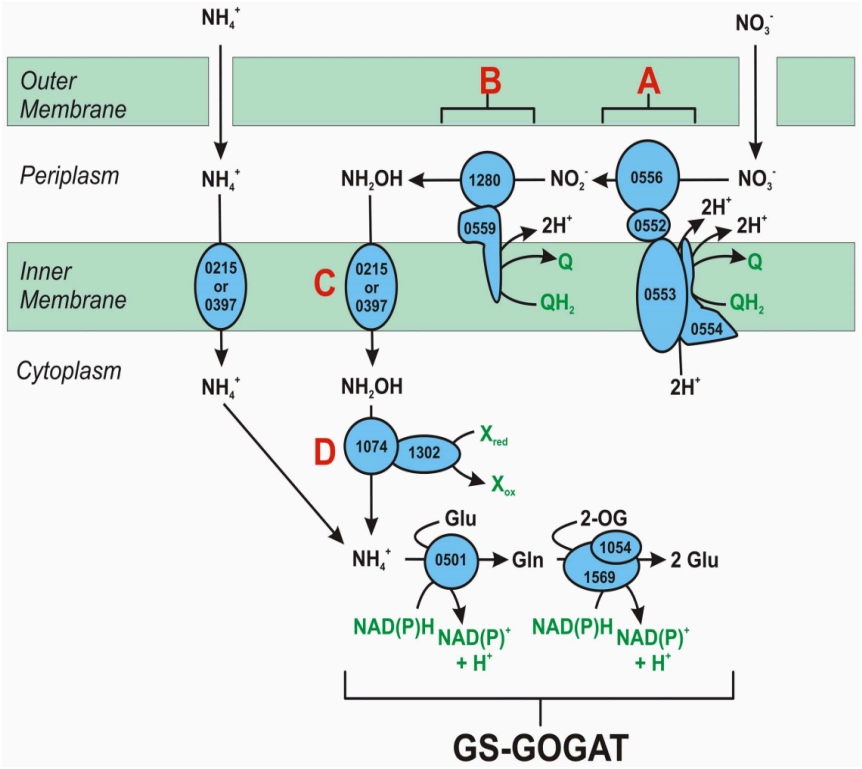


Figure 5.3.1: The reverse-HURM pathway in *N. profundicola*.
Individual steps are noted with capital letters in the figure. Numbers in the depicted proteins refer to locus tags from the AmH genome (i.e., 1280 = NAMH_1280 = HaoA). (A) Nitrate reduction by respiratory periplasmic nitrate reductase (NapABGH, 0556-53), (B) HURM (Hydroxylamine:Ubiquinone Redox Module) comprised of a reversely operating hydroxylamine oxidoreductase (HaoA, 1280) functioning as an ocataheme nitrite reductase, that produces hydroxylamine utilizing electrons donated from a tetra-heme cytochrome c in the NapC/ NrfH/cM552 family (CycB, 0559). (C) Transport of hydroxylamine via ammonia transporters related to AmtB. (D) Reduction of hydroxylamine to ammonium by a hybrid cluster protein/hydroxylamine reductase (Hcp/Har, 1074) utilizing reducing power from a predicted Fe-S containing protein (1302) whose electron donor (X) is currently unknown. Assimilation of ammonium into biomass occurs via glutamine synthetase and glutamine:2-oxoglutarate aminotransferase (GS-GOGAT). (Hanson et al., 2013)

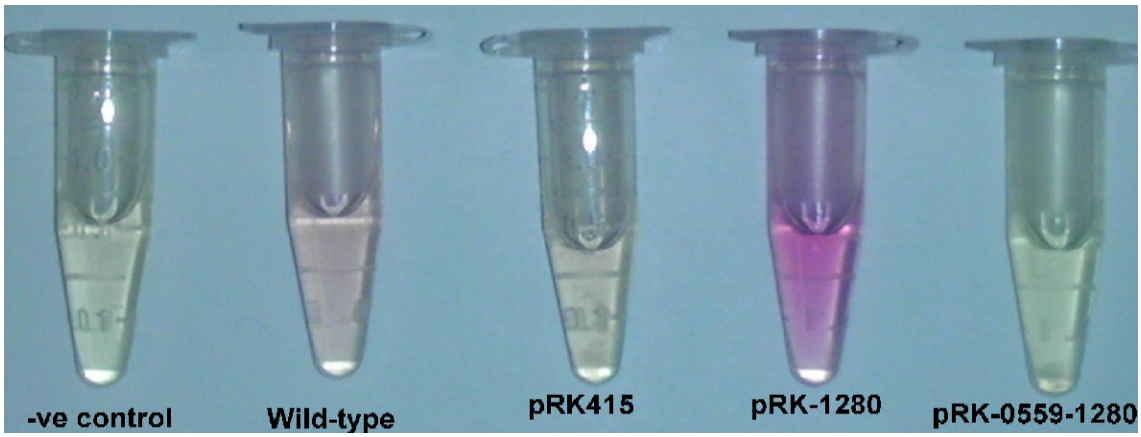


Figure 5.3.2: Preliminary testing of nitrite reduction cassette in *P. chlororaphis* O6. Presence of nitrite is indicated by the pink color, as tested with 1% sulphanilamide and 0.02% N-1 napthylethylene diamine dihydrochloride. –ve control: no bacteria; Wild-type: strain without plasmid; pRK415: strain containing pRK415 plasmid only; pRK-1280: strain containing haoA' only in cloned in pRK415; pRK-0559-1280: strain containing the nitrite reduction cassette cloned into pRK415.

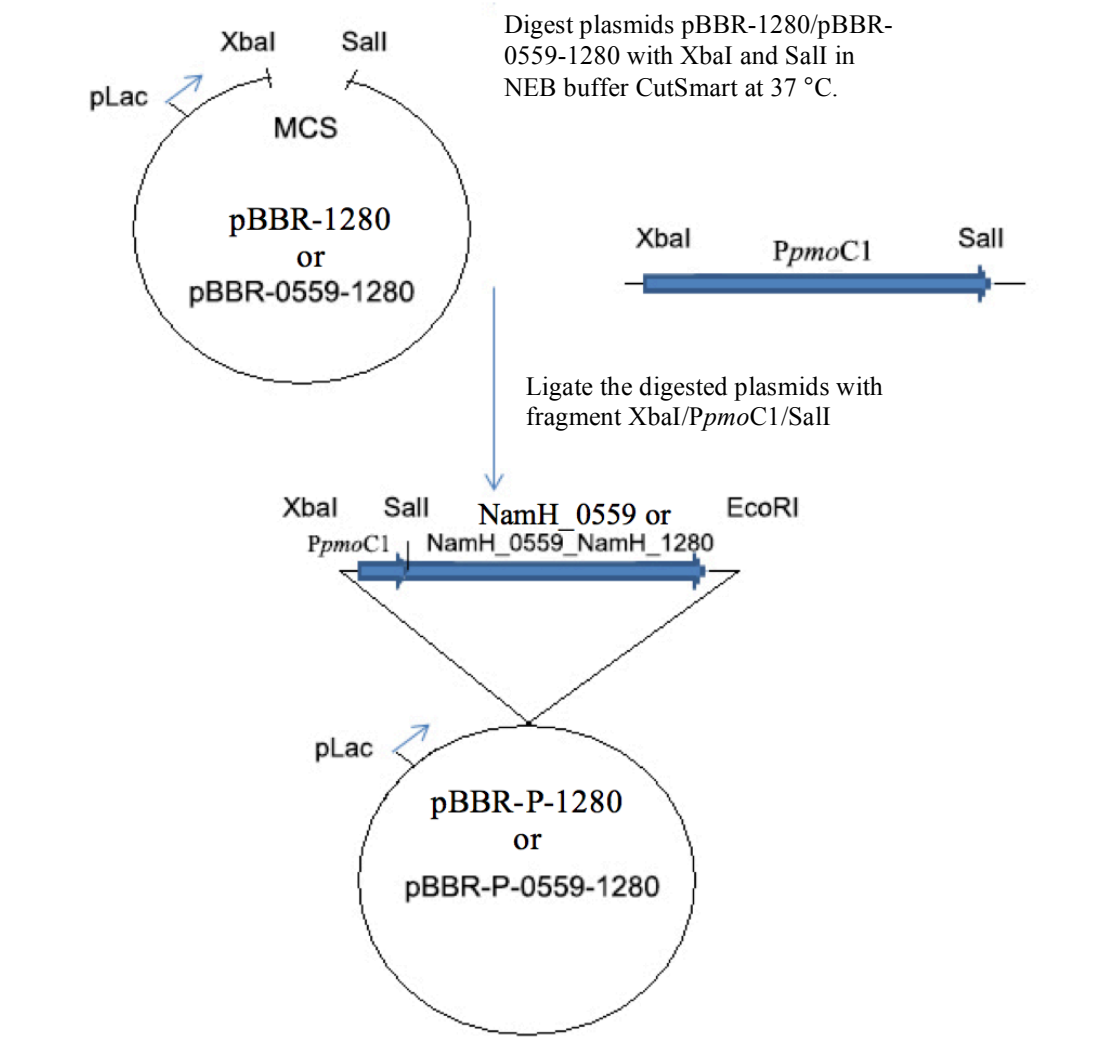


Figure 5.3.3: Workflow of generation of constructs pBBR-P-1280 and pBBR-P-0559-1280.

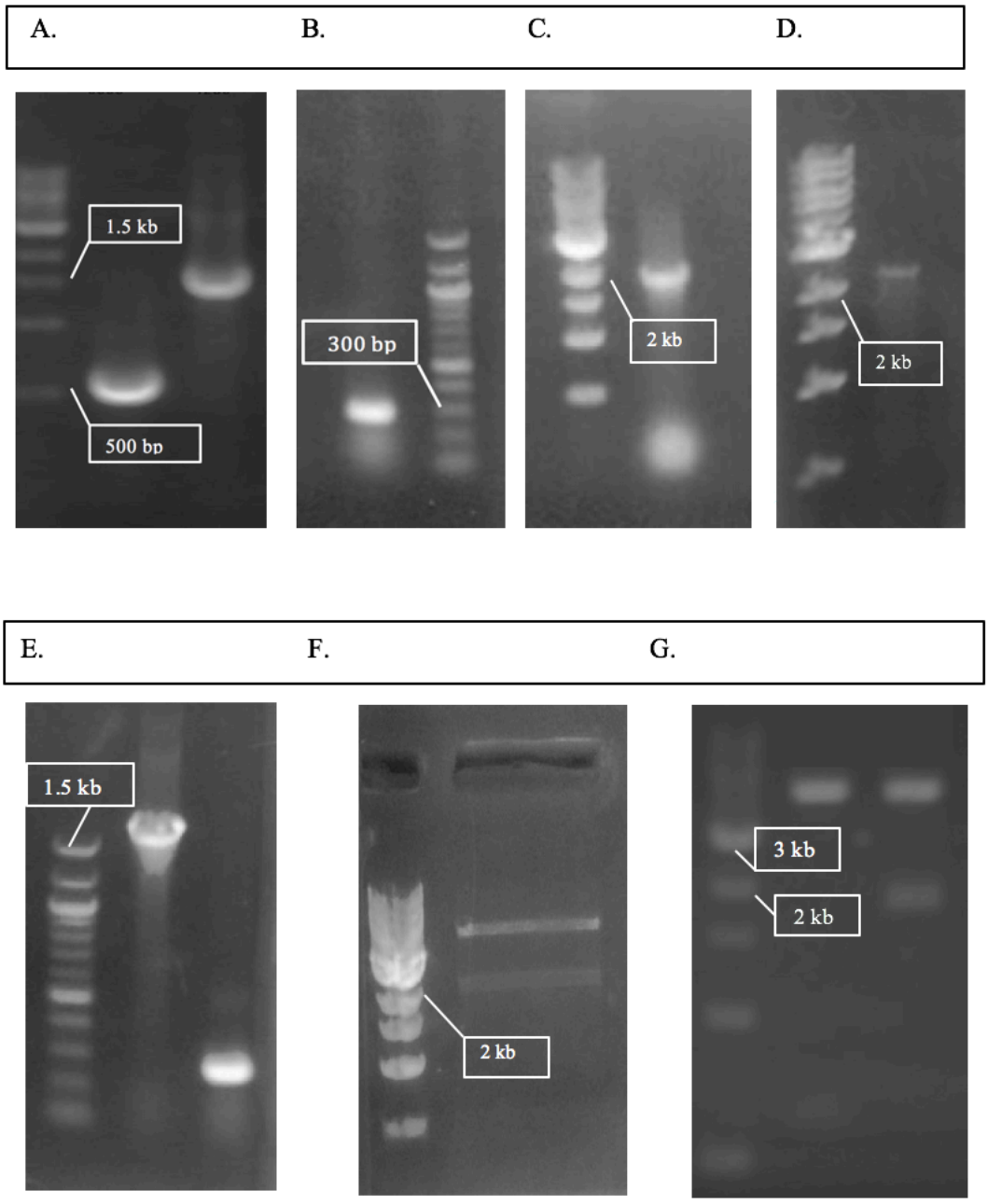


Figure 5.3.4: Agarose-gel electrophoresis and restriction analyses of expression construct pBBR-P-1280 and pBBR-P-0559-1280. A: PCR products of NAMH-0559 (second lane) and NAMH-1280 (third lane) using pC0559-1280 as template. B: PCR product of P*pmo*C1 (left lane) with primers set *pmo*C1F/R using pBBR-P*pmo*C1 as template. C: PCR product of NAMH0559-NAMH1280 (right lane) with primers set NAMH0559F/NAMH1280R using pBBR-P-0559-1280 as template. D: PCR product of P*pmo*C1-NAMH0559-NAMH1280 (right lane) with primers set *pmo*C1F/NAMH1280R using pBBR-P-0559-1280 as template. E: PCR product of NAMH1280 (second lane)

with primers set *pmo*C1F/NAMH1280R and P*pmo*C1 (third lane) with primers set *pmo*C1F/R using pBBR-P-1280 as template. F: Restriction analysis of pBBR-P-0559-1280 with XbaI/EcoRI (NEB). G: Restriction analysis of pBBR-P-1280 with SacI/XbaI (NEB) to linearize the construct (second lane) and XbaI/EcoRI (NEB; third lane).

CHAPTER 6. SUMMARY AND PERSPECTIVES

The elemental biogeochemical nitrogen cycle involves biological transformation of inorganic nitrogen compounds that exist in broad oxidation states ranging from -3 to +5 in all ecosystems facilitated by a plethora of specialized enzymes primarily carried out by microbes (Stein and Klotz 2016). Transformations of these nitrogen compounds are usually coupled with electron transfers, which have the potential to provide sources for nitrogen, energy and reducing power needed for chemolitho(auto)trophic growth. Over the decades, human activities such as excessive use of industrially produced nitrogen fertilizers and fossil fuel combustion have dramatically accelerated the input of fixed nitrogen into the nitrogen cycle, thereby affecting the diversity, abundance and distribution of microorganisms involved in the cycle. Ammonia-oxidizing bacteria (AOB) are able to extract energy and reductant in the process of oxidizing ammonia to nitrite, which is the first step of nitrification. Our understanding of the physiological ecology and molecular basis of obligate chemolithotrophy in AOB is still limited because of the difficulties of cultivating a broad diversity of AOB in the laboratory, although metagenomics studies have begun to provide insights into the structure of AOB communities.

To date, *Nitrosococcus oceani, Nitrosococcus halophilus* and *Nitrosococcus watsonii* are the only three species of gammaproteobacterial AOB maintained in pure

culture that are restricted to marine or saline pond environments (Purkhold et al., 2000). In this study, we successfully isolated into pure culture a new strain (D1FHS) of *Nitrosococcus* from the enrichment culture of sediment sampled from Jiaozhou Bay, China, using serial dilution and semi-solid agar methods. D1FHS was assigned into a new species: *Nitrosococcus wardiae* is classified in the genus *Nitrosococcus*, family *Chromatiaceae*, order *Chromatiales*, class *Gammaproteobacteria,* as the type strain of the species. The isolated strain was purified and cultivated in hydroxylamine-amended ammonium minimum salts (AMS) media (Chapter 2). Strain D1FHS was able to grow at pH 5~9, was tolerant to salt at a concentration up to 1600 mM and its tolerance threshold for ammonium, its sole substrate, was at 300 mM (Koops et al., 1990, Campbell et al., 2011a). Growth of D1FHS was observed at a temperature as high as 45°C with the optimum temperature of 37°C for nitrite production, which is considerably higher than was has been reported for other AOB in the literature (Wang et al., 2016). All of these physiological characteristics featured D1FHS as a strain with unique characteristics when compared with other known *Nitrosococcus* species.

The genome of strain D1FHS was sequenced using the PacBio sequencing technique at the University of Delaware sequencing center. A draft genome was assembled, annotated and analyzed using various bioinformatics software packages. The analysis of its genome provided insight at the gene level into mechanisms that determine the distinct metabolism that enables it to thrive as an ammonia-dependent chemolithotrophic microbe in marine ecosystems. The genome of strain D1FHS contains genes encoding a variety of proteins implicated in the assimilation of nitrogen, carbon, and phosphate. An unexpected finding was the presence of a cyanase gene (*cyn*) in the

genome of D1FHS, which constitutes the first identification of an AOB genome with the genetic potential to encode cyanate hydrolase (EC 4.2.1.104). Physiological and steady-state mRNA level studies indicated that the cyanate hydrolase encoded in the genome of D1FHS is functionally expressed. We also compared four interspecies genomes in genus *Nitrosococcus*: *Nitrosococcus oceani* C-107 (=ATCC 19707=JCM30415) [GCA_000012805.1] (Klotz et al., 2006)*, Nitrosococcus watsonii* C113 [GCA_000143085.1] (Campbell et al., 2011a) *and Nitrosococcus halophilus* Nc4 [GCA_000024725.1] (Campbell et al., 2011a). Comparative analysis of content and organization of the genomes suggests that the *Nitrosococcus* genus is constituted of two eco-genotypes: *N. oceani & N. watsonii* and *N. wardiae & N. halophilus.* The *N. oceani / N. watsonii* eco-genotype is (passive) ammonium/ammonia transporter-negative; in contrast, the *N. wardiae /N. halophilus* eco-genotype is ammonium/ ammonia transport-positive. While the *N. oceani / N. watsonii* genotype is urea-lytic (encoding urea hydrolase) and urea transport-positive (encoding Utp), the *N. wardiae / N. halophilus* genotype is non-urea-lytic (lacks the genetic basis for either form of urea-hydrolytic enzymes) and urea transport-negative. Like the genome of *N. halophilus* Nc4, the D1FHS genome encodes a complete oxidation pipeline of methane to CO_2 (Ward et al., 2004b), which is based on a XoxF-type methanol dehydrogenase (Wu et al., 2015) and the methyl-tetrahydromethanopterin (THMPT) pathway (Chistoserdova et al., 2003). This key inventory is not encoded in genomes of the *N. oceani* (Poret-Peterson et al., 2008) and *N. watsonii* (Campbell et al., 2011a). There are more features differentiated the two genotypes as described in detail in chapter 3.

Together with Dr. Chee Kent Lim, we comparatively analyzed the high quality draft genome sequences of four *Nitrosococcus oceani* strains: AFC27, AFC132, C-27 and the type strain C-107 (=ATCC 19707; JMC30415) with the goal to identify indicators for autochthonous or allochthonous origins of the species. The results suggest that AFC132 is a metabolically more diverse ancestral lineage to the other strains with C-107 potentially being the youngest. Assessment of their genomes will contribute to a better understanding of the molecular evolution of genes and their expression products in microorganisms with key roles in the global nitrogen cycle.

In order to investigate the molecular basis of the ammonia-dependent chemolithotrophic lifestyle in AOB by asking how AOB extract electrons from hydroxylamine by oxidation and relay them to the quinone pool to generate proton motive force that drives the synthesis of ATP, a hydroxylamine:ubiquinone redox module (HURM) consisting of hydroxylamine dehydrogenase (HAO, *hao*A) connected to a cytochrome *c* quinone reductase (c_M552, *cyc*B) was proposed previously (Klotz and Stein 2008, Simon and Klotz 2013). It is still not yet clear whether HAO and c_M552 are interacting directly in an oxidative forward direction to reduce the quinone pool or connect via an additional cytochrome *c* protein electron shuttle such as *c*554 (*cyc*A) (Klotz et al., 2008, Stein et al., 2013). Moreover, the function of gene *hao*B in the *haoA-orf2[haoB]-cycAB* gene cluster and the stem loop in the 5'-end of the *haoB* gene are yet to be determined. Therefore, it is imperative to evaluate the function of the key genes, *hao*A, *hao*B, *cyc*A, *cyc*B and their expression products in context with the structure between *hao*A and *hao*B capable of forming a stem loop. Quantitative PCR analyses of steady-state mRNA levels under various treatments (e.g. ammonium and temperature)

revealed that the genes of interest and additional genes that encode proteins involved in denitrification and N-oxide detoxification were differentially expressed with the external supply of ammonium. Preliminary data lead us to propose that HaoB interacts with the HAO complex leading to less nitrite formation by increased release of NO from HAO into the periplasm. Interestingly, *Nautilia profundicola* encodes a modified HaoA protein capable of forming an alternative HAO' complex (HAO without a tyrosine protein ligand) that was predicted to operate reversely to generate hydroxylamine and ammonium from nitrite (Campbell et al., 2009, Hanson et al., 2013, Simon and Klotz 2013) for both nitrogen assimilation and respiration. This hypothesis has been experimentally verified (Hanson et al., 2013). With the modern molecular approaches such as end-point PCR, cloning and mutation, we have successfully created the desired constructs and host mutants to explore the functions of the genes and enzymes involved in these processes.

Studies in this dissertation have created a diversity of interesting and pertinent questions for further investigation. The interaction between the HAO complex and the quinone/quinol pool and experimentally quantifying under which conditions the "reverse-HURM" pathway in *Nautilia profundicola* produces what ratio of hydroxylamine and ammonium by employing molecular biology, physiology and biochemical experimentation, will help to reveal the precise pathway of electron transfer during ammonia metabolism in nitrifying bacteria. High throughput sequencing and bioinformatics analyses are needed to fully elucidate the molecular inventory that drives cellular functions involved in evolutionary niche differentiation and adaptation of AOB.

REFERENCES

Altschul, S. F., et al. (1997). "Gapped blast and psi-blast - a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389 - 3402.

Alzerreca, J. J., et al. (1999). "The amo operon in marine, ammonia-oxidizing gamma-proteobacteria." FEMS Microbiol Lett **180**(1): 21-29.

Andersson, K. K., et al. (1991). "P460 of hydroxylamine oxidoreductase of nitrosomonas europaea: Soret resonance raman evidence for a novel heme-like structure." Biochem Biophys Res Commun **174**(1): 358-363.

Arciero, D., et al. (1991a). "Spectroscopic and rapid kinetic studies of reduction of cytochrome *c*554 by hydroxylamine reductase from *nitrosomonas europaea*." J. Biol. Chem. **269**: 11878-11886.

Arciero, D. M., et al. (1991b). "Resolution of the four hemes of cytochrome *c*554 from *nitrosomonas europaea* by redox potentiometry and optical spectroscopy." Biochemistry **30**: 11459-11465.

Arciero, D. M. and Hooper, A. B. (1997). "Evidence for a crosslink between *c*-heme and a lysine residue in cytochrome p460 of *nitrosomonas europaea*." FEBS Letters **410**: 457-460.

Arciero, D. M., et al. (2002a). "Nitrosocyanin, a red cupredoxin-like protein from nitrosomonas europaea." Biochemistry **41**(6): 1703-1709.

Arciero, D. M., et al. (2002b). "Nitrosocyanin, a red cupredoxin-like protein from *nitrosomonas europaea*." Biochemistry **41**(6): 1703-1709.

Arp, D. J., et al. (2007). "The impact of genome analyses on our understanding of ammonia-oxidizing bacteria." Annu Rev Microbiol **61**: 503-528.

Arp, D. J., et al. (2002). "Molecular biology and biochemistry of ammonia oxidation by *nitrosomonas europaea*." Archives of Microbiology **178**(4): 250-255.

Aziz, R., et al. (2008). "The rast server: Rapid annotations using subsystems technology." BMC Genomics **9**(1): 75.

Barrangou, R., et al. (2007). "Crispr provides acquired resistance against viruses in prokaryotes." Science **315**(5819): 1709 - 1712.

Barz, M., et al. (2010). "Distribution analysis of hydrogenases in surface waters of marine and freshwater environments." PLoS One **5**(11): e13846.

Bergmann, D. J. and Hooper, A. B. (2003a). "Cytochrome p460 of nitrosomonas europaea." European Journal of Biochemistry **270**: 1935-1941.

Bergmann, D. J. and Hooper, A. B. (2003b). "Cytochrome p460 of nitrosomonas europaea. Formation of the heme-lysine cross-link in a heterologous host and mutagenic conversion to a non-cross-linked cytochrome c'." Eur J Biochem **270**(9): 1935-1941.

Bergmann, D. J., et al. (2005). "Structure and sequence conservation of hao cluster genes of autotrophic ammonia-oxidizing bacteria: Evidence for their evolutionary history." Appl Environ Microbiol **71**(9): 5371-5382.

Bernhard, A. E., et al. (2005). "Loss of diversity of ammonia-oxidizing bacteria correlates with increasing salinity in an estuary system." Environ Microbiol **7**(9): 1289-1297.

Bertram, R. and Schuster, C. F. (2014). "Post-transcriptional regulation of gene expression in bacterial pathogens by toxin-antitoxin systems." Front Cell Infect Microbiol **4**: 6.

Blin, K., et al. (2013). "Antismash 2.0--a versatile platform for genome mining of secondary metabolite producers." Nucleic Acids Res **41**(Web Server issue): W204-212.

Bollmann, A., et al. (2013). "Complete genome sequence of nitrosomonas sp. Is79, an ammonia oxidizing bacterium adapted to low ammonium concentrations." Stand Genomic Sci **7**(3): 469-482.

Buchanan, R. E. (1925). General systematic bacteriology. Baltimore, The Williams and Wilkins Co.

Buist, G., et al. (2008). "Lysm, a widely distributed protein motif for binding to (peptido)glycans." Mol Microbiol **68**(4): 838-847.

Camargo, J. A. and Alonso, A. (2006). "Ecological and toxicological effects of inorganic nitrogen pollution in aquatic ecosystems: A global assessment." Environ Int **32**(6): 831-849.

Campbell, B. J., et al. (2009). "Adaptations to submarine hydrothermal environments exemplified by the genome of nautilia profundicola." PLoS Genet **5**(2): e1000362.

Campbell, M. A., et al. (2011a). "Nitrosococcus watsonii sp. Nov., a new species of marine obligate ammonia-oxidizing bacteria that is not omnipresent in the world's oceans: Calls to validate the names 'nitrosococcus halophilus' and 'nitrosomonas mobilis'." FEMS Microbiol Ecol **76**(1): 39-48.

Campbell, M. A., et al. (2011b). "Model of the molecular basis for hydroxylamine oxidation and nitrous oxide production in methanotrophic bacteria." FEMS Microbiol Lett **322**(1): 82-89.

Carver, T. J., et al. (2005). "Act: The artemis comparison tool." Bioinformatics. **21**: 3422-3423.

Cavanaugh, C. M. and Robinson, J. J. (1996). CO_2 fixation in chemoautotroph-invertebrate symbioses: Expression of form i and form ii rubisco. Microbial growth on c1 compounds. Lidstrom, M. E. and Tabita, F. R. Netherlands, Kluwer Academic Publishers**:** 285-292.

Chain, P., et al. (2003). "Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph nitrosomonas europaea." Journal of Bacteriology **185**(9): 2759-2773.

Chapman, M., et al. (2002). "Role of escherichia coli curli operons in directing amyloid fiber formation." Science **295**: 851 - 855.

Chin, C. S., et al. (2013). "Nonhybrid, finished microbial genome assemblies from long-read smrt sequencing data." Nat Methods **10**(6): 563-569.

Chistoserdova, L., et al. (2003). "Methylotrophy in *methylobacterium extorquens* am1 from a genomic point of view." J Bacteriol **185**(10): 2980 - 2987.

Christinet, L., et al. (2004). "Characterization and functional identification of a novel plant 4,5-extradiol dioxygenase involved in betalain pigment biosynthesis in portulaca grandiflora." Plant Physiol **134**(1): 265-274.

Coleman, M. L. and Chisholm, S. W. (2010). "Ecosystem-specific selection pressures revealed through comparative population genomics." Proc Natl Acad Sci U S A **107**(43): 18634-18639.

Commission (1958). "Opinion 23. Rejection of the generic names *nitromonas* winogradsky 1890 and *nitromonas* orla-jensen 1909, conservation of the generic names *nitrosomonas* winogradsky 1892, *nitrosococcus* winogradsky 1892, and the designation of the type species of these genera." Int. Bull. Bacteriol. Nomencl. Taxon. **8**: 169-170.

Daims, H., et al. (2015). "Complete nitrification by nitrospira bacteria." Nature **528**(7583): 504-509.

Dang, H., et al. (2009a). "Diversity and spatial distribution of *amoa*-encoding archaea in the deep-sea sediments of the tropical west pacific continental margin." Journal of Applied Microbiology **106**(5): 1482-1493.

Dang, H., et al. (2009b). "Diversity and distribution of sediment nirs -encoding bacterial assemblages in response to environmental gradients in the eutrophied jiaozhou bay, china." Microbial Ecology **DOI** 10.1007/s00248-008-9469-5.

Dang, H., et al. (2010a). "Environmental factors shape sediment anammox bacterial communities in hypernutrified jiaozhou bay, china." Appl Environ Microbiol **76**(21): 7036-7047.

Dang, H., et al. (2010b). "Diversity, abundance, and spatial distribution of sediment ammonia-oxidizing betaproteobacteria in response to environmental gradients and coastal eutrophication in jiaozhou bay, china." Appl Environ Microbiol **76**(14): 4691-4702.

Dang, H., et al. (2010c). "Diversity, abundance and distribution of amoa-encoding archaea in deep-sea methane seep sediments of the okhotsk sea." FEMS Microbiol Ecol **72**(3): 370-385.

Dang, H., et al. (2011). "Molecular characterization of putative biocorroding microbiota with a novel niche detection of epsilon- and zetaproteobacteria in pacific ocean coastal seawaters." Environ Microbiol **13**(11): 3059-3074.

Darling, A. C. E., et al. (2004). "Mauve: Multiple alignment of conserved genomic sequence with rearrangements." Genome Research **14**(7): 1394-1403.

Darling, A. E., et al. (2010). "Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement." PLoS One **5**(6): e11147.

De Carvalho, C. C. and Fernandes, P. (2010). "Production of metabolites as bacterial responses to the marine environment." Mar Drugs **8**(3): 705-727.

Dhillon, B. K., et al. (2015). "Islandviewer 3: More flexible, interactive genomic island discovery, visualization and analysis." Nucleic Acids Res **43**(W1): W104-108.

Dorr, T., et al. (2010). "Ciprofloxacin causes persister formation by inducing the tisb toxin in escherichia coli." PLoS Biol **8**(2): e1000317.

Drummond, A. J., et al. (2012). "Bayesian phylogenetics with beauti and the beast 1.7." Molecular Biology and Evolution.

Editorial_Board (1955). "Status of the generic names *nitromonas, nitrosomonas, nitrosococcus* and *nitrobacter*: Editorial board preliminary opinion." Int. Bull. Bacteriol. Nomencl. Taxon. **5**: 27-31.

Eid J, F. A., et al. (2009). "Real-time DNA sequencing from single polymerase molecules." Science **323**(5910): 133-138.

Eijkelkamp, B. A., et al. (2011). "Investigation of the human pathogen acinetobacter baumannii under iron limiting conditions." BMC Genomics **12**: 126.

El Rayes, E. G., et al. (1991). "Methanol metabolism and ammonia assimilation in four methylophilus strains." Acta Biotechnologica **11**(2): 87-93.

El Sheikh, A. F. and Klotz, M. G. (2008). "Ammonia-dependent differential regulation of the gene cluster that encodes ammonia monooxygenase in nitrosococcus oceani atcc 19707." Environ Microbiol **10**(11): 3026-3035.

El Sheikh, A. F., et al. (2008). "Characterization of two new genes, amor and amod, in the amo operon of the marine ammonia oxidizer nitrosococcus oceani atcc 19707." Appl Environ Microbiol **74**(1): 312-318.

Elmore, B. O., et al. (2007). "Cytochromes p460 and c'-beta; a new family of high-spin cytochromes c." FEBS Lett **581**(5): 911-916.

Fitch, W. and Upper, K. (1987). "The phylogeny of trna sequences provides evidence fpr ambiguity reduction in the origin of the genetic code." Cold Spring Harb Quant Biol **52**: 759 - 767.

Furuta, Y., et al. (2010). "Genome comparison and context analysis reveals putative mobile forms of restriction-modification systems and related rearrangements." Nucleic Acids Res **38**(7): 2428-2443.

Graham, J. E., et al. (2011). "Characterizing bacterial gene expression in nitrogen cycle metabolism with rt-qpcr." Methods Enzymol **496**: 345-372.

Garcia, J. C., et al. (2013). "Draft genome sequence of nitrosospira sp. Strain apg3, a psychrotolerant ammonia-oxidizing bacterium isolated from sandy lake sediment." Genome Announc **1**(6).

Goris, J., et al. (2007). "DNA-DNA hybridization values and their relationship to whole-genome sequence similarities." Int J Syst Evol Microbiol **57**: 81-91.

Grissa, I., et al. (2007). "Crisprfinder: A web tool to identify clustered regularly interspaced short palindromic repeats." Nucleic Acids Res **35**(Web Server issue): W52-57.

Gunsalus, R. P. and Wang, H. (2000). "The *nrfa* and *nirb* nitrite reductase operons in *escherichia coli* are expressed differently in response to nitrate than to nitrite." J. Bacteriol. **182**: 5813-5822.

Gurevich, A., et al. (2013). "Quast: Quality assessment tool for genome assemblies." Bioinformatics **29**(8): 1072-1075.

Gutekunst, K., et al. (2014). "The bidirectional nife-hydrogenase in synechocystis sp. Pcc 6803 is reduced by flavodoxin and ferredoxin and is essential under mixotrophic, nitrate-limiting conditions." J Biol Chem **289**(4): 1930-1937.

Head, I. M., et al. (1993). "The phylogeny of autotrophic ammonia-oxidizing bacteria as determined by analysis of 16s ribosomal rna gene sequences." J. Gen. Microbiol. 139: 1147-1153.

Haft, D. H., et al. (2005). "A guild of 45 crispr-associated (cas) protein families and multiple crispr/cas subtypes exist in prokaryotic genomes." PLoS Comput Biol **1**(6): e60.

Hansen, P. A. and Nielsen, V. (1939). "Colorimertric determination of ammonia with thymol-hypobromite reagent."

Hanson, T. E., et al. (2013). "Nitrate ammonification by nautilia profundicola amh: Experimental evidence consistent with a free hydroxylamine intermediate." Front Microbiol **4**: 180.

Hatzenpichler, R., et al. (2008). "A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring." Proc. Natl. Acad. Sci. USA **105**(6): 2134-2139.

Head, I. M., et al. (1993a). "The phylogeny of autotrophic ammonia-oxidizing bacteria as determined by analysis of 16s ribosomal rna gene sequences." J. Gen. Microbiol. **13**: 1147-1153.

Head, I. M., et al. (1993b). "The phylogeny of autotrophic ammonia-oxidizing bacteria as determined by analysis of 16s ribosomal rna gene sequences." J. Gen. Microbiol. **139**: 1147-1153.

Hentzer, M., et al. (2002). "Inhibition of quorum sensing in pseudomonas aeruginosa biofilm bacteria by a halogenated furanone compound." Microbiology **148**(Pt 1): 87-102.

Higuchi, R., et al. (1992). "Simultaneous amplification and detection of specific DNA sequences." Biotechnology (N Y) **10**(4): 413-417.

Higuchi, R., et al. (1993). "Kinetic pcr analysis: Real-time monitoring of DNA amplification reactions." Biotechnology (N Y) **11**(9): 1026-1030.

Holmes, A. J., et al. (1995). "Evidence that particulate methane monooxygenase and ammonia monooxygenase may be evolutionarily related." FEMS Microbiology Letters **132**: 203-208.

Hooper, A. B., et al. (2005). The oxidation of ammonia as an energy source in bacteria in respiration. Respiration in archaea and bacteria: Diversity of procaryotic respiratory systems Zannoni, D. Dordrecht, The Netherlands, Springer. **2:** 121-147.

Hooper, A. B., et al. (1990). Production of nitrite and $n_2o$by the ammonia-oxidizing nitrifiers. Nitrogen fixation: Achievements and objectives. Gresshoff, R., Stacey & Newton. New York • London, Chapman and Hall. **1:** 387-391.

Hooper, A. B. and Terry, K. R. (1979). "Hydroxylamine oxidoreductase of *nitrosomonas:* Production of nitric-oxide from hydroxylamine." Biochimica et Biophysica Acta **571**(1): 12-20.

Hooper, A. B., Vannelli, T., Bergmann, D. J., and Arciero, D. M. (1997). "Enzymology of the oxidation of ammonia to nitrite by bacteria." Antonie Van Leeuwenhoek **71**: 56-67.

Horvath, P. and Barrangou, R. (2010). "Crispr/cas, the immune system of bacteria and archaea." Science **327**(5962): 167-170.

Hyman, M. R. and Wood, P. M. (1983). "Methane oxidation by nitrosomonas europaea." Biochem J **212**(1): 31-37.

Iverson, T. M., et al. (2001). "High-resolution structures of the oxidized and reduced states of cytochrome c554 from nitrosomonas europaea." J Biol Inorg Chem **6**(4): 390-397.

Johnson, C. H., et al. (1996). "Circadian clocks in prokaryotes." Mol Microbiol **21**(1): 5-11.

Kamennaya, N. A. and Post, A. F. (2013). "Distribution and expression of the cyanate acquisition potential among cyanobacterial populations in oligotrophic marine waters." Limnol. Oceanogr. **58**: 1959–1971

Kern, M., et al. (2011). "The wolinella succinogenes mcc gene cluster encodes an unconventional respiratory sulphite reduction system." Mol Microbiol **82**(6): 1515-1530.

Kerscher, S., et al. (2008). The three families of respiratory nadh dehydrogenases. Bioenergetics. Schäfer, G. and Penefsky, H. S. Berlin, Heidelberg, Springer**:** 185-222.

Kim, H. J., et al. (2008). "Membrane tetraheme cytochrome c(m552) of the ammonia-oxidizing nitrosomonas europaea: A ubiquinone reductase." Biochemistry **47**(25): 6539-6551.

Klotz, M. G., et al. (2006). "Complete genome sequence of the marine, chemolithoautotrophic, ammonia-oxidizing bacterium nitrosococcus oceani atcc 19707." Appl Environ Microbiol **72**(9): 6299-6315.

Klotz, M. G. and Norton, J. M. (1998). "Multiple copies of ammonia monooxygenase (amo) operons have evolved under biased at/gc mutational pressure in ammonia-oxidizing autotrophic bacteria." FEMS Microbiol Lett **168**(2): 303-311.

Klotz, M. G., et al. (2008). "Evolution of an octahaem cytochrome c protein family that is key to aerobic and anaerobic ammonia oxidation by bacteria." Environ Microbiol **10**(11): 3150-3163.

Klotz, M. G. and Stein, L. Y. (2011). Genomics of ammonia-oxidizing bacteria and insights to their evolution. Washington, DC., ASM Press,.

Koch, H., et al. (2014). "Growth of nitrite-oxidizing bacteria by aerobic hydrogen oxidation." Science **345**(6200): 1052–1054.

Koch, H., et al. (2015). "Expanded metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus nitrospira." Proc Natl Acad Sci U S A **112**(36): 11371-11376.

Konneke, M., et al. (2005). "Isolation of an autotrophic ammonia-oxidizing marine archaeon." Nature **437**(7058): 543-546.

Konstantinidis, K. T., et al. (2006). "The bacterial species definition in the genomic era." Philos Trans R Soc Lond B Biol Sci **361**(1475): 1929-1940.

Konstantinidis, K. T. and Tiedje, J. M. (2005). "Genomic insights that advance the species definition for prokaryotes." PNAS **102**(7): 2567-2572.

Koops, H.-P., et al. (1976). "Isolation of a moderate halophilic ammonia-oxidizing bacterium, *nitrosococcus mobilis* nov. Sp. ." Archives of Microbiology **107**: 277-282.

Koops, H.-P., et al. (1990). "Description of a new species of *nitrosococcus* " Archives of Microbiology **154**(3): 244-248.

Koper, T. E., et al. (2004). "Urease-encoding genes in ammonia-oxidizing bacteria." Applied and Environmental Microbiology **70**(4): 2342-2348.

Kostera, J., et al. (2010). "Enzymatic interconversion of ammonia and nitrite: The right tool for the job." Biochemistry **49**(39): 8546-8553.

Kostera, J., et al. (2008). "Kinetic and product distribution analysis of no· reductase activity in *nitrosomonas europaea* hydroxylamine oxidoreductase." Journal of Biological Inorganic Chemistry **13**(7): 1073-1083.

Kovach, M. E., et al. (1995). "Four new derivatives of the broad-host-range cloning vector pbbr1mcs, carrying different antibiotic-resistance cassettes." Gene **166**: 175-176.

Kowalchuk, G. A. and Stephen, J. R. (2001). "Ammonia-oxidizing bacteria: A model for molecular microbial ecology." Annu Rev Microbiol **55**: 485-529.

Kozlowski, J. A., et al. (2016). "Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and thaumarchaeota." ISME J **10**(8): 1836-1845.

Kumagai, H., et al. (1997). "Membrane localization, topology, and mutual stabilization of the *rnfabc* gene products in *rhodobacter capsulatus* and implications for a new family of energy-coupling nadh oxidoreductases." Biochemistry **36**(18): 5509-5521.

Kuypers, M. M., Lavik, G., Woebken, D., Schmid, M., Fuchs, B. M., Amann, R., Et Al. (2005). "Massive nitrogen loss from the benguela upwelling system through anaerobic ammonium oxidation." Proc. Natl. Acad. Sci. U.S.A. **102**: 6478–6483.

Kuypers, M. M., Sliekers, A. O., Lavik, G., Schmid, M., Jørgensen, B. B., Kuenen, J. G., Et Al. (2003). "Anaerobic ammonium oxidation by anammox bacteria in the black sea." Nature **422**: 608-611.

Lampson, B. C., et al. (2005). "Retrons, msdna, and the bacterial genome." Cytogenet Genome Res **110**(1-4): 491-499.

Lauro, F. M., et al. (2014). "Ecotype diversity and conversion in photobacterium profundum strains." PLoS One **9**(5): e96953.

Lawton, T. J., et al. (2009). "Crystal structure of a two-domain multicopper oxidase: Implications for the evolution of multicopper blue proteins." J. Biol. Chem. %R 10.1074/jbc.M900179200: M900179200.

Leigh, J. A. (2000). "Nitrogen fixation in methanogens: The archaeal perspective."

Li, H., et al. (2010). "Molecular detection of anaerobic ammonium-oxidizing (anammox) bacteria in high-temperature petroleum reservoirs." Microbial Ecology **Online Early**: 1-13.

Li, W., et al. (2015). "The embl-ebi bioinformatics web and programmatic tools framework." Nucleic Acids Res **43**(W1): W580-584.

Limpiyakorn, T., et al. (2005). "Communities of ammonia-oxidizing bacteria in activated sludge of various sewage treatment plants in tokyo." FEMS Microbiol Ecol **54**(2): 205-217.

Lipschultz, F., et al. (1981). "Production of no and n_2o by soil nitrifying bacteria." Nature **294**: 641-643.

Livak, K. J. a. T. D. S. (2001). "Analysis of relative gene expression data using real-time quantitative pcr and the 2(-delta delta c(t)) method. ." Methods **25**: 402-408.

Loper, J. E., et al. (2012). "Comparative genomics of plant-associated pseudomonas spp.: Insights into diversity and inheritance of traits involved in multitrophic interactions." PLoS Genet **8**(7): e1002784.

Lopez-Perez, M., et al. (2013). "Genomic diversity of "deep ecotype" alteromonas macleodii isolates: Evidence for pan-mediterranean clonal frames." Genome Biol Evol **5**(6): 1220-1232.

Lucker, S., et al. (2013). "The genome of nitrospina gracilis illuminates the metabolism and evolution of the major marine nitrite oxidizer." Front Microbiol **4**: 27.

Luque-Almagro, V. M., et al. (2008). "Characterization of the pseudomonas pseudoalcaligenes cect5344 cyanase, an enzyme that is not essential for cyanide assimilation." Appl Environ Microbiol **74**(20): 6280-6288.

Lynch, M. D. and Gill, R. T. (2006). "Broad host range vectors for stable genomic library construction." Biotechnol Bioeng **94**(1): 151-158.

Levene, M. J. et al. (2003). " Zero-mode waveguides for single-molecule analysis at high concentrations." Science **299**(5607): 682–686.

Ma, J. F., et al. (1999). "Bacterioferritin a modulates catalase a (kata) activity and resistance to hydrogen peroxide in pseudomonas aeruginosa." J Bacteriol **181**(12): 3730-3742.

Magnuson, R. D. (2007). "Hypothetical functions of toxin-antitoxin systems." J Bacteriol **189**(17): 6089-6092.

Makarova, K. S., et al. (2013). "Comparative genomics of defense systems in archaea and bacteria." Nucleic Acids Res **41**(8): 4360-4377.

Oshiki, M., et al. (2016). "Hydroxylamine-dependent anaerobic ammonium oxidation (anammox) by "candidatus brocadia sinica"." Environ Microbiol **18**(9): 3133-3143.

Mcwilliam, H., et al. (2013). "Analysis tool web services from the embl-ebi." Nucleic Acids Res **41**(Web Server issue): W597-600.

Migula, W. (1900). System der bakterien. Jena, Gustav Fischer Verlag.

Murphy, R. R., et al. (2015). "Nxrepair: Error correction in de novo sequence assembly using nextera mate pairs." PeerJ **3**: e996.

Murray, R. G. E. and Watson, S. W. (1962). "Structure of *nitrocystis oceanus* and comparison with *nitrosomonas* and *nitrobacter*." Journal of Bacteriology **89**(6): 1594-1609.

Murrell, J. C. and Holmes, A. J. (1996). Molecular biology of particulate methane monooxygenase. Microbial growth on c1 compounds. Lidstrom, M. E. and Tabita, F. R. Netherlands, Kluwer Academic Publishers**:** 133-140.

Nakamura, K., et al. (2003). "Novel types of two-domain multi-copper oxidases: Possible missing links in the evolution." FEBS Lett **553**(3): 239-244.

Nevison, D. C. and Holland, E. A. (1997). "A reexamination of the impact of anthropogenically fixed nitrogen on atmospheric n_2o and the stratospheric o_3 layer. ." Journal of Geophysical Research **102**: 25519-25536.

Nicholas, D. J. D. and Nason, A. (1957). "Determination of nitrate and nitrite." Methods of Enzymology **3**: 981-984.

Nina A. Kamennaya, et al. (2008). "The cyanate utilization capacity of marine unicellular cyanobacteria." Limnology and Oceanography **53** (6): 2485-2494.

Norton, J. M., et al. (2002). "Diversity of ammonia monooxygenase operon in autotrophic ammonia-oxidizing bacteria." Arch Microbiol **177**(2): 139-149.

Norton, J. M., et al. (2008). "Complete genome sequence of nitrosospira multiformis, an ammonia-oxidizing bacterium from the soil environment." Appl Environ Microbiol **74**(11): 3559-3572.

Numata, M., et al. (1990). "Cytochrome p-460 of nitrosomonas europaea: Further purification and further characterization." J. Biochem. **108**: 1016-1023.

Oberto, J. (2013). "Synttax: A web server linking synteny to prokaryotic taxonomy." BMC Bioinformatics **14**: 4.

Olson, J. W., et al. (2001). "Requirement of nickel metabolism proteins hypa and hypb for full activity of both hydrogenase and urease in helicobacter pylori." Mol Microbiol **39**(1): 176-182.

Palatinszky, M., et al. (2015). "Cyanate as an energy source for nitrifiers." Nature **524**(7563): 105-108.

Pertea, M., et al. (2009). "Operondb: A comprehensive database of predicted operons in microbial genomes." Nucleic Acids Res **37**(Database issue): D479-482.

Perez-Novo, C. A., et al. (2005). "Impact of rna quality on reference gene expression stability." Biotechniques **39**(1): 52, 54, 56.

Pfaffl, M. W., et al. (2004). "Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: Bestkeeper--excel-based tool using pair-wise correlations." Biotechnol Lett **26**(6): 509-515.

Phornphisutthimas, S., et al. (2007). "Conjugation in escherichia coli: A laboratory exercise." Biochem Mol Biol Educ **35**(6): 440-445.

Poret-Peterson, A. T., et al. (2008). "Transcription of nitrification genes by the methane-oxidizing bacterium, *methylococcus capsulatus* strain bath." ISME J. **2**: 1213-1220.

Postgate, J. R. (1970). "Biological nitrogen fixation." Nature **226**: 25-27.

Prosser, J. I. (1989). "Autotrophic nitrification in bacteria." Adv. Microb. Physiol. **30**: 125-181.

Purkhold, U., et al. (2000). "Phylogeny of all recognized species of ammonia oxidizers based on comparative 16s rrna and *amoa* sequence analysis: Implications for molecular diversity surveys." Appl. Environ. Microbiol. **66**(12): 5368-5382.

Purkhold, U., et al. (2003). "16s rrna and amoa-based phylogeny of 12 novel betaproteobacterial ammonia-oxidizing isolates: Extension of the dataset and proposal of a new lineage within the nitrosomonads." Int. J. Syst. Evol. Microbiol. **53**(Pt 5): 1485-1494.

Ravishankara, A. R., et al. (2009). "Nitrous oxide (n2o): The dominant ozone-depleting substance emitted in the 21st century." Science **326**: 123-125.

Richter, M. and Rossello-Mora, R. (2009). "Shifting the genomic gold standard for the prokaryotic species definition." Proc Natl Acad Sci U S A **106**(45): 19126-19131.

Rissman, A. I., et al. (2009). "Reordering contigs of draft genomes using the mauve aligner." Bioinformatics **25**(16): 2071-2073.

Roberts, R. J., et al. (2003). "A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes." Nucleic Acids Res **31**(7): 1805-1812.

Rotthauwe, J. H., et al. (1997). "The ammonia monooxygenase structural gene *amoa* as a functional marker: Molecular fine-scale analysis of natural ammonia-oxidizing populations." Appl. Environ. Microbiol **63**(12): 4704-4712.

Ryther, J. H. and Dunstan, W. M. (1971). "Nitrogen, phosphorus, and eutrophication in the coastal marine environment." Science **171**(3975): 1008-1013.

Sanger, F., et al. (1977). "DNA sequencing with chain-terminating inhibitors." Biotechnology **24**: 104-108.

Sayavedra-Soto, L. A., et al. (2011a). "The membrane-associated monooxygenase in the butane-oxidizing gram-positive bacterium *nocardioides* sp. Strain cf8 is a novel member of the amo/pmo family." Environ Microbiol Reports **3**(3): 1758-2229.

Sayavedra-Soto, L. A., et al. (2011b). Dissecting iron uptake and homeostasis in *nitrosomonas europaea*. Methods in enzymology. Klotz, M. G., Academic Press. Volume **486:** 403-428.

Schmehl, M., et al. (1993). "Identification of a new class of nitrogen fixation genes in *rhodobacter capsulatus*: A putative membrane complex involved in electron transport to nitrogenase." Mol. Gen. Genet. **241**: 602-615.

Schmid, M. C., et al. (2008). "Environmental detection of octahaem cytochrome c hydroxylamine/hydrazine oxidoreductase genes of aerobic and anaerobic ammonium-oxidizing bacteria." Environ Microbiol **10**(11): 3140-3149.

Schmidt, I., et al. (2001). "Ammonia oxidation by *nitrosomonas eutropha* with no_2 as oxidant is not inhibited by acetylene." Microbiology **147**(8): 2247-2253.

Scott, K. M., et al. (2006). "The genome of deep-sea vent chemolithoautotroph thiomicrospira crunogena xcl-2." PLoS Biol **4**(12): e383.

Sevin, E. W. and Barloy-Hubler, F. (2007). "Rasta-bacteria: A web-based tool for identifying toxin-antitoxin loci in prokaryotes." Genome Biol **8**(8): R155.

Shapiro, B. J., et al. (2012). "Population genomics of early events in the ecological differentiation of bacteria." Science **336**(6077): 48-51.

Shen, Z., et al. (2006). "Nutrient structure of seawater and ecological responses in jiaozhou bay, china." Estuarine, Coastal and Shelf Science **69**(1-2): 299-307.

Shen, Z. L. (2001). "Historical changes in nutrient structure and its influences on phytoplantkon composition in jiaozhou bay." Estuarine, Coastal and Shelf Science **52**(2): 211-224.

Siebers, B., et al. (2011). "The complete genome sequence of *thermoproteus tenax*: A physiologically versatile member of the crenarchaeota." PLoS One **6**(10): e24222.

Sievers, F., et al. (2011). "Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega." Mol Syst Biol **7**: 539.

Silva, P. J., et al. (2000). "Enzymes of hydrogen metabolism in pyrococcus furiosus." Eur J Biochem **267**(22): 6541-6551.

Simon, D. M. and Zimmerly, S. (2008). "A diversity of uncharacterized reverse transcriptases in bacteria." Nucleic Acids Res **36**(22): 7219-7229.

Simon, J. and Klotz, M. G. (2013). "Diversity and evolution of bioenergetic systems involved in microbial nitrogen compound transformations." Biochim Biophys Acta **1827**(2): 114-135.

Simon, R., et al. (1983). "A broad host range mobilization system for in vivo genetic engineering: Transposon mutagenesis in gram negative bacteria." Bio-Technology **1**: 784-791.

Skerman, V. B. D., et al. (1980). "Approved lists of bacterial names." International Journal of Systematic Bacteriology **30**: 225-420.

Smith, J. L., et al. (2008). "Nautilia profundicola sp. Nov., a thermophilic, sulfur-reducing epsilonproteobacterium from deep-sea hydrothermal vents." Int J Syst Evol Microbiol **58**(Pt 7): 1598-1602.

Spreitzer, R. J. (2003). "Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase." Arch Biochem Biophys **414**(2): 141-149.

Spring, S., et al. (2013). "Taxonomy and evolution of bacteriochlorophyll a-containing members of the om60/nor5 clade of marine gammaproteobacteria: Description of luminiphilus syltensis gen. Nov., sp. Nov., reclassification of haliea rubra as pseudohaliea rubra gen. Nov., comb. Nov., and emendation of chromatocurvus halotolerans." BMC Microbiol **13**: 118.

Stahl, D. A. and De La Torre, J. R. (2012). "Physiology and diversity of ammonia-oxidizing archaea." Annu Rev Microbiol **66**: 83-101.

Stein, L. Y., et al. (2007). "Whole-genome analysis of the ammonia-oxidizing bacterium, nitrosomonas eutropha c91: Implications for niche adaptation." Environ Microbiol **9**(12): 2993-3007.

Stein, L. Y., et al. (2011). "Genome sequence of the methanotrophic alphaproteobacterium methylocystis sp. Strain rockwell (atcc 49242)." J Bacteriol **193**(10): 2668-2669.

Stein, L. Y., et al. (2013). "Energy-mediated vs. Ammonium-regulated gene expression in the obligate ammonia-oxidizing bacterium, nitrosococcus oceani." Front Microbiol **4**: 277.

Stein, L. Y. and Klotz, M. G. (2011). "Nitrifying and denitrifying pathways of methanotrophic bacteria." Biochem Soc Trans **39**(6): 1826-1831.

Stein, L. Y. and Klotz, M. G. (2016). "The nitrogen cycle." Curr Biol **26**(3): R94-98.

Stephen, J. R., et al. (1999). "Effect of toxic metals on indigenous soil beta-subgroup proteobacterium ammonia oxidizer community structure and protection against toxicity by inoculated metal-resistant bacteria." Appl Environ Microbiol **65**(1): 95-101.

Stewart Ac, O. B., Read Td (2009). "Diya: A bacterial annotation pipeline for any genomics lab. ." Bioinformatics. **25** 962-963.

Strous, M. and Jetten, M. S. (2004). "Anaerobic oxidation of methane and ammonium." Annu Rev Microbiol **58**: 99-117.

Strous, M., et al. (1999). "Missing lithotroph identified as new planctomycete." Nature **400**: 446-449.

Suwa, Y., et al. (2011). "Genome sequence of nitrosomonas sp. Strain al212, an ammonia-oxidizing bacterium sensitive to high levels of ammonia." J Bacteriol **193**(18): 5047-5048.

Tarca, A. L., et al. (2006). "Analysis of microarray experiments of gene expression profiling." Am J Obstet Gynecol **195**(2): 373-388.

Thompson, C. C., et al. (2013). "Genomic taxonomy of the genus prochlorococcus." Microb Ecol **66**(4): 752-762.

Tittsler, R. P. and Sandholzer, L. A. (1936). "The use of semi-solid agar for the detection of bacterial motility." J Bacteriol **31**(6): 575-580.

Trotsenko, Y. A. and Murrell, J. C. (2008). "Metabolic aspects of aerobic obligate methanotrophy." Adv Appl Microbiol **63**: 183-229.

Urakawa, H., et al. (2015). "Nitrosospira lacus sp. Nov., a psychrotolerant, ammonia-oxidizing bacterium from sandy lake sediment." Int J Syst Evol Microbiol **65**(Pt 1): 242-250.

Van Kessel, M. A., et al. (2015). "Complete nitrification by a single microorganism." Nature **528**(7583): 555-559.

Vignais, P.M. and Billoud, B. (2007). "Occurrence, classification, and biological function of hydrogenases: An overview." Chem Rev. **107**(10):4206-72.

Vignais, P. M., et al. (2001). "Classification and phylogeny of hydrogenases." FEMS Microbiol Rev 25(4): 455-501. Vitousek, P. M., et al. (1997). "Human alteration of the global nitrogen cycle: Sources and consequences." Ecol. Applic. **7**: 737-750.

Vitousek, P. M., et al. (1997). "Human alterations of the global nitrogen cycle: Sources and consequences."

Voss, M., et al. (2013). "The marine nitrogen cycle: Recent discoveries, uncertainties and the potential relevance of climate change." Philos Trans R Soc Lond B Biol Sci **368**(1621): 20130121.

Walker, C. B., et al. (2010). "Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea." Proc Natl Acad Sci U S A **107**(19): 8818-8823.

Wang, L., et al. (2016). "D1FHS, the type strain of the ammonia-oxidizing bacterium nitrosococcus wardiae spec. Nov.: Enrichment, isolation, phylogenetic, and growth physiological characterization." Front Microbiol **7**: 512.

Wang, Y., et al. (2015). "Orthovenn: A web server for genome wide comparison and annotation of orthologous clusters across multiple species." Nucleic Acids Res **43**(W1): W78-84.

Wang, Z., et al. (2009). "Rna-seq: A revolutionary tool for transcriptomics." Nat Rev Genet **10**(1): 57-63.

Ward, B. B. (2002). Nitrification in aquatic systems. Encyclopedia of environmental microbiology. Capone, D. A. New York, Wiley & Sons**:** 2144-2167.

Ward, B. B., et al. (2007). "Ammonia-oxidizing bacterial community composition in estuarine and oceanic environments assessed using a functional gene microarray." Environmental Microbiology **9**(10): 2522-2538.

Ward, B. B. and O'mullan, G. D. (2002). "Worldwide distribution of nitrosococcus oceani, a marine ammonia-oxidizing -proteobacterium, detected by pcr and sequencing of 16s rrna and amoa genes." Applied and Environmental Microbiology **68**(8): 4153-4157.

Ward, N., et al. (2004). "Genomic insights into methanotrophy: The complete genome sequence of methylococcus capsulatus (bath)." PLoS Biol **2**(10): e303.

Watanabe, S., et al. (2012). "Structural basis of [nife] hydrogenase maturation by hyp proteins." Biol Chem **393**(10): 1089-1100.

Watson, S. (1965). "Characteristics of a marine nitrifying bacterium, *nitrosocystis oceanus* sp. N." Limnol. Oceanogr. **10**: R274-289.

Watson, S. W. and Mandel, M. (1971). "Comparison of the morphology and deoxyribonucleic acid composition of 27 strains of nitrifying bacteria." Journal of Bacteriology **107**: 563-569.

Wayne, L. G., et al. (1987). "Report of the ad hoc committee on reconciliation of approaches to bacterial systematics " Int J Syst Bacteriol **37**: 463-464.

Wei, X., et al. (2006). "Transcript profiles of nitrosomonas europaea during growth and upon deprivation of ammonia and carbonate." FEMS Microbiol Lett **257**(1): 76-83.

Welander, P. V. and Summons, R. E. (2012). "Discovery, taxonomic distribution, and phenotypic characterization of a gene required for 3-methylhopanoid production." Proc Natl Acad Sci U S A **109**(32): 12905-12910.

Whelan, S. and Goldman, N. (2001). "A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach." Mol Biol Evol **18**(5): 691-699.

Whittaker, M., et al. (2000). "Electron transfer during the oxidation of ammonia by the chemolithotrophic bacterium *nitrosomonas europaea*." Biochimica et Biophysica Acta **1459**: 346-355.

Whittenbury, R., et al. (1970). "Enrichment, isolation and some properties of methane-utilizing bacteria." Journal of General Microbiology **61**: 205-218.

Wiedenheft, B., et al. (2012). "Rna-guided genetic silencing systems in bacteria and archaea." Nature **482**(7385): 331-338.

Winogradsky, S. (1892). "Contributions à la morphologie des organismes de la nitrification. ." Archives des Sciences Biologique **1**: 88-37.

Wu, M. and Eisen, J. A. (2008). "A simple, fast, and accurate method of phylogenomic inference." Genome Biology **9**(10): R151-R151.

Wu, M. L., et al. (2015). "Xoxf-type methanol dehydrogenase from the anaerobic methanotroph "*candidatus* methylomirabilis oxyfera"." Applied and Environmental Microbiology **81**(4): 1442-1451.

Zehr, J. P., et al. (2007). "Low genomic diversity in tropical oceanic n2-fixing cyanobacteria." Proc Natl Acad Sci U S A **104**(45): 17807-17812.

Zhang, K., et al. (2006). "Sequencing genomes from single cells by polymerase cloning." Nat Biotechnol **24**(6): 680-686.

Zhou, Y., et al. (2011). "Phast: A fast phage search tool." Nucleic Acids Res **39**(Web Server issue): W347-352.

Zumft, W. G. (1997). "Cell biology and molecular basis of denitrification." Microbiol Mol Biol Rev. **61**: 522-616.

APPENDIX A: PRIMERS USED IN THIS DISSERTATION

Name	Sequence (5'-3')	Used for	Used in
M13F	GTAAAACGACGGCCAG	Sequencing	Chapter 2, 5.1, 5.2, 5.3
M13R	CAGGAAACAGCTATGA	Sequencing	Chapter 2, 5.1, 5.2, 5.3
27F	AGAGTTTGATCMTGGCTCAG	Sequencing; PCR	Chapter 2
1492R	GGTTACCTTGTTACGACTT	Sequencing; PCR	Chapter 2
gamma-haoAF	YTGYCAYAAYGGRGYNGAYCAYAAYGAGT	PCR	Chapter 2
gamma-haoAR	TTRTARWGCTKGAKSANRTGMTGYTCCCACAT	PCR	Chapter 2
gamma-16SF	CAATCTGAGCCATGATCAAAC	PCR	Chapter 2
gamma-16SR	CCTACGGCTACCTTGTTACG	PCR	Chapter 2
q16SF	TCAGCTCGTGTCGTGAGATG	qPCR	Chapter 3
q16SR	AAGGGCCATGACTTGACG	qPCR	Chapter 3
qcynF	TATGATCCCAGCATTCTCC	qPCR	Chapter 3
qhaoAF	TGCGAAGGGTAATAAGATGG	qPCR	Chapter 3
qhaoAR	TGATATGACACATGGTGCAA	qPCR	Chapter 3
qorf2F	CTGCTTATACCTGATCCGTG	qPCR	Chapter 3
qorf2R	CAGTGAACAACAACATAGGC	qPCR	Chapter 3
qcynR	TTGAAATCGATAGCGGAC	qPCR	Chapter 3
qamoAF	TATTCAGGAGTACCCGTC	qPCR	Chapter 3
qamoAR	GGTCAAACCATAGCTCTTG	qPCR	Chapter 3
RE1haoArev	NNNTCCGGAGTCTCGCCCGAC	Cloning	Chapter 5.1
RE1haoBfwd	NNNTCCGGAAGGAGCACACCCCAATGTACATATCTGAATTAAAGCGGTCTCGG	Coning	Chapter 5.1
RE3PhaoF	NNNTCTAGAGAGCACCTCGGAGTGACACC	PCR	Chapter 5.1
RE1PhaoR	NNNCAATTGTCCTTCGGCAATCCATCCTG	PCR	Chapter 5.1
RE1cycA(B)F	NNNCAATTGATCGTTCAGATGAGGACACG	PCR	Chapter 5.1
RE2cyc(A)BR	NNNNATCGATTCGCCATTGGCAGTAATAGG	PCR	Chapter 5.1
RE1cycBF	NNNCAATTGTAAATGGATGGAGTAAAATAATATGC	PCR	Chapter 5.1
RE2cycAR	NNNATCGATACTCCTACCATTTATTTTTTCTGC	PCR	Chapter 5.1
qnoccycAF	ATCCTGAGTGCGTCGGTTG	qPCR	Chapter 5.1

qnoccycAR	CATCTTATGGTTATGCTTGT	qPCR	Chapter 5.1
qnoccycBF	TTATCAACGACCACCTTTTG	qPCR	Chapter 5.1
qnoccycBR	AAGTTCCCTAACCCCTGA	qPCR	Chapter 5.1
Fqmca16S	GCACCTCAGCGTCAGTGTT	qPCR	Chapter 5.1; 5.2
Rqmca16S	CGTAGGCGGTTTGATAAGTC	qPCR	Chapter 5.1; 5.2
MCA0956F	GCCAGGCGTTCTGGAGTCCG	Cloning	Chapter 5.2
MCA0956R	GGAGAGGGCTTGGGTGAGGGG	Cloning	Chapter 5.2
q2haoF	GCTCTACAAGGGGCTGGTC	qPCR	Chapter 5.2
q2haoR	CCGTGGGCGGTTGATAGA	qPCR	Chapter 5.2
q2orf2F	TACCTGCTCGTCTGTCTGGA	qPCR	Chapter 5.2
q2orf2R	AGATAGCTGGCCTGATCGAC	qPCR	Chapter 5.2
Fqmca0956	GCTCTACAAGGGGCTGGTC	qPCR	Chapter 5.2
Rqmca0956	GCAAACGGGTGTTCTCGTC	qPCR	Chapter 5.2
F2qmca955	CCCTGTTCCTCGCCTGGT	qPCR	Chapter 5.2
R2qmca955	TACGCTCCGCCTCCTCATC	qPCR	Chapter 5.2
F3qmca955	TACCTGCTCGTCTGTCTGGA	qPCR	Chapter 5.2
R3qmca955	AGATAGCTGGCCTGATCGAC	qPCR	Chapter 5.2
qMcaheme8F	GCCGACAAGGGCACACTC	qPCR	Chapter 5.2
qMcaheme8R	CGCAGGGCGGTTGGTCTTCT	qPCR	Chapter 5.2
Fqmca2394	CAAGGTGAAATACCCCGATG	qPCR	Chapter 5.2
Rqmca2394	TAGTTGCCGCCCCTCAGT	qPCR	Chapter 5.2
qMca0590-F	TGGTGACCGTACCGTCTTTC	qPCR	Chapter 5.2
qMca0590-R	GTCTGGATCATGGCCACCAA	qPCR	Chapter 5.2
FqnirB	GCATCCGTCCCAACATAGAA	qPCR	Chapter 5.2
RqnirB	CATTCACCGACCGCATAGA	qPCR	Chapter 5.2
Fqhaoterm	CCGGAAAGGATGACTCGAAC	qPCR	Chapter 5.2
Rqhaoterm	GGAACCGTCGTACAGCA	qPCR	Chapter 5.2
Fqmca2120	ACGGACATCCACCAGTGTTC	qPCR	Chapter 5.2
Rqmca2120	GTGGCGCTGTCGATACTCTT	qPCR	Chapter 5.2
Fqmca2130	CTGCGATTCTGGGTCTTGG	qPCR	Chapter 5.2
Rqmca2130	TGATAAGCGGTGAAATGGA	qPCR	Chapter 5.2
Fqmca0563	GAGACTTCCAGACCGACCACG CA	qPCR	Chapter 5.2
Rqmca0563	CCAGACGCAGCAGGTAAGGCA	qPCR	Chapter 5.2
Fqmca1677	CGAAACCCCTGGTCGGAGACAA	qPCR	Chapter 5.2
Rqmca1677	CGGACAACCCGCCGTAAAGGTC	qPCR	Chapter 5.2
PpmoC1F	NNNNCTCGAGTCTAGAGAAGG GATGTCATCGGGAGA	cloning	Chapter 5.3
PpmoC1R	NNNNGTCGACCAATTGCCTCC TAAAGTGATGGTTGAC	cloning	Chapter 5.3
NamH-0559F	NNNGTCGACGGTGAAGATAAAA	cloning	Chapter 5.3

	T AAAGGTG		
NamH-0559R	NNNGGATCCTTATTCTCCTTCAT A TCTTTCAAGG	cloning	Chapter 5.3
NamH-1280F	NNNGGATCCCTATTGGCACTTTT A CAC	cloning	Chapter 5.3
NamH-1280R	NNNNNGAATTCAAGCTGCTTTTC A CAGAC	cloning	Chapter 5.3

APPENDIX B: SEQUENCES OF CONSTRUCTS BUILT IN CHAPTER 5.1

1. Sequence of pBAB1 ∝

TCTAGAGAGCACCTCGGAGTGACACCGGACGAGCCGGCGTTCCGCCCGACCG

XbaI RE3PhaoF

GGGCGCCTCGAACGACGCCTCTTGATTGTTGACGTCATTTGGTTTTGTTCCGA

-35 -10

TACTTTGTACCGATAACGACCAGGATGGATTGGCGAAGGACAATTGATCGTT

RE1PhaoR MfeI

CAGATGAGGACACGATCCGCACCTATTTTTTGACTGGCGACACTGATAAAAA

RE1cycABF

TGCTCTGATAGTCAAACTGCTACCGTTTAGCACCTCAAACCCAATGGAACTTG
AAGTGCTCAAGCTGGTCGCTCAGTATGGCAGTTATTGGGTTTATTCGTTGCCG
CCTGCTAATGATGCTGGTTAATTAATGTTTTCCTCTAAAGGAGAAAATTTATG

RBS *

AACCACATAATTAGATATACATTCTTTTATGCTCTATTGGCGATCGCTATGGG
AACGGGGTGGGCTCAAGGGGAACCACCCTACCATGGCCTTAAGAAATGTAAA
AGTTGCCATGAGTCTCAGTATGATTCATGGCTTGAAACGGATCATGGACAGG
CGATGAAGTCTTTAGAGGCTGGCGAAGAGGTAGAGGTCAACGAAAAAGCAG
GGCTGGACCCTGATGAGGATTATACTGAAGATCCTGAGTGCGTCCGTTGCCA
TGTGACCGGGTTTAGGAAAGATGAGGGGCTACTAAATTGATCTCTCCAAACC
GCGTTGAAATATGTATTTACAGGGGTGTTGGTTGCGAATCCTGCTCATGGTCC
TGGCGTGCCAGGGGACGAGCATGACCATAAAGATGCGGGCCATAGAAGTTTT
GTAAGAGTCCCAGGAGACGTACGTTCTTCGGCAGAGAGCTCGCTGCTATCGA
GCAAGGTTTGTAGCCGACTTATGGAATTTGATGACGTGCAATGCTTGTCACCG
TTAACTACGGACGTCGCCATGGACGTCTTGTCTAAGTAGGCCCTAGAGCCCC
CAT……
……....…GGCATTAATAAAGGCTTA
GAGGGCGTTTTTTCTGGTTGATTCGGTCGGCATCCGTTGCATGAAAGAATCCA
AAAGGCATGCTGCGCCACCGCAGAAAAAAATAAATGGTAGGAGTAAAAATA

RE2cycAR ClaI

AATATGCTATCTAAATTAAACAGTGGTGGCAAGCCCATTGTTGAGTATTTTAA
AGTTCATTGGCGCCCGGTTGCGGTCGGAGCGGGGGGGGCTATTTGCTCTAATT
GTTATCGTTTTTGGAGGCGAGGCGGCGTTATCAACGACCACCTTTTGCACTAG
TTGTCATTCGATGTCTTATCCGTATGAGGAATTGCAAGAGTCCGCTCATTATG
GCAGGTTTGGTTTGCAACCCCAATGTAAGGATTGCCATATTCCTCAGGGGTTA
GGGAACTTCCACAAGGCAGTTTATACTCATGTAGTGGATGGGGCGCGAGAGC
TTTATTTAGAACTTGTTAACGATTATTCTACACTTGAAAAATTTAATGAGCGC
CGGCAAGAAATGGCCCATCATGCCCGCATGAATCTAAAGGGATGGGATAGCG
TGACCTGCCGGGATTGCCATCAGGATCCACAACCAGCAGCTAAATCTGGCCA
GCGGGCTCACAAAAAGCTTGATGACGGCTGGACCTGCATTGATTGCCATCAA
AACCTTGTGCATGCTCCTGTGGAGCATACCGATCTTGATGCAAGTGAGCGGC
GTGGCGAGATGGTCATCATGGATGAGATAGATTGGACAGACCGTAACCCGCT
AGCCGGAGTGGAGACTCGTTCGATAAGGAATTAAATATAAGAGCGAGCGGTC
GCAATATTATTTGTTAGATCTTATAAAAGCCCCAGCTAAAGCCGGGGCTTTTT
TTGAGCTGACGAGCGGAAAAATACCTATTACTGCCAATGGCGAATCGAT

RE2cycABR **ClaI**

2. Sequence of pBAB2 ∝

TCTAGAGAGCACCTCGGAGTGACACCGGACGAGCCGGCGTTCCGCCAGACCG
XbaI **RE3PhaoF**
GGGCGCCTCGAACGACGCCTCTTGATTGTTGACGTCATTTGGTTTTGTTCCGA
-35 **-10**
TACTTTGTACCGATAACGACCAGGATGGATTGCCGAAGGACAATTGTAAATG
RE1PhaoR **MfeI** **RE1cycBF**
GTAGGAGTAAAATAATATGCTATCTAAATTAAAACAATGGTGGCAAGCCATT
RBS *****
GTTGAGTATTTTAAAGTTCATTGGCGCCCGGTTGCGGTCGGAGCGGGGGGGG
CCATTTGCTCTAATTGTTATCGTTTTTGGAGGCGAGGCGGCGTTATCAACGAC
CACCTTTTGCACTAGTTGTCATTCGATGTCATATCCGTATGAGGAATTGCAAG
AGTCCGCTCATTATGGCAGGTTTGGTTTGCAACCCCAATGTAAGGATTGCCAT
ATTCCTCAGGGGTTAGGGAACTTCCACAAGGCAGTTTATACTCATGTAGTGG
ATGGGGCGCGAGAGCTTTATTTAGAACTTGTTAACGATTATTCTACACTTGAA
AAATTTAATGAGCGCCGGCAAGAAATGGCCCATCATGCCCGCATGAATCTAA
AGGGATGGGATAGCGTGACCTGCCGGGATTGCCATCAGGATCCACAACCAGC

AGCTAAATCTGGCCAGCGGGCTCACAAAAAGCTTGATGACGGCTGGACCTGC
ATTGATTGCCATCAAAACCTTGTGCATGCTCCTGTGGAGCATACCGATCTTGA
TGCAAGTGAGCGGCGTGGCGAGATGGTCATCATGGATGAGATAGATTGGACA
GACCGTAACCCGCTAGCCGGAGTGGAGACTCGTTCGATAAGGAATTAAATAT
AAGAGCGAGCGGTCGCAATATTTTTTGTTAGATCTTATAAAAGCCCCAGCTA
AAGCCGGGGCTTTTTTTGAGCTGACGAGCGGAAAAATACCTATTACTGCCAA
TGGCGAATCGAATCGAT

RE1cycBR **ClaI**

3. Sequence of pBAB3 ∝

TCTAGAAGCACCTCGGAGTGACACCGGACGAGCCGGCGTTCCGCCAGACCGG
XbaI **RE3PhaoF**
GGCGCCTCGAACGACGCCTCTTGATTGTTGACGTCATTTGGTTTTGTTCCGAT
-35 **-10**
ACTTTGTACCGATAACGACCAGGATGGATTGCCGAAGGACAATTGATCGTTC
RE1PhaoR **MfeI**
AGATGAGGACACGATCCGCACCTATTTTTTGACTGGCGACACTGATAAAAAT
RE1cycAF
GCTCTGATAGTCAAACTGCTACCGTTTAGCACCTCAAACCCAATGGAACTTGA
AGTGCTCAAGCTGGTCGCTCAGTATGGCAGTTATTGGGTTTATTCGTTGCCGC
CTGCTAATGATGCTGGTTAATTAGTGTTTTCCTCTAAAGGAGAAAATTTATGA
RBS *****
ACCAGATAATTAGATATACATTCTTTTATGCTCTATTGGCGATCGCTATGGGA
ACGGTGTGGGCTCAAGGGGAACCACCCTACGATGGCCTTAAGAAATGTAAGA
GTTGCCATGAGTCTCAGTATGATTCATGGCTTGAAACAGATCATGGACAGGC
GATGAAGTCTTTAGAGCCTGGCGAAGAGGTAGAGGCCAAGGAAAAAGCAGG
GCTGGACCCTGATGAGGATTATACTGAAGATCCTGAGTGCGTCGGTTGCCAT
GTGACCGGGTTTAGGAAAGATGGGGGCTACGAAATTGATCTCTCCAACCGCT
TGAAAAAGTATTTACAGGGTGTTGGTTGCGAATCCTGCCATGGTCCTGGGTCC
AGGTACAAGCATAACCATAAAGATACGGCCAAAAAGTTTGAGAGTCCCAAG
GAAACTACTTCTCGGCAGGAGCTCGCCGCTATCGGGCAGGTTTTTACCGATA
AGGAATTTGAAGAGCGTTGCAATGCTTGTCACCTTAACTACGAAGGTTCGCC
ATGGCCTCATGCTAAAGAGCCCTATACCCCCTTCACTCCTGAAGTGGATCCAA
AATACGAGTTCAATTTTGGAGAAGCGGTGCATAATGATGAAGCGATGCACAA

GCATTATAAGCTAGAGGGCGTTTTTTCTGGTGATCCGGTCGCATCGTTCCATG

AAGAATTCCAAAAGCATGCTGCGCCACCGCAGAAAAATAAATGGTAGGAGT

RE1cycAR

ATCGAT

ClaI

∝: The following regions were labeled in above sequences: -35 region, -10 region (a.k.a Pribnow box), start codon (*) in DNA sequence, Ribosomal Binding Site (RBS) in DNA sequence, cloning primers and RE sites.