Applied and Environmental Microbiology

Characterization of two new genes, amoR and amoD, in the amo operon of the marine ammonia oxidizer Nitrosococcus oceani ATCC 19707

Amal F. El Sheikh, Amisha T. Poret-Peterson, and Martin G. Klotz

Prepublished November 9, 2007;
doi:10.1128/AEM.01654-07.

Submitted for publication as a NOTE in Applied and Environmental Microbiology

Title: Characterization of two new genes, *amoR* and *amoD*, in the *amo* operon of the marine ammonia oxidizer *Nitrosococcus oceani* ATCC 19707.

Authors: Amal F. El Sheikh[1], Amisha T. Poret-Peterson[1], and Martin G. Klotz[1,2,*],

Affiliations: [1]Evolutionary and Genomic Microbiology Laboratory, Department of Biology and [2]Department of Microbiology & Immunology, University of Louisville, Louisville, KY 40292, USA.

Running title: The *amoRCABD* operon in nitrosococci.

Keywords: *Nitrosococcus*, ammonia oxidation, *amo* gene expression

***Corresponding author:**

Martin G. Klotz, Ph.D.

University of Louisville

Department of Biology

139 Life Science Building

Louisville, KY 40292

502.852.7779 (O), 502.852.6773 (L), 502.852.0725 (F)

aem@mgklotz.com (e-mail)

ABSTRACT

Molecular analysis of the *amo* gene cluster in *Nitrosococcus oceani* revealed that it consists of 5 instead of the three known *amoCAB* genes. The two additional genes, *orf1* and *orf5*, were introduced as the *amoR* and *amoD* genes, respectively. Putative functions of the AmoR and AmoD proteins are discussed.

INTRODUCTION

Nitrosococcus oceani ATCC 19707 is a marine aerobic ammonia-oxidizing bacterium (AOB) that belongs to the class *Gammaproteobacteria*, order *Chromatiales* (the purple sulfur bacteria). Its complete genome sequence has been published recently (15). In AOB, ammonia oxidation proceeds in two consecutive steps. First, ammonia is converted to hydroxylamine via the multi-subunit, membrane-bound enzyme Ammonia Monooxygenase (AMO) in the following reaction: $NH_3 + 2e^- + O_2 + 2H^+ \rightarrow NH_2OH + H_2O$ (10). The subsequent oxidation of hydroxylamine to nitrite is facilitated by the soluble periplasmic enzyme, hydroxylamine oxidoreductase (HAO): $NH_2OH + H_2O \rightarrow NO_2^- + 5H^+ + 4e^-$ (10). The oxidation of hydroxylamine to nitrite yields four electrons, two of which are returned to the upstream monooxygenase reaction, and the remaining two are the sole source for generating useable energy and reductant. The mechanism of returning the electrons to AMO is unknown (2).

AMO is encoded by at least three contiguous genes, *amoCAB*, arranged in a gene cluster that is conserved in all investigated genomes of AOB (3, 8, 12, 14, 15, 20-22). Prior work had identified a conserved open reading frame (ORF) following the terminator downstream of the single cluster of *amoCAB* genes in *N. oceani* (1, 20). A recent analysis of available genome sequences revealed that all *amo* gene clusters in betaproteobacterial AOB (Beta-AOB) genomes, which encode two or three copies of nearly identical gene clusters, are actually succeeded by two conserved ORFs, *orf4* and *orf*5, except for the *amo-hao* super cluster in *N. multiformis* (2, 12, 22). In contrast, the terminator downstream of the *amoB* gene in gammaproteobacterial AOB (Gamma-AOB) is succeeded only by the *orf5* gene (1), which is conserved at the levels of DNA and protein sequence in all AOB. Our recent analysis of the *N. oceani* genome sequence revealed

52 that the intergenic region between the *amoB* terminator and *orf5* did not contain a promoter
53 consensus sequence. Furthermore, examination of upstream flanking sequence of the *N. oceani*
54 *amoCAB* cluster revealed an additional 213-bp open reading frame (*orf1*) that has no homologue
55 in the non-redundant GenBank database including the published (8, 22) and unpublished (12)
56 genomes of Beta-AOB. These aforementioned gene-structural differences between Gamma- and
57 Beta-AOB indicate a possible divergent expression of *amo* genes in Gamma- and Beta-AOB,
58 which may correspond to their respective niche adaptation in the environment (i.e., Gamma-
59 AOB are restricted to marine oligotrophic environments).

60 We report in this paper that the *orf1* and *orf5* genes are co-transcribed with *amoCAB* in
61 *N. oceani,* designate these ORFs as *amoR* and *amoD, respectively,* and propose that these 5
62 genes constitute the Gamma-AOB-typical *amo* operon, *amoRCABD*.

63

64 In our experiments *Nitrosococcus oceani* C-107 ATCC 19707 (18, 25) was grown in
65 artificial seawater as 200 ml - 400 ml batch cultures in 2-L Erlenmeyer flasks for three weeks at
66 30 ºC in the dark without shaking as described previously (1). Genomic DNA was isolated from
67 cells in stationary growth phase using Wizard genomic DNA purification kit (Promega, Madison,
68 WI) according to manufacturer's recommendations. For RNA preparations, *N. oceani* cells were
69 harvested at mid to late exponential growth phase, and resuspended for 24 hours prior to RNA
70 isolation in 200 ml of fresh marine medium. RNA was isolated using Fast RNA Pro Blue Kit
71 (Q-BIOgene, Solon, Ohio) according to manufacturer's guidelines. Before cDNA synthesis, the
72 RNA preparations were treated with RNase-free RQ1 DNase (Promega, Madison, WI) according
73 to the manufacturer's protocol.

74

Northern hybridization. Approximately 2 μg/lane of total RNA was resolved by electrophoresis at 4.5 volts/cm for 5-6 hours on 0.9% agarose gel made with 1X MOPS buffer and 6% formaldehyde. An ethidium bromide-stained 9-kb RNA ladder (Ambion, Austin, Texas) was used as a size estimate. DIG-labeled dsDNA probes *amoC, amoA, amoB,* and *orf5* were generated using the PCR DIG probe synthesis kit (Boehringer Mannheim, Germany) according to manufacturer's protocol, specific primers (Biosynthesis, Lewisville, TX) (Table 1) and approximately 50 ng of genomic DNA as the template. Hybridizing RNA fragments were detected using anti-digoxigenin alkaline phosphatase antibodies (Boehringer Mannheim) and alkaline phosphatase chemiluminescent substrate CSPD (Roche) according to manufacturers' guidelines, and visualized with autoradiography film. The three probes were designed to target the three presently known genes in the *amo* gene cluster, *amoC, amoA,* and *amoB* ((14) and Fig. 3A), each hybridized to two RNAs of approximately 3.9 kb and 4.6 kb in size (Figure 1, lanes C, A, and B). This means that *N. oceani* expresses two different polycistronic transcripts that contain the *amoCAB* genes. In addition, hybridization to individual *amoC* and *amoA* transcripts was observed, which had been predicted before (1). A comparison with the genome sequence (15) revealed that the smaller 3.9-kb transcript extended from a transcriptional start point (*tsp*) approximately 600 bps upstream of the *amoC* gene to the intrinsic transcriptional terminator we identified downstream of *amoB*, and included the *amoCAB* genes. Surprisingly, the probe designed to target the *orf5* gene hybridized also to the larger 4.6-kb transcript but not to the shorter one (Figure 1, lane 4). While the intensity of the 4.6-kb band was rather low, presence of the transcript and residence of *orf5* on this transcript was verified by RT-PCR (see below). Thus we conclude that the *orf5* gene is expressed together with the *amoCAB* genes and resides on a transcript that begins at the same *tsp* as the shorter transcript, extends beyond the *amoB*

98 terminator and ends at a terminator downstream of *orf5*. Because (i) there were two transcripts
99 with only the larger one including the *orf5* gene, (ii) the larger band was of lesser intensity than
100 the smaller band, and (iii) the *in silico* size difference between *amoCAB* and *amoCAB-orf5*
101 transcripts accounted for the size difference between the smaller and the larger bands in the
102 Northern blot (Fig. 1), we propose that *orf5* is expressed by partial read-through of the *amoB*
103 terminator as a member of the *amo* operon. Because *orf5* is ancestral to the *orf4* genes found in
104 Beta-AOB (see below) and the *orf4* gene is missing in Gamma-AOB, we designate this gene
105 *amoD*. Both transcripts contained a leader sequence of approximately 0.6 kb that is large enough
106 to contain the small *orf1* gene that we identified *in silico* upstream of *amoC*.

107 **Primer extension analysis.** To verify these conclusions, we conducted primer extension
108 experiments to determine the transcriptional start point as well as RT-PCR to confirm the
109 residence of *amoD* on the larger one of the two observed transcripts. To this end, a synthetic 29-
110 mer oligonucleotide, X1 (Table 1, Fig. 2A), was 5' end-labeled with (γ^{32}P) ATP (Perkin Elmer,
111 Boston, MA, specific activity, 3000 Ci mmol^{-1}) using T4 polynucleotide kinase (Promega). X1
112 was complementary to positions 91 - 63 in the nucleotide sequence upstream of *orf*1 (Fig. 2A).
113 Primer extension with total RNA was conducted using SuperScript III RT (Invitrogen) following
114 the manufacturer's protocol. The extension products were electrophoresed in 10% denaturing
115 polyacrylamide 10x10 cm mini gel, exposed to PhosphoImager screen, and analyzed using
116 molecular imager FX and Quantity One image analysis software (BIO-RAD, Hercules, CA). A
117 5'-end-labeled FX174 *Hinf* I DNA marker (Promega) was used as a size estimate. The resulting
118 cDNA product was 160 - 180 nucleotides in size (Fig. 2B) and identified the transcriptional start
119 point between nucleotides 223 and 232 upstream of the translational start for *orf1* (Fig. 2A). The
120 - 10 and -35 consensus sequences of the putative operon promoter was preceded by an A-track

(Fig. 2A), which is known to enhance transcription (9). In contrast to Beta-AOB where the distal *amo* promoter is located 166 nucleotides upstream of *amoC* gene (6), the operon promoter identified in *N. oceani* is much further upstream of the *amoC* gene thereby generating a leader sequence of at least 600 bp in size. In addition to the identified *orf1*, this leader contained 220 - 230 nucleotides of untranslated RNA located upstream of *orf1*. While the function of *orf1* and the significance of the 5' UTR were outside the scope of this study, the co-transcription of *orf1* with the *amoCAB* genes, its size and predicted topology as a small cytoplasmic alpha-helical protein (using PSIPRED (7)), and its uniqueness to nitrosococci suggest a regulatory involvement in ammonia catabolism of Gamma-AOB. For these reasons, we consider the *orf1*gene a member of the *amo* operon in *N. oceani* and designate it as *amoR*.

RT-PCR. To test our hypothesis that the *amoR* and *amoD* genes are part of the *amo* mRNA, specific primers were designed to amplify the four intergenic regions of the predicted *amo* operon (Fig. 3A) and an additional fragment that included the putative transcriptional start point (*tsp*), the entire *amoR* gene, and the 5'-end of *amoC*. The target positions and the sequences of the primers are provided in Fig. 3A and Table 1, respectively. The *tsp*F primer was selected based on PCR amplifications of the cDNA preparations with several tandem forward primers paired with the same *R-C*R reverse primer (data not shown)**.** The most upstream oligonucleotide primer that resulted in a high intensity PCR product was selected as *tsp*F primer for the final multiplex PCR.

First strand cDNA synthesis from *N. oceani* total RNA was carried out with SuperScript II (Invitrogen) and random nonamer primers according to manufacturer's protocol. Second strand synthesis was conducted in 100 µl reactions with the entire volume of the previous RT reaction, and 15 U of Klenow fragment (Promega, Madison, WI) according to the manufacturer's

144 protocol.

145 All final multiplex PCR assays were carried out using GoTaq Flexi DNA polymerase
146 (Promega, Madison, WI) with one of the following templates: (i) 8-10 ng of the cDNA, (ii)
147 DNase treated RNA preparations obtained before the reverse transcription for the negative
148 controls, (iii) approximately 10 ng of gDNA for the positive controls. Primers targeting the 16S
149 ribosomal sequence were included as an internal standard in every multiplex PCR amplification
150 reaction, including the positive and the negative controls. Negative control PCR reactions did not
151 result in any PCR products, which indicated the absence of DNA carryover in these preparations
152 (data not shown). Positive control PCR reactions resulted in bands of expected size (Fig. 3C),
153 indicating the specificity of the primers. The multiplex PCR assays conducted with cDNA as the
154 template are presented in Fig. 3B. Amplification the upstream leader sequence using primer *tsp*F
155 (Table 1; complementary to nucleotide positions 560-539 upstream of AmoC start codon) and
156 the reverse primer *amoR-C*R (Table 1; complementary to nucleotide positions -12 to +26 with
157 respect to AmoC start codon) resulted in an approximately 585-bp band (Fig. 3B, lane 2).
158 Amplification of the *amoR-amoC* intergenic region was conducted with forward primer *amoR-*
159 *C*F (Table 1, complementary tonucleotide position +56 to +82 within *amoR*), and the reverse
160 primer *amoR-C*R yielded a PCR product of 347 bp in size (Fig. 3B, lane 3). These results
161 indicate that *amoC* resides on a transcript with a leader of at least 600 bp in size including the
162 *amoR* gene; therefore, *amoR* and *amoC* are transcriptionally linked. Amplification of the *amoC-*
163 *amoA* and *amoA-amoB* intergenic spacers resulted in 290-bp and 450-bp bands, respectively
164 (Fig. 3B, lanes 4 and 5). This confirms that, similarly to Beta-AOB (19), the three *amoCAB*
165 genes reside on a common transcript in *N. oceani* and that all four amplicons (Fig. 3B, lanes 2, 3,
166 4, and 5) originated from one or both transcripts observed in the Northern hybridization

experiment (Fig. 1).

To investigate the difference between the two observed transcripts, the *amoB-amoD* intergenic spacer was amplified with a forward *B-D*F primer that targets the C-terminus of *amoB* and a reverse primer *B-D*R complementary to the N-terminus of the *amoD* gene. The observed band of ~ 330-bp (Fig. 3B, lane 6) indicates that *amoB* and *amoD* are transcriptionally linked, in that the *amoRCAB* transcript extends beyond the *amoB* terminator, the *amoB-amoD* intergenic spacer and the *amoD* gene, and terminates at an *in silico*-identified *rho*-independent terminator downstream of *amoD* (ΔG = -35.3 kcal/mol; start at nt. 636,765 in the *N. oceani* genome sequence, Genbank CP000127 (15)) thereby resulting in a 4.6-kb *amoRCABD* transcript as observed in the Northern hybridization experiment (Fig. 1).

AmoR and AmoD sequence comparison and analysis. Sequence similarities of AmoR and AmoD with proteins in the non-redundant database were investigated initially using the NCBI BLAST program (1). Protein sequences were also analyzed with the PSORT (19), PSIPRED (7), and Phobius (http://phobius.cgb.ki.se/ (13)) servers to identify the secondary structure and hydrophobic domains that could serve as signal peptides for export into the periplasm or constitute membrane-spanning domains.

None of the searched databases contained a sequence with significant similarity to the AmoR protein, hence this protein appears to be unique to *Nitrosococcus*. The AmoR orf is preceded by a Shine-Dalgarno sequence (Fig. 2A) and the protein is predicted to be cytoplasmic (19). Its sequence of 71 amino acids (Fig. 2A) contains three small helices of nine, four and nine amino acid residues art the N-terminus whereas the C-terminus contains a larger helical domain of 19 residues. Using a combination of sequence and predicted structure of the AmoR protein for a search of the PDB (http://www.rcsb.org/pdb/home/home.do) database, the N-terminal folding

190 domain of frizzled-related protein 3 (PDB entry 1ijxA0) was found to be nearly identical in its
191 fold and 23% identical to primary sequence of AmoR, extending from residue 25 to 105 (126
192 residues total). The deduced AmoR protein sequence contains three cysteine residues that are not
193 part of a known coordination motif. By analogy to the structure of frizzled-related protein 3
194 (1ijxA0), cysteine-37 and cysteine-71 of AmoR could form a disulfide bond. Thus AmoR could
195 serve as a cytoplasmic redox sensor in that this disulfide bond would lock AmoR into a particular
196 reactive secondary structure if the cytoplasm becomes less reducing. The N-terminus of frizzled-
197 related protein 3 is involved in protein-protein interactions. Based on this analogy and the fact
198 that the *amoR* gene is a member of the *amo* operon, we speculate that the AmoR protein
199 participates in the regulation of ammonia catabolism in *Nitrosococcus*. Experiments have been
200 initiated to test this hypothesis.

201 A multiple sequence alignment was produced from a total of 25 (20 full length and 5
202 partial) available protein sequences using the deduced full-length AmoD protein sequence from
203 *N. oceani* ATCC19707, respective BLAST hits retrieved from the GenBank/EMBL database and
204 unpublished genome sequences using ClustalX v.1.83 (24). Based on this alignment a distance
205 neighbor-joining tree was constructed with the BioNJ function in PAUP* v. 4.10b (23) and used
206 as a guide tree for manual refinement of the ClustalX alignment. Sources for the protein
207 sequences from AOB and other organisms used in the alignments are indicated in the presented
208 phylogenetic tree (Fig. 4). The alignment was subjected to a Bayesian inference of phylogeny
209 (MrBayes v. 3.0b4; written by Huelsenbeck and Ronquist; http://morphbank.ebc.uu.se/mrbayes)
210 using four equally heated Markov Chains over 1,000,000 generations in three independent runs.
211 The searches were conducted assuming an equal or a gamma distribution of rates across sites,
212 sampling every 100th generation and using the WAG empirical amino acid substitution model

(26). A 50% majority rule consensus phylogram was constructed that displayed the mean branch lengths and posterior probability values of the observed clades.

It has been previously reported that the gene (*orf5*) encoding the AmoD protein in Beta-AOB is in a tandem arrangement with another gene, *orf4* (2, 15)(20). The phylogenetic tree (Fig. 4) rooted in the AmoD homologue, which is expressed from the gene cluster that encodes particular methane monooxygenase (*pmo*) in the alphaproteobacterial methanotroph, *Methylosinus trichosporium*, demonstrates that all *orf4* and *orf5* genes are homologues, whether they are part of *amo* and *pmo* gene clusters in AOB and methane-oxidizing bacteria (MOB), respectively, or singletons.

Very recent work demonstrated that expression of the betaproteobacterial *orf4* and *orf5* gene tandems is significantly up-regulated during recovery from ammonia starvation (5). Interestingly, all *orf4-orf5* gene tandems are flanked downstream by copper resistance genes (*copCD*) in the genome sequences of Beta-AOB (2). Furthermore, our database search revealed that homologues of *amoD* (but not *orf4*) genes also reside in the vicinity of genes encoding copper enzymes in some genomes of MOB. In Alpha-MOB, *amoD* homologues reside downstream of *pmo* genes (CAJ01564-62, AAF37892-94), which encode a homologue of AMO (1, 20). In Gamma-MOB, an *amoD* homologue was found adjacent to genes encoding multi-copper oxidases (MCA2129, MCA2128) that are homologues of the *copA* copper oxidase gene upstream of the *amoRCABD* operon in *N. oceani*.

In silico analysis of the deduced AmoD protein structure revealed that AmoD is likely exported to the periplasm but stays anchored with its C-terminus in the inner membrane (19). Likewise, the AmoD protein is likely transported to and integrated into the extensive intra-cytoplasmic membrane system of AOB where it may interact with electron transfer proteins and

enzymes that facilitate the oxidation of ammonia or their maturation. This hypothesis is based on our finding that the *amoD* gene is an expressed member of the *amo* operon in *Nitrosococcus*, that it is also expressed in Beta-AOB (5) and that it is conserved in sequence and synteny in the *amo* gene clusters of all AOB; all of which suggests that the *amo* cluster genes encode proteins that interact physically (4, 11, 17).

In summary of these observations and results, we hypothesize that the *amoD* gene is distributed in AOB and MOB similar to other inventory that is involved in nitrification (i.e., *amo*/*pmo*, *hao* (2, 4, 16)) and that the AmoD protein is unique to organisms capable of ammonia-oxidation and nitrification. Because the *amoD* gene is ancestral to *orf4* (Fig. 4), which has likely arisen by complete gene duplication from *amoD* in the ancestor of all Beta-AOB, we also propose to name the *orf4* gene of Beta-AOB "*amoE*."

CONCLUSIONS

We have discovered that the *amo* gene cluster in *N. oceani* consists of five instead of the three known *amoCAB* genes and introduced the *amoR* and *amoD* genes. The transcription of the five genes in the *amoRCABD* gene cluster is not equal because other promoters in this cluster have been formerly identified in AOB (1, 14, 20, 21) and leaky terminators downstream of the *amoC* (21) and *amoB* (this study) genes will halt some but not all RNA polymerases that transcribe the genes in this cluster. We have shown that expression from a promoter at least 600 bp upstream of *amoC* generates an *amoRCABD* transcript in *N. oceani* and that these five genes constitute an operon when expressed under normal growth conditions. The unique presence of AmoR and absence of AmoE in *Nitrosococcus* (this study), the significant difference in amo operon copy number as well as the difference in organization of transcriptional units between

259 *Nitrosococcus* and Beta-AOB (this study and (2, 3, 5, 6)) suggest that the transcription of
260 ammonia-catabolic genes is differently regulated in Beta-AOB and *Nitrosococcus*. Putative
261 functions of the AmoR and AmoD proteins have been proposed but future studies at the protein
262 level are needed to determine their roles in the process of ammonia oxidation by *Nitrosococcus*.
263

264 **ACKNOWLDGMENTS**

265

266 We would like to thank D.J. Schultz (UofL) for valuable advice on the RNA manipulations,
267 L.Y. Stein (UC-Riverside) and three anonymous reviewers for critical reading of the manuscript.
268 This project was supported, in part, by incentive funds provided by the UofL-EVPR office, the
269 KY Science and Engineering Foundation (KSEF-787-RDE-007), and the National Science
270 Foundation (EF-0412129).

271 **REFERENCES**

272

273 1. **Alzerreca, J. J., J. M. Norton, and M. G. Klotz.** 1999. The *amo* operon in marine,
274 ammonia-oxidizing Gammaproteobacteria. FEMS Microbiology Letters **180:**21-29.
275 2. **Arp, D. J., P. S. G. Chain, and M. G. Klotz.** 2007. The impact of genome analyses on
276 our understanding of ammonia-oxidizing bacteria. Ann. Rev. Microbiol. **61:**21-58.
277 3. **Arp, D. J., L. A. Sayavedra-Soto, and N. G. Hommes.** 2002. Molecular biology and
278 biochemistry of ammonia oxidation by *Nitrosomonas europaea*. Arch. Microbiol.
279 **178:**250-255.
280 4. **Bergmann, D. J., A. B. Hooper, and M. G. Klotz.** 2005. Structure and sequence
281 conservation of genes in the *hao* cluster of autotrophic ammonia-oxidizing bacteria:
282 Evidence for their evolutionary history. Appl. Environ. Microbiol. **71:**5371-5382.
283 5. **Berube, P. M., S. C. Proll, and D. A. Stahl.** 2007. Genome wide transcriptional analysis
284 following the recovery of *Nitrosomonas europaea* from ammonia starvation. Abstracts of
285 the 107th General Meeting American Society for Microbiology**:**H-105.
286 6. **Berube, P. M., R. Samudrala, and D. A. Stahl.** 2007. Transcription of all *amoC* copies
287 is associated with recovery of *Nitrosomonas europaea* from ammonia starvation. J.
288 Bacteriol. **189:**3935-3944.
289 7. **Bryson, K., L. J. McGuffin, R. L. Marsden, J. J. Ward, J. S. Sodhi, and D. T. Jones.**
290 2005. Protein structure prediction servers at University College London. Nucl. Acids Res.
291 **33:**W36-38.
292 8. **Chain, P., J. Lamerdin, F. Larimer, W. Regala, V. Lao, M. Land, L. Hauser, A.**
293 **Hooper, M. Klotz, J. Norton, L. Sayavedra-Soto, D. Arciero, N. Hommes, M.**

Whittaker, and D. Arp. 2003. Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph *Nitrosomonas europaea*. J. Bacteriol. **185:**2759-2773.

9. **Estrem, S. T., W. Ross, T. Gaal, Z. W. S. Chen, W. Niu, R. H. Ebright, and R. L. Gourse.** 1999. Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. Genes Dev. **13:**2134-2147.

10. **Hooper, A. B., D. M. Arciero, D. Bergmann, and M. P. Hendrich.** 2005. The oxidation of ammonia as an energy source in bacteria in respiration., p. 121-147. *In* D. Zannoni (ed.), Respiration in archaea and bacteria: Diversity of procaryotic respiratory systems, vol. 2. Springer, Dordrecht, The Netherlands.

11. **Huynen, M. A., and P. Bork.** 1998. Measuring genome evolution. PNAS **95:**5849-5856.

12. **IMG.** 2006. *Nitrosospira multiformis* ATCC 25196. DOE Joint Genome Institute http://img.jgi.doe.gov/cgi-bin/pub/main.cgi?page=taxonDetail&taxon_oid=637000197.

13. **Kall, L., A. Krogh, and E. L. L. Sonnhammer.** 2004. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. **338:**1027.

14. **Klotz, M. G., J. Alzerreca, and J. M. Norton.** 1997. A gene encoding a membrane protein exists upstream of the *amoA*/*amoB* genes in ammonia oxidizing bacteria: a third member of the *amo* operon? FEMS Microbiol. Lett. **150:**65-73.

15. **Klotz, M. G., D. J. Arp, P. S. G. Chain, A. F. El-Sheikh, L. J. Hauser, N. G. Hommes, F. W. Larimer, S. A. Malfatti, J. M. Norton, A. T. Poret-Peterson, L. M. Vergez, and B. B. Ward.** 2006. Complete genome sequence of the marine,

chemolithoautotrophic, ammonia-oxidizing bacterium *Nitrosococcus oceani* ATCC 19707. Appl. Environ. Microbiol. **72:**6299-6315.

16. **Klotz, M. G., and L. Y. Stein.** 2007. Nitrifier genomics and evolution of the N-cycle. FEMS Microbiol. Lett. **277:**(In Press).

17. **Lathe, I., Warren C., B. Snel, and P. Bork.** 2000. Gene context conservation of a higher order than operons. Trends in Biochemical Sciences **25:**474.

18. **Murray, R. G. E., and S. W. Watson.** 1962. Structure of *Nitrocystis oceanus* and comparison with *Nitrosomonas* and *Nitrobacter*. Journal of Bacteriology **89:**1594-1609.

19. **Nakai, K., and M. Kanehisa.** 1991. Expert system for predicting protein localization sites in Gram-negative bacteria. Proteins **11:**95-110.

20. **Norton, J. M., J. J. Alzerreca, Y. Suwa, and M. G. Klotz.** 2002. Diversity of ammonia monooxygenase operon in autotrophic ammonia-oxidizing bacteria. Arch. Microbiol. **177:**139-149.

21. **Sayavedra-Soto, L. A., N. G. Hommes, J. J. Alzerreca, D. J. Arp, J. M. Norton, and M. G. Klotz.** 1998. Transcription of the *amoC*, *amoA* and *amoB* genes in *Nitrosomonas europaea* and *Nitrosospira* sp. NpAV. FEMS Microbiol. Lett. **167:**81-88.

22. **Stein, L. Y., D. J. Arp, P. M. Berube, P. S. G. Chain, L. J. Hauser, M. S. M. Jetten, M. G. Klotz, F. W. Larimer, J. M. Norton, H. J. M. Op den Camp, M. Shin, and X. Wei.** 2007. Whole-genome analysis of the ammonia-oxidizing bacterium, *Nitrosomonas eutropha* C91: Implications for niche adaptation. Environ. Microbiol. **9:**(In Press).

23. **Swofford, D. L.** 1999. PAUP (phylogenetic analysis using parsimony)*, vol. 4.0.b10. Sinauer Associates, Sunderland, MA.

24. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.**

1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. **24:**4876-4882.

25. **Watson, S.** 1965. Characteristics of a marine nitrifying bacterium, *Nitrosocystis oceanus* sp. n. Limnol. Oceanogr. R274-289.

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

TABLE 1

Primers for Northern Analysis (Fig. 1):

Target	Primers
amoC	5'-TGCCTGGCGTGGCATGTGGTTAG-3'; forward
	5'-AATAACCCAACGCCATAAACAACCCA-3'; reverse
amoA	5'-GCTAAAGTCTTTAGAACGTTGGA-3'; forward
	5'-TCACCTGCTAACACCCCTAGCGT-3'; reverse
amoB	5'-TATCAGCATGACGGTTGAAATCAC-3'; forward
	5'-TCTCATTCCCCTCTGGATCAAC-3'; reverse
orf5 (*amoD*)	5'-AGCTGCCTTGTATCGTTTGGA-3' ; forward
	5'-TGGTAAAATCGGTATCAAGCTCA-3'; reverse

Primers for primer extension (Fig. 2):

amoC leader	5'-TGGCTACGCTTATTCTTCAAGGACCCCGA-3'; X1

Primers for RT-PCRs and PCR from genomic DNA (Fig. 3)

amoC upstream (Figs. 3B&C, Lanes 2)	
	5'-GGTTGCTTGCCATAAAGCCGA-3'; *tsp*-F
	5'-CTACAGCTCTACTAGTTGCAGCCATATTGATAGCCTCCT-3'; *R-C*R
amoR-amoC intergenic spacer (Figs. 3B&C, Lanes 3)	
	5'-AAAAGCTTAATGGTGCCCCAAGCTCGTGGGCGT-3'; *R-C*F
	5'-CTACAGCTCTACTAGTTGCAGCCATATTGATAGCCTCCT-3'; *R-C*R

amoC-amoA intergenic spacer (Figs. 3B&C, Lanes 4)

5'-TTAGCGAAGGGTTGAATAGAAGGG-3'; *C-A*F

5'-AGTGCACTCATTAAACCTGCCCTCC-3'; *C-A*R

amoA-amoB intergenic spacer (Figs. 3B&C, Lanes 5)

5'-CCGCTGGTTCTCCAAGGACTAC-3'; *A-B*F

5'-TCGAACGAGGGACGAACATACCAT-3'; *A-B*R

amoB-orf5 (*amoD*) intergenic spacer (Figs. 3B&C, Lanes 6)

5'-CCTATCGGCGGTCCATTAGTTCCCA-3'; *B-D*F

5'-CCCCATGGGCCATGGCGGAAGT-3'; *B-D*R

16S rRNA gene (Figs. 3B&C, Lanes 2 - 6)

5'-GTTTGATCATGGCTCAGATTG-3'; F1-16S

(hybridizing to GenBank: NC_007484; 999386-999406)

5'-CACTGGTGTTCCTTCTTCCGATA-3'; R2-16S

(hybridizing to GenBank: NC_007484; 1000104 – 1000085)

389 **FIGURE LEGENDS**

390

391 **Figure 1.** Northern hybridization analysis using probes based on the *amoC*, *amoA*, *amoB* and
392 *orf5* genes. The sizes of the observed bands are indicated based on a 9-kb RNA ladder. The blot
393 with the probe targeting *orf5* yielded only the larger 4.6-kb band.

394

395 **Figure 2.** Primer extension experiment performed to identify the transcriptional start points
396 (*tsp*) of the two transcripts observed in Northern analysis (Fig. 1). **Panel A** shows the annotated
397 sequence upstream of the *amoC* gene, including *orf1 (amoR)*. **Panel B:** Lane 1 contains the
398 single-stranded DNA ladder and lane 2 contains the extension product obtained with primer X1
399 (see Panel A and Table 1).

400

Figure 3. Map of the *amo* gene cluster based on the genome sequence of *Nitrosococcus oceani* ATCC 19707 (A) and PCR amplification of intergenic regions in the *amo* gene cluster using cDNA (B) and gDNA (C) as the template. **Panel A:** The location of primers used for amplification of intergenic sequence from cDNA are indicated above and the locations of the sequence complementary to probes used for Northern analysis are indicated as horizontal bars within the arrows that indicate length and location of the genes in the *amo* gene cluster. Transcriptional start sites and terminators are indicated by flags and circles, respectively. Open circles indicate leaky terminators. **Panels B and C:** cDNA and gDNA, respectively, were PCR-amplified with primers indicated in (A) and listed in Table 1. Each reaction also included primers for amplification of the 16S rRNA gene. Lane 1, ladder; lane 2 *tsp*F *&* *R-C*R; lane 3, *R-C*F & *R-C*R; lane 4, *C-A*F & *C-A*R; lane 5, *A-B*F & *A-B*R; lane 6, *B-D*F & *B-D*R.

Figure 4. Phylogenetic consensus tree constructed after Bayesian analysis of an alignment of available AmoD (*orf5*) and AmoE (*orf4*) protein sequences. The tree was rooted using the AmoD homologue encoded adjacent to particular methane monooxygenase genes (*pmoCAB*) in the alphaproteobacterial methanotroph, *Methylosinus trichosporium*. Posterior probability values are indicated at the nodes, branch lengths reflect the evolutionary distance based on the standard provided for 50 changes over time. Labels indicate sequence source and location in the source genome; the asterisk indicates gammaproteobacterial methanotroph sources and shaded labels identify singleton status of the encoding genes. Structural analysis identified the AmoD and AmoE proteins as periplasmic membrane proteins.

A

```
                                                                                | -164 ----------------------------
CATGTCAGTAACAAAAAGGCGTATCTGCTCAGGGTGGtAGTTCTCGGGAAAAAGGTTAAatATCAAGGCTATTATGGGTTGAGCCAAGTGTCGCCAAGGTT
                                    -35                        -10         +1
--------------------------------------------------------------------------------------------------------
TGAGGGGAAAATGGTTGCTTGCCATAAAGCCGATAATGGGCTCCGTGTGAGTGAAGCTCGGTATTTGTTTGTTGGGAAATTTTATTAGCGAGATCCAGAG
            |======= tspF =====>>
----<================ X1 ========|                                                              AmoR
TTAATCGGGGTCCTTGAAGAATAAGCGTAGCCATGGATGAGAATTTTTACCGAAATTATAGAGATCCCAAATCATAACATTCCAAAAAAAAGATTGATGAG
                                                |========== R-CF =======>>         S/D       M  R
AAATAGAAGCACCAAAGAAGTGAGGAGACAGGCTGTCCCTTTCCAGCAGATTAATGGTGCCCCAAGCTCGTGGGCGTTATCAGCTATGATGAACTCTCCA
 N  R  S  T  K  E  V  R  R  Q  A  V  P  F  Q  Q  I  N  G  A  P  S  S  W  A  L  S  A  M  M  N  S  P
TACTGTACGAGCCATTGGATTTTTCCAAAAGTAGATGGTGAAAAAAGTATTAAAAATTCGTGTTTATTATGGCTTGACAGGCTTTACACGAGTGTCACAC
Y  C  T  S  H  W  I  F  P  K  V  D  G  E  K  S  I  K  N  S  C  L  L  W  L  D  R  L  Y  T  S  A  T  L
TTAGTTGCTAAAGGACTTGCATGGTTATAAGAAAGGATTTGATATTCCCTCAAAGTAATAAACTTAAAACGAGGGAAGCATATAGGAAGTGAGGGCCTAA
  S  C  *                                                    <==================== R-CR ======|
GGGCCGGTGTTTCATGGCTGGTCAAGTGTCATCACTATTACAATGGTTTCAGTCAAATAGGAGGCTATCAATATGGCTGCAACTAGTAGAGCTGTAGCGC
                                                          S/D          M  A  A  T  S  R  A  V  A  Q
                                                                       AmoC
```

B