Selective Enrichment of *Nitrososphaera viennensis*-like Ammonia-Oxidizing Archaea over Ammonia-Oxidizing Bacteria from Drinking Water Biofilms

Yissue Woo\textsuperscript{a,b,c}, Mercedes Cecilia Cruz\textsuperscript{a#}, and Stefan Wuertz\textsuperscript{a,c}* 

\textsuperscript{a}Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore

\textsuperscript{b}Singapore Centre for Environmental Life Sciences Engineering, Interdisciplinary Graduate Programme, Nanyang Technological University, Singapore

\textsuperscript{c}School of Civil and Environmental Engineering, Nanyang Technological University, Singapore

* Address correspondence to: Stefan Wuertz, swuertz@ntu.edu.sg

\textsuperscript{#} Present Address: (1) Instituto de Investigaciones para la Industria Química (INIQUI), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) National Scientific and Technical Research Council of Argentina (CONICET), Godoy Cruz 2290, Piso 9 (C1425FQB), Ciudad Autónoma de Buenos Aires, República Argentina. (2) Department of Biological Sciences, Marquette University, 1428 W Clybourn Street, Milwaukee, Wisconsin 53233, US. Mail: mccruz@conicet.gov.ar
Figure S1. Maximum likelihood phylogenetic tree based on archaeal 16S rRNA gene nucleotide sequences. The cloned archaeal 16S rRNA sequences are most related to the 16S rRNA gene sequence from *Nitrososphaera viennensis* strain EN76 (closed circle). The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model (80). The unrooted tree with the highest log likelihood (-8752.35) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.0804)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 34 nucleotide sequences. All positions with less than 90% site coverage were eliminated, i.e., fewer than 10% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There was a total of 1397 positions in the final dataset. The number of bootstrap replications was 500. Evolutionary analyses were conducted in MEGA X (67, 81).
Figure S2. Maximum likelihood phylogenetic tree based on archaeal amoA gene nucleotide sequences. The cloned archaeal amoA sequences (closed triangles) are most closely related to the amoA sequence from *Nitrososphaera viennensis* strain EN76 (closed circle). The evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible model (80). The unrooted tree with the highest log likelihood (-4008.34) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.0704)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 37.69% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 30 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions with less than 90% site coverage were eliminated, i.e., fewer than 10% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There was a total of 600 positions in the final dataset. The number of bootstrap replications was 500. Evolutionary analyses were conducted in MEGA X (67, 81).
Figure S3. Maximum likelihood phylogenetic tree based on translated peptide sequences of cloned archaeal amoA nucleotide sequences. The translated peptide sequences of the cloned archaeal amoA are closely related to the amoA peptide sequences from the AOA genera *Nitrososphaera*, including the species *Nitrososphaera viennensis* (closed circle). The cloned sequences are highlighted by closed triangles. The evolutionary history were inferred using the Maximum Likelihood method (80) and Le_Gascuel_2008 model (82). The unrooted tree with the highest log likelihood (-1321.91) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.2277)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 36 amino acid sequences. All positions with less than 90% site coverage were eliminated, i.e., fewer than 10% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). The final dataset had a total of 210 positions. The number of bootstrap replications was 500. Evolutionary analyses were conducted in MEGA X (67, 81).
- Singapore DWDS AOB 16s ASV5
- Singapore DWDS AOB 16s ASV7
- Singapore DWDS AOB 16s ASV2
- Singapore DWDS AOB 16s ASV1
- Singapore DWDS AOB 16s ASV4
- Singapore DWDS AOB 16s ASV3
- Singapore DWDS AOB 16s ASV6

AY123810.1:576-896 Nitrosonomas sp. Nm47 16S rRNA gene partial sequence
  - AY958677.1:488-806 Nitrosonomas sp. NL7 16S rRNA gene partial sequence
  - EF016119.1:567-887 Nitrosonomas oligotropha isolate AS1 16S rRNA gene partial sequence
  - NR 104820.1:574-894 Nitrosonomas oligotropha strain Nm45 16S rRNA gene partial sequence
  - AJ621027.1:574-894 Nitrosonomas sp. Is32 16S rRNA gene isolate Is32
    - AF272422.1:593-913 Nitrosonomas oligotropha 16S rRNA gene partial sequence
    - KF228157.1:594-865 Nitrosonomas oligotropha clone G10-0AF4B 11038 16S rRNA gene partial sequence
        - EU849155.1:576-832 Nitrosonomas sp. VKMM063 16S rRNA gene partial sequence
          - AY123797.1:576-896 Nitrosonomas sp. Nm84 16S rRNA gene partial sequence
            - KF228156.1:581-901 Nitrosonomas aestuarilae clone H11-0AF4A 11038 16S rRNA gene partial sequence
              - AY123811.1:576-896 Nitrosonomas sp. Nm59 16S rRNA gene partial sequence
                - AB000700.1:594-914 Nitrosonomas sp. JL21 gene for 16S rRNA partial sequence
                  - AY123798.1:576-896 Nitrosonomas sp. Nm86 16S rRNA gene partial sequence
                    - AF272423.1:594-914 Nitrosonomas cryotolerans 16S rRNA gene partial sequence
                      - M96403.1:587-909 Nitrosococcus mobilis 16S rRNA (16S rRNA) gene sequence
                        - EF015571.1:572-892 Nitrososppila sp. 17SS 16S rRNA gene partial sequence
                          - AY123813.1:574-894 Nitrososppila sp. Nsp65 16S rRNA gene partial sequence
                            - NR 119316.1:505-827 Nitrosonomas marina strain Nm22 16S rRNA partial sequence
                              - AF272413.1:593-915 Nitrosonomas halophila 16S rRNA gene partial sequence
                                - NR 114767.1:572-894 Nitrosonomas halophila strain Nm1 16S rRNA partial sequence
Figure S4. Maximum likelihood phylogenetic tree based on 16S amplicon sequence variant (ASV) nucleotide sequences, which are closely related to the 16S sequences from the *Nitrosomonas oligotropha* group (closed circles) of AOB. The ASV sequences are highlighted by the filled diamonds. The evolutionary history was inferred by using the Maximum Likelihood method and Kimura 2-parameter model (80). The unrooted tree with the highest log likelihood (-883.41) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.7320)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 43.93% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 27 nucleotide sequences. All positions with less than 90% site coverage were eliminated, i.e., fewer than 10% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There were a total of 321 positions in the final dataset. The number of bootstrap replications was 500. Evolutionary analyses were conducted in MEGA X (67, 81).
- Singapore DWDS AOB amoA Clone 3
- Singapore DWDS AOB amoA Clone 4
  - KU747129:1:269-759 Nitrosomonas oligotropha strain Nm75 ammonia monoxygenase A (amoA) and ammonia monoxygenase B (amoB) genes partial cds
  - KU747138:1:146-636 Nitrosomonas sp. strain Nm173 ammonia monoxygenase A (amoA) and ammonia monoxygenase B (amoB) genes partial cds
  - CP002876:1:506672-507162 Nitrosomonas sp. Is79A3 complete genome
  - AF327918:1:326-816 Nitrosomonas sp. AL212 ammonia monoxygenase subunit A (amoA) gene complete cds
  - LT907782:1:191582-192072 Nitrosomonas ureae strain Nm15 genome assembly chromosome: I
  - KU747130:1:203-693 Nitrosomonas ureae strain Nm10 ammonia monoxygenase A (amoA) and ammonia monoxygenase B (amoB) genes partial cds
  - Singapore DWDS AOB amoA Clone 2
  - Singapore DWDS AOB amoA Clone 9
    - HM345616:1 Nitrosomonas aestuarii isolate SF AOB C09 ammonia monoxygenase subunit A (amoA) gene partial cds
    - HM345621:1 Nitrosomonas marina isolate SF AOB F12 ammonia monoxygenase subunit A (amoA) gene partial cds
    - EU670848:1 Nitrosomonas sp. N2005 ammonia monoxygenase subunit A (amoA) gene partial cds
      - AF314753:1:1149-1636 Nitrosomonas cryotolerans ammonia monoxygenase operon partial sequence and unknown gene
      - DQ228464:1:148-626 Nitrosospira sp. NIJS18 ammonia monoxygenase subunit A (amoA) gene partial cds
      - DQ228463:1:148-626 Nitrosospira sp. NIJS16 ammonia monoxygenase subunit A (amoA) gene partial cds
      - CP021106:3:575447-575937 Nitrosospira lacus strain APG3 complete genome
      - AJ298700:1:1-410 Nitrosococcus oceani partial amoA gene for ammonia monoxygenase subunit A strain Nc1
  - Singapore DWDS AOB amoA Clone 5
    - KU747127:1:203-693 Nitrosomonas nitrosa strain Nm90 ammonia monoxygenase A (amoA) and ammonia monoxygenase B (amoB) genes partial cds
    - KU747126:1:203-683 Nitrosomonas communis strain Nm2 ammonia monoxygenase A (amoA) and ammonia monoxygenase B (amoB) genes partial cds
    - JN999309:1:2-492 Nitrosomonas europaea ATCC 19178 ammonium monoxygenase (amoA) gene partial cds
      - JN387455:1 Nitrosococcus sp. LT-3 ammonia monoxygenase (amoA) gene partial cds
Figure S5. Maximum likelihood phylogenetic tree based on the cloned bacterial amoA nucleotide sequences, which are diversely related to amoA nucleotide sequences from the AOB genera *Nitrosomonas* and *Nitrosococcus*, including the species *Nitrosomonas oligotropha* (closed circle). The cloned sequences are highlighted by the filled diamonds. The evolutionary history was inferred by using the Maximum Likelihood method and Kimura 2-parameter model (80). The unrooted tree with the highest log likelihood (-4809.00) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.0460)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 25.57% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 28 nucleotide sequences. All positions with less than 90% site coverage were eliminated, i.e., fewer than 10% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There were a total of 479 positions in the final dataset. The number of bootstrap replications was 500. Evolutionary analyses were conducted in MEGA X (67, 81).
Figure S6. Maximum likelihood phylogenetic tree based on translated peptide sequences of cloned bacterial amoA nucleotide sequences. The translated peptide sequences of the cloned bacterial amoA are closely related to the amoA peptide sequences from the AOB genera *Nitrosomonas*, including *Nitrosomonas oligotropha* (closed circle). The cloned sequences are highlighted by closed diamonds. One of the cloned sequences is diversely related to the amoA peptide sequences from the AOB genera *Nitrosomonas*, *Nitrosospira* and *Nitrosococcus*. The evolutionary history was inferred using the Maximum Likelihood method (80) and Le_Gascuel_2008 model (82). The unrooted tree with the highest log likelihood (-1454.32) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.7441)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 33 amino acid sequences. All positions with less than 90% site coverage were eliminated, i.e., fewer than 10% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There was a total of 144 positions in the final dataset. The number of bootstrap replications was 500. Evolutionary analyses were conducted in MEGA X (67, 81).
Figure S7. Comparison of ddPCR reactions using NSS_amoAF/R and Arch_amoAF/R primers. 1D plots from the archaeal amoA ddPCR droplet reads showed good signal separation for the positive droplets from the negative droplets of the reactions using the designed primers (A), but the signal separation between the positive and negative droplets of the reactions using the published primers (71) was poorer (B).
Figure S8. Comparison of ddPCR reactions using NSM_amoAF/R and amoA332F/822R PCR reactions. 1D plots from the bacterial amoA ddPCR droplet reads showed good signal separation for the positive droplets from the negative droplets of the reactions using the designed primers (A), but the signal separation between the positive and negative droplets of the reactions using the published primers (72) was poorer and the positive control plasmids failed to amplify (B).
Figure S9. The total taxa sequenced from the enrichment without treatment. The taxa are faceted by genera and colored by phyla. Proteobacteria phylum appeared to be the dominant taxa, with *Hyphomicrobium* being the dominant genus in the enrichment. NA refers to taxa not identifiable to genus level.
Figure S10. The total taxa sequenced from the enrichment with DMTU and pyruvate treatment. The taxa are faceted by genera and colored by phyla. AOA (Thaumarchaeota) phylum and Bacteroidetes phylum appeared to be the dominant taxa. NA refers to taxa not identifiable to genus level.
Figure S11. The total taxa sequenced from the enrichment with DMTU and pyruvate treatment and filtration with 0.45 µm membrane filter. The taxa are faceted by genera and colored by phyla. AOA (Thaumarchaeota) phylum and Proteobacteria phylum appeared to be the dominant taxa. NA refers to taxa not identifiable to genus level.
Figure S12. Set up of the enrichment culture vessel: Nalgene™ Polycarbonate Magnetic Culture Vessel; Silicon heating jacket with thermostat; Sampling tube with valve; filtered air supply tube with aeration stone; filtered air vent.
Figure S13. Fluorescent in-situ hybridization of the AOA enrichment observed with confocal laser scanning microscopy under 100x objective with oil immersion. The scale bar is at 1 µm. AOB cells were not detected, and total cells were stained with DAPI. The AOA cells have a diameter of approximately < 0.5 µm.
Figure S14. Fluorescent *in-situ* hybridization of the AOB enrichment observed with confocal laser scanning microscopy under 100x objective with oil immersion. The scale bar is at 1 µm. AOA cells were not detected although faint Archaea 16S signals were generated, and total cells were stained with DAPI. The AOB cells are approximately 0.5 – 1 µm long.