Supplementary information

Genomic Characteristics of a Novel Species of Ammonia-oxidizing Archaea from the Jiulong River Estuary

Dayu Zou[1,2,3], **Ru Wan**[4], **Lili Han**[4], **Min Nina Xu**[4], **Yang Liu**[2], **Hongbin Liu**[3,5], **Shuh-Ji Kao**[4*], **Meng Li**[1, 2 *]

[1]SZU-HKUST Joint PhD Program in Marine Environmental Science, Shenzhen University, Shenzhen, China

[2]Shenzhen Key Laboratory of Marine Microbiome Engineering, Institute for Advanced Study, Shenzhen University, Shenzhen, China

[3]Department of Ocean Science, The Hong Kong University of Science and Technology, Hong Kong SAR, China

[4]State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, China

[5]Hong Kong Branch of Southern Marine Science & Engineering Guangdong Laboratory, The Hong Kong University of Science and Technology, Hong Kong SAR, China

Running Head: Novel AOA species from the Jiulong River Estuary

Keywords: Nitrification, ammonia-oxidizing archaeon, metagenome, Jiulong River Estuary

***Corresponding author:** Meng Li (limeng848@szu.edu.cn), Shuh-Ji Kao (sjkao@xmu.edu.cn).

List of Tables:

List of Figures:

Figure S8 COG comparisons between the new AOA genus and other AOA genus.

Figure S9 The pan-genomic analysis of AOA MAGs in this study (in light blue) and other AOA isolates.

Supplementary Tables

Table S1 Sequences for each plasmid used in constructing the standard curve.

Target genes	Sequences
Bacterial and Archaeal 16S	TGTGCCAGCAGCCGCGGTAACACCGGCTCTTCTAGTGGTGGCCGTTTTTATTGGGCTTAAAGCGTTCGTAGCCGGATTTGTAAGTTCCTGGCGAAATTCTGTCGCTTAACGATAGAGCTTGCTGGGAATACTGCTTATCTTGAGACCGGGAGAGGCTAGAGGTACTCCTAGGGTAGGGGTGAAATCCGTTAATCCTAGGGGGACCACCTGTGGTGAAGACGTCTAGCTGGAACGGGTCTGACGGTGAGGGACGAAACCCAGGGGAGCAAACCGGATTAGATACCCGTGTAGTCCA
Archaeal 16S	TCCCGCCAATTCCTTTAAGTTTCAGCCTTGCGGCCGTACTCCCCAGGCGGCAGGCTTAACGGCTTCCCTGCGGCACTAGAACGGCACGTAGCCGTCCAAACACCTAGCCTGCATCGTTTACAGCTGGGACTACCCGGGTATCTAATCCGGTTCGCTCCCCCAGCTTTCATCCCTCACCGTCGGGCGCGTTCTAGCCGACAGCCTTCGCCACTGGTGGTCTTTCCGGGATTATAGGATTTCACCCCTACCCCGAAAATACCGTCGGCCTCTCCCGTCCCCTAGCCAAGCAGTATCCCCCGCACACCAACGATTAAATCGGTGGATTTAACAGGGGACTTGCCCAGCCAGCTACGGATGCTTTAGGCCCAATAAACATCCACACCACTCGGAGAGCTGGTATTACCGCGGCGGCTGA
AOB *amoA*	TGGGGTTTTTACTGGTGGTCCCATTATCCGATTAACTTTGTTCTGCCCTCCACGATGATTCCCGGAGCGCTGATCATGGATACATGTCTGTTGTTAACCCGTAACTGGATGATTACAGCACTGTTCGGTGGCGGAGCATTTGGACTATTGTTCTACCCGGGTAACTGGCCGATCTTCGGACCAACTCACCTTCCATTAGTAGTTGAAGGCGTATTGCTGTCATTGGCTGATTACACTGGATTCCTTTATGTTCGTACTGGTACACCTGAGTATGTTCGGTTAATTGAACAAGGATCGCTACGTACCTTCGGTGGTCACACTACGGTAATTGCTGCATTCTTTGCGGCTTTTGTATCAATGTTAATGTTCGTTGTCTGGTGGTACCTGGGAAGGTTCTATTGCACCTCGTTCTATTACGTCAAAGGTCCTAGAGGACGGATAACAGAGAAGAATGATGTTACCGCTTTTGGCGAAGAAGGCTTTCCCGAGGGGA
AOA *amoA*	TGCCATCCACCTGTATGTCCACGTAGTTAGTTTACATCCCAACGCCGTAAACGTTACAGATAACACTGCTCCAGCTCCTGCGCCAAGCGCAACTGGGCTGTTATAGAACTTTCCTACCTGCGGTTCTATCGGTGTCATGTATGGAGGCAACGTAGGTCTTGGATACTTGAATGCAGTCTCCAATGGATCTGCTACTGTAATCAGGTTTACCATGTTGAACAATGGTAACGACATTCCACACAGTACTCCGCCGAACAGTATCAGAGAGTGCTTGTTCTTCTTTGTCGCCCAATATGCCAAATCCAACAGCATTGCCGATGGAATCCACACTGGTGTTACAATGAAGTCATATGGATATCCTAGTGCAAACCATGCACCTTTTGCTACCCATGTATACACCGTCATAATCAACGCGTAGTATGTTGCTGTTCCTGGAACTCCTGTAAATGTTAGGTAATATGTTGCACCTACACAAAGCATCAAAGTTTGTGAGATGGAAAATACTACAAAGGATGTCCAAGCCCAATCTGTGTAGAAGATATAGTCACCTGCATTAATTGTAAGCAGAGTAGAATTGACTGCTACCACTACTATGAATAAGTAATGAGTACATCGTCTAAGCCAGACCATA

Table S2 Detailed qPCR results of samples.

	W1(copies/mL)	W2(copies/mL)	W3(copies/mL)	S1(copies/g)	S2(copies/g)	S3(copies/g)
16S rRNA (Bacteria+ Archaea)	7.25e5(2.78e4)	7.91e5(2.64e4)	7.86e5(7.61e4)	4.46e9(2.66e8)	4.80e9(3.35e7)	1.05e9(9.99e7)
16S rRNA (Archaea)	1.41e5(2.38e4)	2.32e5(2.15e4)	3.09e5(3.66e4)	1.40e9(2.55e8)	2.14e9(2.04e8)	3.72e8(1.87e7)
amoA (Bacteria)	1.38e3(3.68e1)	1.24e3(1.10e2)	3.46e2(1.70e1)	2.27e7(2.25e6)	6.80e6(7.34e5)	4.62e5(4.53e4)
amoA (Archaea)	4.69e2(4.21e1)	3.54e4(1.81e3)	1.81e5(9.46e3)	1.10e7(6.74e5)	1.01e7(4.98e5)	1.80e6(1.09e5)

Table S3 Summary of assembled scaffolds.

	W1	W2	W3	S1	S2	S3
Raw reads (GB)	119.86	127.32	122.87	139.37	135.05	142.44
Assembled scaffolds (GB)	3.51	3.41	3.10	4.03	4.38	4.22
No. of assembled scaffolds	3184430	3194728	2933303	2933303	4133450	3517306
N50 (bp)	1556	1485	1482	1384	1354	1658
Longest scaffolds (bp)	358766	871881	1248239	535704	330828	826502

Table S4 CAZys identified in AOA MAGs in this study.

MAGs	Scaffold No.	Category	Scaffold No.	Category
S1bin1	**contig_1_45**	**GT2**	**contig_10_37**	**GT2**
	contig_186_4	**GT2**	**contig_332_1**	**GT2**
	contig_12_3	**GT1**	**contig_219_3**	**GT1**
	contig_18_26	AA6	contig_1_69	GH18
	contig_2_75	GH101		

S2bin1	**contig_106_4**	**GT2**	**contig_17_12**	**GT2**
	contig_38_1	**GT2**	**contig_30_6**	**GT2**
	contig_78_3	**GT2**	**contig_85_6**	**GT66**
	contig_1_7	**GT1**	contig_367_1	GT1
	contig_18_3	AA6	contig_9_15	GH18
	contig_15_5	GH101		
S3bin1	**contig_24_2**	**GT2**	**contig_18_9**	**GT2**
	contig_7_21	**GT2**	**contig_115_4**	**GT1**
	contig_124_6	**GT1**	contig_189_4	GT1
	contig_1_29	AA6	contig_259_2	GH18
	contig_150_2	GH101		
W1bin1	**scaffold_8857_10**	**GT2**	**scaffold_1213_32**	**GT2**
	scaffold_13204_12	**GT2**	**scaffold_140512_4**	**GT2**
	scaffold_84949_2	**GT2**	**scaffold_1213_15**	**GT1**
	scaffold_7106_19	AA6	scaffold_3058_8	GH18
	scaffold_8859_2	GH101		
W2bin3	**contig_177_3**	**GT2**	**contig_437_3**	**GT2**
	contig_470_1	**GT2**	**contig_476_3**	**GT66**
	contig_446_3	**GT1**	contig_117_7	GH5
	contig_1984_2	GH18	contig_991_2	GH101
	contig_813_2	GH101		
W2bin1	**contig_187_6**	**GT2**	**contig_218_1**	**GT2**

	contig_434_4	**GT1**	**contig_395_6**	**GT1**
	contig_9_2	GT1	contig_232_7	GH101
	contig_1_36	GH101		
W2bin2	contig_209_6	GT1	contig_46_9	GT1
	contig_6_20	GH101	contig_373_1	GH101
W3bin1	**contig_37_12**	**GT2**	**contig_16_26**	**GT2**
	contig_194_2	**GT2**	contig_1_44	GH101
W3bin2	**contig_21_29**	**GT2**	**contig_376_1**	**GT2**
	contig_163_5	**GT2**	**contig_68_15**	**GT1**
	contig_6_4	GT1	contig_115_7	GT1
	contig_4_26	GH101		

* Scaffold No. with bold black indicates the results were verified by at least 2 out of 3 methods in the dbCAN (i.e, HMMER, DIAMOND and Hotpep).

Table S5 Extracellular peptidases and enzymes identified in AOA MAGs in this study.

MAG	Scaffold No.	Annotation
S1bin1	S1bin1_contig_40_10	Gamma-glutamyltranspeptidase / glutathione hydrolase [EC:2.3.2.2 3.4.19.13]
	S1bin1_contig_147_2	Surface layer-associated subtilisin-like serine protease
	S1bin1_contig_644_3	Superoxide dismutase, Fe-Mn family [EC:1.15.1.1]
	S1bin1_contig_440_3	Ribonuclease P protein subunit POP4 [EC:3.1.26.5]
S2bin1	S2bin1_contig_81_1	Gamma-glutamyltranspeptidase / glutathione hydrolase [EC:2.3.2.2 3.4.19.13]

	S2bin1_contig_155_3	Ribonuclease P protein subunit POP4 [EC:3.1.26.5]
	S2bin1_contig_50_9	Surface layer-associated subtilisin-like serine protease
	S2bin1_contig_123_2	Gamma-glutamyltranspeptidase / glutathione hydrolase [EC:2.3.2.2 3.4.19.13]
	S2bin1_contig_64_6	Belongs to the peptidase S8 family
	S2bin1_contig_64_7	Superoxide dismutase, Fe-Mn family [EC:1.15.1.1]
	S2bin1_contig_49_6	Superoxide dismutase, Fe-Mn family [EC:1.15.1.1]
	S2bin1_contig_51_6	nirK; nitrite reductase (NO-forming) [EC:1.7.2.1]
S3bin1	S3bin1_contig_427_3	Ribonuclease P protein subunit POP4 [EC:3.1.26.5]
	S3bin1_contig_134_2	Surface layer-associated subtilisin-like serine protease
	S3bin1_contig_80_2	Superoxide dismutase, Fe-Mn family [EC:1.15.1.1]
	S3bin1_contig_109_3	Belongs to the peptidase S8 family
	S3bin1_contig_109_4	Superoxide dismutase, Fe-Mn family [EC:1.15.1.1]
	S3bin1_contig_87_4	nirK; nitrite reductase (NO-forming) [EC:1.7.2.1]
	S3bin1_contig_154_7	Thiol-disulfide isomerase or thioredoxin
W1bin1	W1bin1_scaffold_3608_37	Superoxide dismutase, Fe-Mn family [EC:1.15.1.1]
	W1bin1_scaffold_17674_14	Belongs to the peptidase S8 family
	W1bin1_scaffold_17674_15	Superoxide dismutase, Fe-Mn family [EC:1.15.1.1]
	W1bin1_scaffold_18718_9	Ribonuclease P protein subunit POP4 [EC:3.1.26.5]
	W1bin1_scaffold_28602_7	Gamma-glutamyltranspeptidase / glutathione hydrolase [EC:2.3.2.2 3.4.19.13]
	W1bin1_scaffold_45384_6	nirK; nitrite reductase (NO-forming) [EC:1.7.2.1]
	W1bin1_scaffold_50458_2	Surface layer-associated subtilisin-like serine protease
W2bin3	W2bin3_contig_420_1	Surface layer-associated subtilisin-like serine protease

	W2bin3_contig_238_10	Gamma-glutamyltranspeptidase / glutathione hydrolase [EC:2.3.2.2 3.4.19.13]
	W2bin3_contig_1577_2	Superoxide dismutase, Fe-Mn family [EC:1.15.1.1]
	W2bin3_contig_735_3	Superoxide dismutase, Fe-Mn family [EC:1.15.1.1]
	W2bin3_contig_735_4	Belongs to the peptidase S8 family
	W2bin3_contig_474_4	Superoxide dismutase, Fe-Mn family [EC:1.15.1.1]
	W2bin3_contig_492_2	Serine protease, subtilisin family
	W2bin3_contig_692_2	Serine protease, subtilisin family
W2bin1	W2bin1_contig_83_3	Superoxide dismutase, Fe-Mn family [EC:1.15.1.1]
	W2bin1_contig_83_4	Belongs to the peptidase S8 family
W2bin2	W2bin2_contig_78_1	Formylglycine-generating sulfatase enzyme
	W2bin2_contig_531_1	Surface layer-associated subtilisin-like serine protease
	W2bin2_contig_1099_1	vgb; virginiamycin B lyase [EC:4.2.99.-]
W3bin1	W3bin1_contig_1364_2	vgb; virginiamycin B lyase [EC:4.2.99.-]
	W3bin1_contig_5_11	Superoxide dismutase, Fe-Mn family [EC:1.15.1.1]
	W3bin1_contig_288_5	Superoxide dismutase, Fe-Mn family [EC:1.15.1.1]
	W3bin1_contig_847_1	DsbG-like Protein-disulfide isomerase
W3bin2	W3bin2_contig_52_15	pepF, pepB; oligoendopeptidase F [EC:3.4.24.-], Peptidase family M3
	W3bin2_contig_22_3	Superoxide dismutase, Fe-Mn family [EC:1.15.1.1]
	W3bin2_contig_22_4	Belongs to the peptidase S8 family

Table S6 Unique gene categories encoded in estuarine AOA MAGs compared with marine AOA.

COG No.	Category	Annotations
Unique gene categories in estuarine AOA		
COG0033	G	Phosphoglucomutase
COG0058	G	Glucan phosphorylase
COG0068	O	Hydrogenase maturation factor HypF (carbamoyltransferase)
COG0198	J	Ribosomal protein L24
COG0208	F	Ribonucleotide reductase beta subunit, ferritin-like domain
COG0210	L	Superfamily I DNA or RNA helicase
COG0213	F	Thymidine phosphorylase
COG0216	J	Protein chain release factor A
COG0235	G	Ribulose-5-phosphate 4-epimerase/Fuculose-1-phosphate aldolase
COG0242	J	Peptide deformylase
COG0255	J	Ribosomal protein L29
COG0276	H	Protoheme ferro-lyase (ferrochelatase)
COG0290	J	Translation initiation factor IF-3
COG0291	J	Ribosomal protein L35
COG0292	J	Ribosomal protein L20
COG0297	G	Glycogen synthase
COG0305	L	Replicative DNA helicase
COG0331	I	Malonyl CoA-acyl carrier protein transacylase
COG0370	P	Fe2+ transport system protein B
COG0404	E	Glycine cleavage system T protein (aminomethyltransferase)
COG0408	H	Coproporphyrinogen III oxidase
COG0450	V	Alkyl hydroperoxide reductase subunit AhpC (peroxiredoxin)
COG0466	O	ATP-dependent Lon protease, bacterial type
COG0477	GEPR	MFS family permease
COG0515	T	Serine/threonine protein kinase
COG0523	R	GTPase, G3E family
COG0551	L	ssDNA-binding Zn-finger and Zn-ribbon domains of topoisomerase 1
COG0557	K	Exoribonuclease R
COG0566	J	tRNA G18 (ribose-2'-O)-methylase SpoU
COG0568	K	DNA-directed RNA polymerase, sigma subunit (sigma70/sigma32)
COG0590	J	tRNA(Arg) A34 adenosine deaminase TadA
COG0612	R	Predicted Zn-dependent peptidase
COG0617	J	tRNA nucleotidyltransferase/poly(A) polymerase
COG0618	F	nanoRNase/pAp phosphatase, hydrolyzes c-di-AMP and oligoRNAs
COG0633	C	Ferredoxin
COG0655	C	Multimeric flavodoxin WrbA
COG0664	T	cAMP-binding domain of CRP or a regulatory subunit of cAMP-dependent protein kinases
COG0669	H	Phosphopantetheine adenylyltransferase
COG0711	C	FoF1-type ATP synthase, membrane subunit b or b'
COG0721	J	Asp-tRNAAsn/Glu-tRNAGln amidotransferase C subunit
COG0727	R	Fe-S-cluster containining protein

COG0742	J	16S rRNA G966 N2-methylase RsmD
COG0747	E	ABC-type transport system, periplasmic component
COG0749	L	DNA polymerase I - 3'-5' exonuclease and polymerase domains
COG0756	FV	dUTPase
COG0776	L	Bacterial nucleoid DNA-binding protein
COG0777	I	Acetyl-CoA carboxylase beta subunit
COG0789	K	DNA-binding transcriptional regulator, MerR family
COG0794	GM	D-arabinose 5-phosphate isomerase GutQ
COG0800	G	2-keto-3-deoxy-6-phosphogluconate aldolase
COG0834	ET	ABC-type amino acid transport/signal transduction system, periplasmic component/domain
COG0859	M	ADP-heptose:LPS heptosyltransferase
COG0861	P	Membrane protein TerC, possibly involved in tellurium resistance
COG1002	V	Type II restriction/modification system, DNA methylase subunit YeeA
COG1045	E	Serine acetyltransferase
COG1055	P	Na+/H+ antiporter NhaD or related arsenite permease
COG1090	R	NAD dependent epimerase/dehydratase family enzyme
COG1107	L	Archaea-specific RecJ-like exonuclease, contains DnaJ-type Zn finger domain
COG1122	PR	Energy-coupling factor transporter ATP-binding protein EcfA2
COG1123	O	ABC-type glutathione transport system ATPase component, contains duplicated ATPase domain
COG1145	C	Ferredoxin
COG1194	L	Adenine-specific DNA glycosylase, acts on AG and A-oxoG pairs
COG1222	O	ATP-dependent 26S proteasome regulatory subunit
COG1228	Q	Imidazolonepropionase or related amidohydrolase
COG1232	H	Protoporphyrinogen oxidase
COG1246	E	N-acetylglutamate synthase or related acetyltransferase, GNAT family
COG1249	C	Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component or related enzyme
COG1283	P	Na+/phosphate symporter
COG1341	J	Polynucleotide 5'-kinase, involved in rRNA processing
COG1376	M	Lipoprotein-anchoring transpeptidase ErfK/SrfK
COG1401	V	5-methylcytosine-specific restriction endonuclease McrBC, GTP-binding regulatory subunit McrB
COG1406	N	Chemotaxis protein CheX, a CheY~P-specific phosphatase
COG1463	M	ABC-type transporter Mla maintaining outer membrane lipid asymmetry, periplasmic component MlaD
COG1468	V	CRISPR/Cas system-associated exonuclease Cas4, RecB family
COG1489	GT	DNA-binding protein, stimulates sugar fermentation
COG1502	I	Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthase or related enzyme
COG1512	S	Uncharacterized membrane protein YgcG, contains a TPM-fold domain
COG1519	M	3-deoxy-D-manno-octulosonic-acid transferase
COG1520	M	Outer membrane protein assembly factor BamB, contains PQQ-like beta-propeller

		repeat
COG1547	S	Predicted metal-dependent hydrolase
COG1549	R	Predicted RNA-binding protein, contains uracil-DNA-glycosylase-like and PUA domains
COG1552	J	Ribosomal protein L40E
COG1560	I	Lauroyl/myristoyl acyltransferase
COG1566	V	Multidrug resistance efflux pump
COG1591	L	Holliday junction resolvase, archaeal type
COG1617	J	tRNA threonylcarbamoyladenosine modification (KEOPS) complex, Cgi121 subunit
COG1704	S	Uncharacterized conserved protein
COG1778	MR	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase KdsC and related HAD superfamily phosphatases
COG1783	X	Phage terminase large subunit
COG1812	H	Archaeal S-adenosylmethionine synthetase
COG1843	N	Flagellar hook assembly protein FlgD
COG1845	C	Heme/copper-type cytochrome/quinol oxidase, subunit 3
COG1871	NT	Chemotaxis receptor (MCP) glutamine deamidase CheD
COG1879	G	ABC-type sugar transport system, periplasmic component, contains N-terminal xre family HTH domain
COG1893	H	Ketopantoate reductase
COG1913	R	Predicted Zn-dependent protease
COG1924	I	Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)
COG1970	M	Large-conductance mechanosensitive channel
COG1981	S	Uncharacterized membrane protein
COG2009	C	Succinate dehydrogenase/fumarate reductase, cytochrome b subunit
COG2015	Q	Alkyl sulfatase BDS1 and related hydrolases, metallo-beta-lactamase superfamily
COG2057	I	Acyl CoA:acetate/3-ketoacid CoA transferase, beta subunit
COG2061	R	Uncharacterized conserved protein, contains ACT domain
COG2072	P	Predicted flavoprotein CzcO associated with the cation diffusion facilitator CzcD
COG2105	R	Uncharacterized conserved protein YtfP, gamma-glutamylcyclotransferase (GGCT)/AIG2-like family
COG2126	J	Ribosomal protein L37E
COG2142	C	Succinate dehydrogenase, hydrophobic anchor subunit
COG2153	R	Predicted N-acyltransferase, GNAT family
COG2175	Q	Taurine dioxygenase, alpha-ketoglutarate-dependent
COG2177	D	Cell division protein FtsX
COG2217	P	Cation transport ATPase
COG2230	I	Cyclopropane fatty-acyl-phospholipid synthase and related methyltransferases
COG2251	R	Predicted nuclease, RecB family
COG2263	R	Predicted RNA methylase
COG2267	I	Lysophospholipase, alpha-beta hydrolase superfamily
COG2268	S	Uncharacterized membrane protein YqiK, contains Band7/PHB/SPFH domain
COG2271	G	Sugar phosphate permease

COG2274	V	ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal double-glycine peptidase domain
COG2319	R	WD40 repeat
COG2332	CO	Cytochrome c-type biogenesis protein CcmE
COG2333	R	Metal-dependent hydrolase, beta-lactamase superfamily II
COG2339	T	Membrane proteinase PrsW, cleaves anti-sigma factor RsiW, M82 family
COG2345	K	Predicted transcriptional regulator, ArsR family
COG2433	R	Possible nuclease of RNase H fold, RuvC/YqgF family
COG2469	S	Uncharacterized protein, contains HTH domain
COG2510	S	Uncharacterized membrane protein
COG2604	S	Uncharacterized conserved protein
COG2609	C	Pyruvate dehydrogenase complex, dehydrogenase (E1) component
COG2761	O	Predicted dithiol-disulfide isomerase, DsbA family
COG2811	C	Archaeal/vacuolar-type H+-ATPase subunit H
COG2815	M	PASTA domain, binds beta-lactams
COG2816	F	NADH pyrophosphatase NudC, Nudix superfamily
COG2870	M	ADP-heptose synthase, bifunctional sugar kinase/adenylyltransferase
COG2884	D	ABC-type ATPase involved in cell division
COG2885	M	Outer membrane protein OmpA and related peptidoglycan-associated (lipo)proteins
COG2951	M	Membrane-bound lytic murein transglycosylase B
COG2961	J	23S rRNA A2030 N6-methylase RlmJ
COG3072	F	Adenylate cyclase
COG3087	D	Cell division protein FtsN
COG3096	D	Chromosome condensin MukBEF, ATPase and DNA-binding subunit MukB [Escherichia coli str. K-12 substr. MG1655
COG3108	S	Uncharacterized conserved protein YcbK, DUF882 family
COG3153	R	Predicted N-acetyltransferase YhbS
COG3177	K	Fic family protein
COG3212	S	Uncharacterized membrane protein YkoI
COG3291	S	PKD repeat
COG3316	X	Transposase (or an inactivated derivative)
COG3325	G	Chitinase, GH18 family
COG3333	R	TctA family transporter
COG3352	N	Archaellum component FlaC
COG3355	K	Predicted transcriptional regulator
COG3378	X	Phage- or plasmid-associated DNA primase
COG3409	M	Peptidoglycan-binding (PGRP) domain of peptidoglycan hydrolases
COG3480	T	Predicted secreted protein containing a PDZ domain
COG3497	X	Phage tail sheath protein FI
COG3562	M	Capsule polysaccharide modification protein KpsS
COG3565	R	Predicted dioxygenase of extradiol dioxygenase family
COG3577	R	Predicted aspartyl protease
COG3623	G	L-ribulose-5-phosphate 3-epimerase UlaE

COG3628	X	Phage baseplate assembly protein W
COG3677	X	Transposase
COG3721	P	Putative heme iron utilization protein
COG3752	R	Steroid 5-alpha reductase family enzyme
COG3772	M	Phage-related lysozyme (muramidase), GH24 family
COG3773	DM	Cell wall hydrolase CwlJ, involved in spore germination
COG3829	KT	Transcriptional regulator containing PAS, AAA-type ATPase, and DNA-binding Fis domains
COG3832	S	Uncharacterized conserved protein YndB, AHSA1/START domain
COG3881	S	Uncharacterized protein YrrD, contains PRC-barrel domain
COG4049	R	Uncharacterized protein, contains archaeal-type C2H2 Zn-finger
COG4175	E	ABC-type proline/glycine betaine transport system, ATPase component
COG4245	S	Uncharacterized conserved protein YegL, contains vWA domain of TerY type
COG4249	R	Uncharacterized protein, contains caspase domain
COG4276	R	Ligand-binding SRPBCC domain
COG4293	S	Uncharacterized protein
COG4294	L	UV DNA damage repair endonuclease
COG4346	O	Predicted membrane-bound dolichyl-phosphate-mannose-protein mannosyltransferase
COG4372	S	Uncharacterized conserved protein, contains DUF3084 domain
COG4373	X	Mu-like prophage FluMu protein gp28
COG4430	S	Uncharacterized conserved protein YdeI, YjbR/CyaY-like superfamily, DUF1801 family
COG4487	S	Uncharacterized protein
COG4531	P	ABC-type Zn2+ transport system, periplasmic component/surface adhesin
COG4558	P	ABC-type hemin transport system, periplasmic component
COG4627	R	Predicted SAM-depedendent methyltransferase
COG4641	D	Spore maturation protein CgeB
COG4875	S	Uncharacterized protein
COG4942	D	Septal ring factor EnvC, activator of murein hydrolases AmiA and AmiB
COG4982	I	3-oxoacyl-ACP reductase domain of yeast-type FAS1
COG4990	S	Uncharacterized protein YvpB
COG5002	T	Signal transduction histidine kinase
COG5023	S	Uncharacterized protein
COG5148	S	Uncharacterized protein
COG5276	S	Uncharacterized conserved protein
COG5425	S	Usg protein (tryptophan operon, function unknown)
COG5566	K	Transcriptional regulator, Middle operon regulator (Mor) family
COG5632	M	N-acetylmuramoyl-L-alanine amidase CwlA
COG5635	T	Predicted NTPase, NACHT family domain
COG5637	S	Uncharacterized membrane protein
COG5653	N	Acetyltransferase involved in cellulose biosynthesis, CelD/BcsL family
COG5663	S	Uncharacterized protein, HAD superfamily
Unique gene categories in marine AOA		

COG0022	C	Pyruvate/2-oxoglutarate/acetoin dehydrogenase complex, dehydrogenase (E1) component
COG0160	E	4-aminobutyrate aminotransferase or related aminotransferase
COG0236	IQ	Acyl carrier protein
COG0295	F	Cytidine deaminase
COG0426	C	Flavorubredoxin
COG0471	G	Di- and tricarboxylate transporter
COG0582	LX	Integrase
COG0637	GR	Beta-phosphoglucomutase or related phosphatase, HAD superfamily
COG0654	HC	2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases
COG0675	X	Transposase
COG0733	R	Na+-dependent transporter, SNF family
COG0737	FV	2',3'-cyclic-nucleotide 2'-phosphodiesterase/5'- or 3'-nucleotidase, 5'-nucleotidase family
COG0847	L	DNA polymerase III, epsilon subunit or related 3'-5' exonuclease
COG1004	M	UDP-glucose 6-dehydrogenase
COG1044	M	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
COG1071	C	TPP-dependent pyruvate or acetoin dehydrogenase subunit alpha
COG1080	G	Phosphoenolpyruvate-protein kinase (PTS system EI component in bacteria)
COG1083	M	CMP-N-acetylneuraminic acid synthetase
COG1099	R	Predicted metal-dependent hydrolase, TIM-barrel fold
COG1134	GM	ABC-type polysaccharide/polyol phosphate transport system, ATPase component
COG1179	J	tRNA A37 threonylcarbamoyladenosine dehydratase
COG1210	M	UTP-glucose-1-phosphate uridylyltransferase
COG1309	K	DNA-binding transcriptional regulator, AcrR family
COG1335	HR	Nicotinamidase-related amidase
COG1359	C	Quinol monooxygenase YgiN
COG1403	V	5-methylcytosine-specific restriction endonuclease McrA
COG1427	R	Predicted periplasmic solute-binding protein
COG1432	S	Uncharacterized conserved protein, LabA/DUF88 family
COG1497	K	Predicted transcriptional regulator
COG1518	V	CRISPR/Cas system-associated endonuclease Cas1
COG1524	R	Predicted pyrophosphatase or phosphodiesterase, AlkP superfamily
COG1680	V	CubicO group peptidase, beta-lactamase class C family
COG1682	GM	ABC-type polysaccharide/polyol phosphate export permease
COG1715	V	Restriction endonuclease Mrr
COG1737	K	DNA-binding transcriptional regulator, MurR/RpiR family, contains HTH and SIS domains
COG1776	T	Chemotaxis protein CheY-P-specific phosphatase CheC
COG1777	K	Predicted transcriptional regulator
COG2071	E	Gamma-glutamyl-gamma-aminobutyrate hydrolase PuuD (putrescine degradation), contains GATase1-like domain
COG2120	G	N-acetylglucosaminyl deacetylase, LmbE family

COG2249	R	Putative NADPH-quinone reductase (modulator of drug activity B)
COG2303	IR	Choline dehydrogenase or related flavoprotein
COG2357	FT	ppGpp synthetase catalytic domain (RelA/SpoT-type nucleotidyltranferase)
COG2706	G	6-phosphogluconolactonase, cycloisomerase 2 family
COG2717	C	Periplasmic DMSO/TMAO reductase YedYZ, heme-binding membrane subunit
COG2865	K	Predicted transcriptional regulator, contains HTH domain
COG2932	X	Phage repressor protein C, contains Cro/C1-type HTH and peptisase s24 domains
COG2957	E	Agmatine/peptidylarginine deiminase
COG3055	M	N-acetylneuraminic acid mutarotase
COG3083	M	Membrane-anchored periplasmic protein YejM, alkaline phosphatase superfamily
COG3239	I	Fatty acid desaturase
COG3401	R	Fibronectin type 3 domain
COG3430	N	Archaeal flagellin (archaellin), FlaG/FlaF family
COG3465	S	Uncharacterized protein YwgA
COG3584	S	3D (Asp-Asp-Asp) domain
COG3609	K	Transcriptional regulator, contains Arc/MetJ-type RHH (ribbon-helix-helix) DNA-binding domain
COG3883	S	Uncharacterized N-terminal domain of peptidoglycan hydrolase CwlO
COG3980	M	Spore coat polysaccharide biosynthesis protein SpsG, predicted glycosyltransferase
COG4254	R	Uncharacterized conserved protein, contains LysM and FecR domains
COG4274	S	Uncharacterized protein, contains GYD domain
COG4343	V	CRISPR/Cas system-associated exonuclease Cas4, RecB family
COG4422	X	Bacteriophage protein gp37
COG4447	R	Uncharacterized protein related to plant photosystem II stability/assembly factor
COG4477	D	Septation ring formation regulator EzrA
COG4565	KT	Response regulator of citrate/malate metabolism
COG4675	S	Microcystin-dependent protein (function unknown)
COG4973	L	Site-specific recombinase XerC
COG5099	S	Uncharacterized protein
COG5278	T	Extracellular (periplasmic) sensor domain CHASE3 (specificity unknown)
COG5285	Q	Ectoine hydroxylase-related dioxygenase, phytanoyl-CoA dioxygenase (PhyH) family
COG5301	X	Phage-related tail fibre protein
COG5422	S	Uncharacterized protein

Table S7 Unique gene categories encoded in MAGs of the new AOA species in this study compared with other AOA genus or species.

COG No.	Category	Annotations
Jiulong River estuary AOA vs. Nitrosopumilus		
COG0199	J	Ribosomal protein S14

COG0306	P	Phosphate/sulfate permease
COG0500	QR	SAM-dependent methyltransferase
COG0515	T	Serine/threonine protein kinase
COG0655	C	Multimeric flavodoxin WrbA
COG0735	P	Fe2+ or Zn2+ uptake regulation protein
COG1180	O	Pyruvate-formate lyase-activating enzyme
COG1232	H	Protoporphyrinogen oxidase
COG1355	R	Predicted class III extradiol dioxygenase, MEMO1 family
COG1475	D	Chromosome segregation protein Spo0J, contains ParB-like nuclease domain
COG1549	R	Predicted RNA-binding protein, contains uracil-DNA-glycosylase-like and PUA domains
COG1683	S	Uncharacterized conserved protein YbbK, DUF523 family
COG1721	S	Uncharacterized conserved protein, DUF58 family, contains vWF domain
COG1846	K	DNA-binding transcriptional regulator, MarR family
COG1889	J	Fibrillarin-like rRNA methylase
COG1913	R	Predicted Zn-dependent protease
COG2078	S	Uncharacterized conserved protein, AMMECR1 domain
COG2443	U	Preprotein translocase subunit Sss1
COG2770	T	HAMP domain
COG2811	C	Archaeal/vacuolar-type H+-ATPase subunit H
COG2815	M	PASTA domain, binds beta-lactams
COG3119	P	Arylsulfatase A or related enzyme
COG3153	R	Predicted N-acetyltransferase YhbS
COG3325	G	Chitinase, GH18 family
COG3881	S	Uncharacterized protein YrrD, contains PRC-barrel domain
COG4294	L	UV DNA damage repair endonuclease
COG4372	S	Uncharacterized conserved protein, contains DUF3084 domain
COG4639	R	Predicted kinase
COG5002	T	Signal transduction histidine kinase
COG5653	N	Acetyltransferase involved in cellulose biosynthesis, CelD/BcsL family
Jiulong River estuary AOA vs. *Ca.* Nitrosoarchaeum		
COG0312	T	Serine/threonine protein kinase
COG0405	D	Chromosome segregation protein Spo0J, contains ParB-like nuclease domain
COG0421	R	Predicted RNA-binding protein, contains uracil-DNA-glycosylase-like and PUA domains
COG0515	S	Uncharacterized conserved protein YbbK, DUF523 family
COG0636	S	Uncharacterized conserved protein, DUF58 family, contains vWF domain
COG0640	R	Predicted Zn-dependent protease
COG1011	C	Archaeal/vacuolar-type H+-ATPase subunit H
COG1181	M	PASTA domain, binds beta-lactams
COG1301	R	Predicted N-acetyltransferase YhbS
COG1475	G	Chitinase, GH18 family
COG1549	S	Uncharacterized protein YrrD, contains PRC-barrel domain
COG1605	L	UV DNA damage repair endonuclease

COG1683	S	Uncharacterized conserved protein, contains DUF3084 domain
COG1721	T	Signal transduction histidine kinase
COG1828	N	Acetyltransferase involved in cellulose biosynthesis, CelD/BcsL family
COG1913	R	Predicted Zn-dependent protease or its inactivated homolog
COG2015	E	Gamma-glutamyltranspeptidase
COG2222	E	Spermidine synthase
COG2260	C	FoF1-type ATP synthase, membrane subunit c/Archaeal/vacuolar-type H+-ATPase, subunit K
COG2329	K	DNA-binding transcriptional regulator, ArsR family
COG2430	H	FMN phosphatase YigB, HAD superfamily
COG2811	MR	D-alanine-D-alanine ligase and related ATP-grasp enzymes
COG2815	C	Na+/H+-dicarboxylate symporter
COG3153	E	Chorismate mutase
COG3221	F	Phosphoribosylformylglycinamidine (FGAM) synthase, PurS component
COG3325	Q	Alkyl sulfatase BDS1 and related hydrolases, metallo-beta-lactamase superfamily
COG3576	M	Fructoselysine-6-P-deglycase FrlB and related proteins with duplicated sugar isomerase (SIS) domain
COG3614	J	rRNA maturation protein Nop10, contains Zn-ribbon domain
COG3638	H	Heme-degrading monooxygenase HmoA and related ABM domain proteins
COG3639	S	Uncharacterized protein
COG3741	P	ABC-type phosphate/phosphonate transport system, periplasmic component
COG3881	R	Predicted flavin-nucleotide-binding protein, pyridoxine 5'-phosphate oxidase superfamily
COG4262	T	Extracellular (periplasmic) sensor domain CHASE1 (specificity unknown)
COG4294	P	ABC-type phosphate/phosphonate transport system, ATPase component
COG4372	P	ABC-type phosphate/phosphonate transport system, permease component
COG4880	E	N-formylglutamate amidohydrolase
COG5002	R	Predicted spermidine synthase with an N-terminal membrane domain
COG5207	R	Secreted protein containing C-terminal beta-propeller domain distantly related to WD-40 repeats
COG5431	R	Uncharacterized Zn-finger protein, UBP-type
COG5653	R	Predicted nucleic acid-binding protein, contains Zn-finger domain
Jiulong River estuary AOA vs. *Ca. Nitrosomarinus* (SPOT01)		
COG0312	T	Serine/threonine protein kinase
COG0405	D	Chromosome segregation protein Spo0J, contains ParB-like nuclease domain
COG0421	R	Predicted RNA-binding protein, contains uracil-DNA-glycosylase-like and PUA domains
COG0515	S	Uncharacterized conserved protein YbbK, DUF523 family
COG0636	S	Uncharacterized conserved protein, DUF58 family, contains vWF domain
COG0640	R	Predicted Zn-dependent protease
COG1011	C	Archaeal/vacuolar-type H+-ATPase subunit H
COG1181	M	PASTA domain, binds beta-lactams
COG1301	R	Predicted N-acetyltransferase YhbS

COG1475	G	Chitinase, GH18 family
COG1549	S	Uncharacterized protein YrrD, contains PRC-barrel domain
COG1605	L	UV DNA damage repair endonuclease
COG1683	S	Uncharacterized conserved protein, contains DUF3084 domain
COG1721	T	Signal transduction histidine kinase
COG1828	N	Acetyltransferase involved in cellulose biosynthesis, CelD/BcsL family
COG1913	R	Predicted Zn-dependent protease or its inactivated homolog
COG2015	E	Gamma-glutamyltranspeptidase
COG2222	E	Spermidine synthase
COG2260	C	FoF1-type ATP synthase, membrane subunit c/Archaeal/vacuolar-type H+-ATPase, subunit K
COG2329	K	DNA-binding transcriptional regulator, ArsR family
COG2430	H	FMN phosphatase YigB, HAD superfamily
COG2811	MR	D-alanine-D-alanine ligase and related ATP-grasp enzymes
COG2815	C	Na+/H+-dicarboxylate symporter
COG3153	E	Chorismate mutase
COG3221	F	Phosphoribosylformylglycinamidine (FGAM) synthase, PurS component
COG3325	Q	Alkyl sulfatase BDS1 and related hydrolases, metallo-beta-lactamase superfamily
COG3576	M	Fructoselysine-6-P-deglycase FrlB and related proteins with duplicated sugar isomerase (SIS) domain
COG3614	J	rRNA maturation protein Nop10, contains Zn-ribbon domain
COG3638	H	Heme-degrading monooxygenase HmoA and related ABM domain proteins
COG3639	S	Uncharacterized protein
COG3741	P	ABC-type phosphate/phosphonate transport system, periplasmic component
COG3881	R	Predicted flavin-nucleotide-binding protein, pyridoxine 5'-phosphate oxidase superfamily
COG4262	T	Extracellular (periplasmic) sensor domain CHASE1 (specificity unknown)
COG4294	P	ABC-type phosphate/phosphonate transport system, ATPase component
COG4372	P	ABC-type phosphate/phosphonate transport system, permease component
COG4880	E	N-formylglutamate amidohydrolase
COG5002	R	Predicted spermidine synthase with an N-terminal membrane domain
COG5207	R	Secreted protein containing C-terminal beta-propeller domain distantly related to WD-40 repeats
COG5431	R	Uncharacterized Zn-finger protein, UBP-type
COG5653	R	Predicted nucleic acid-binding protein, contains Zn-finger domain

Table S8 Detailed annotations of unique genes encoded in the new AOA species

Gene cluster	MAGs	Contig_id	Annotations
GC_00001152	S1bin1	contig_38	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
GC_00001152	S1bin1	contig_35	DNA-binding response regulator, OmpR family,

			contains REC and winged-helix (wHTH) domain
GC_00001152	S2bin1	contig_19	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
GC_00001152	S3bin1	contig_65	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
GC_00001152	S3bin1	contig_481	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
GC_00001152	S3bin1	contig_21	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
GC_00001152	W1bin1	scaffold_2754	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
GC_00001152	W1bin1	scaffold_2754	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
GC_00001152	W1bin1	scaffold_12641	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
GC_00001152	W2bin3	contig_408	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
GC_00001152	W2bin3	contig_64	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
GC_00001152	W2bin3	contig_2251	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
GC_00001152	W2bin3	contig_147	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
GC_00001438	S1bin1	contig_80	DNA polymerase elongation subunit (family B)
GC_00001438	S1bin1	contig_13	DNA polymerase elongation subunit (family B)
GC_00001438	S2bin1	contig_3	DNA polymerase elongation subunit (family B)
GC_00001438	S3bin1	contig_264	DNA polymerase elongation subunit (family B)
GC_00001438	S3bin1	contig_149	DNA polymerase elongation subunit (family B)
GC_00001438	W1bin1	scaffold_35710	DNA polymerase elongation subunit (family B)
GC_00001438	W1bin1	scaffold_35710	DNA polymerase elongation subunit (family B)
GC_00001438	W2bin3	contig_280	DNA polymerase elongation subunit (family B)
GC_00001438	W2bin3	contig_307	DNA polymerase elongation subunit (family B)
GC_00001589	S1bin1	contig_1	Aryl-phospho-beta-D-glucosidase BglC, GH1 family|Chitinase, GH18 family|SFI04223.1 Chitinase, GH18 family [Planctomicrobium piriforme]
GC_00001589	S2bin1	contig_9	Aryl-phospho-beta-D-glucosidase BglC, GH1 family|Chitinase, GH18 family|SFI04223.1 Chitinase, GH18 family [Planctomicrobium piriforme]
GC_00001589	S3bin1	contig_259	Aryl-phospho-beta-D-glucosidase BglC, GH1 family|Chitinase, GH18 family|SFI04223.1 Chitinase, GH18 family [Planctomicrobium piriforme]

GC_00001589	W1bin1	scaffold_3058	Aryl-phospho-beta-D-glucosidase BglC, GH1 family|Chitinase, GH18 family|SFI04223.1 Chitinase, GH18 family [Planctomicrobium piriforme]
GC_00001589	W2bin3	contig_117	Aryl-phospho-beta-D-glucosidase BglC, GH1 family|Chitinase, GH18 family|SFI04223.1 Chitinase, GH18 family [Planctomicrobium piriforme]
GC_00001589	W2bin3	contig_1984	Aryl-phospho-beta-D-glucosidase BglC, GH1 family|Chitinase, GH18 family|SFI04223.1 Chitinase, GH18 family [Planctomicrobium piriforme]
GC_00001606	S1bin1	contig_1	Ig (Immunoglobulin) family protein [Thaumarchaeota archaeon CSP1-1]
GC_00001606	S2bin1	contig_31	Ig (Immunoglobulin) family protein [Thaumarchaeota archaeon CSP1-1]
GC_00001606	S3bin1	contig_41	Ig (Immunoglobulin) family protein [Thaumarchaeota archaeon CSP1-1]
GC_00001606	W1bin1	scaffold_8857	Ig (Immunoglobulin) family protein [Thaumarchaeota archaeon CSP1-1]
GC_00001606	W2bin3	contig_1236	Ig (Immunoglobulin) family protein [Thaumarchaeota archaeon CSP1-1]
GC_00001606	W2bin3	contig_43	Ig (Immunoglobulin) family protein [Thaumarchaeota archaeon CSP1-1]
GC_00001621	S1bin1	contig_38	YIP1 family protein [Jeotgalibacillus campisalis]
GC_00001621	S3bin1	contig_100	YIP1 family protein [Jeotgalibacillus campisalis]
GC_00001621	S3bin1	contig_21	YIP1 family protein [Jeotgalibacillus campisalis]
GC_00001621	W1bin1	scaffold_29150	YIP1 family protein [Jeotgalibacillus campisalis]
GC_00001621	W2bin3	contig_450	YIP1 family protein [Jeotgalibacillus campisalis]
GC_00001621	W2bin3	contig_128	YIP1 family protein [Jeotgalibacillus campisalis]
GC_00001627	S1bin1	contig_89	Uncharacterized conserved protein YfaS, alpha-2-macroglobulin family
GC_00001627	S2bin1	contig_88	Uncharacterized conserved protein YfaS, alpha-2-macroglobulin family
GC_00001627	S2bin1	contig_129	Uncharacterized conserved protein YfaS, alpha-2-macroglobulin family
GC_00001627	S3bin1	contig_64	Uncharacterized conserved protein YfaS, alpha-2-macroglobulin family
GC_00001627	W1bin1	scaffold_38255	Uncharacterized conserved protein YfaS, alpha-2-macroglobulin family
GC_00001627	W2bin3	contig_1821	Uncharacterized conserved protein YfaS, alpha-2-macroglobulin family
GC_00001661	S1bin1	contig_213	tetratricopeptide repeat protein [Candidatus Nitrosopelagicus brevis]

GC_00001661	S2bin1	contig_73	tetratricopeptide repeat protein [Candidatus Nitrosopelagicus brevis]
GC_00001661	W1bin1	scaffold_26013	tetratricopeptide repeat protein [Candidatus Nitrosopelagicus brevis]
GC_00001661	W2bin3	contig_708	tetratricopeptide repeat protein [Candidatus Nitrosopelagicus brevis]
GC_00001661	W2bin3	contig_653	tetratricopeptide repeat protein [Candidatus Nitrosopelagicus brevis]
GC_00001661	W2bin3	contig_21	tetratricopeptide repeat protein [Candidatus Nitrosopelagicus brevis]
GC_00001677	S1bin1	contig_131	thr operon leader peptide, partial [Nitrosopumilales archaeon CG11_big_fil_rev_8_21_14_0_20_33_24]
GC_00001677	S2bin1	contig_20	thr operon leader peptide, partial [Nitrosopumilales archaeon CG11_big_fil_rev_8_21_14_0_20_33_24]
GC_00001677	S3bin1	contig_593	thr operon leader peptide, partial [Nitrosopumilales archaeon CG11_big_fil_rev_8_21_14_0_20_33_24]
GC_00001677	W1bin1	scaffold_17716	thr operon leader peptide, partial [Nitrosopumilales archaeon CG11_big_fil_rev_8_21_14_0_20_33_24]
GC_00001677	W2bin3	contig_740	thr operon leader peptide, partial [Nitrosopumilales archaeon CG11_big_fil_rev_8_21_14_0_20_33_24]
GC_00001678	S1bin1	contig_18	polyketide cyclase [Candidatus Nitrosotenuis sp. AQ6f]
GC_00001678	S2bin1	contig_18	polyketide cyclase [Candidatus Nitrosotenuis sp. AQ6f]
GC_00001678	S3bin1	contig_1	polyketide cyclase [Candidatus Nitrosotenuis sp. AQ6f]
GC_00001678	W1bin1	scaffold_7106	polyketide cyclase [Candidatus Nitrosotenuis sp. AQ6f]
GC_00001678	W2bin3	contig_715	polyketide cyclase [Candidatus Nitrosotenuis sp. AQ6f]
GC_00001680	S1bin1	contig_172	FMN phosphatase YigB, HAD superfamily
GC_00001680	S2bin1	contig_98	FMN phosphatase YigB, HAD superfamily
GC_00001680	S3bin1	contig_76	FMN phosphatase YigB, HAD superfamily
GC_00001680	W1bin1	scaffold_86398	FMN phosphatase YigB, HAD superfamily
GC_00001680	W2bin3	contig_1164	FMN phosphatase YigB, HAD superfamily
GC_00001685	S1bin1	contig_1	Sugar-specific transcriptional regulator TrmB
GC_00001685	S2bin1	contig_31	Sugar-specific transcriptional regulator TrmB
GC_00001685	S3bin1	contig_28	Sugar-specific transcriptional regulator TrmB
GC_00001685	W1bin1	scaffold_8857	Sugar-specific transcriptional regulator TrmB

GC_00001685	W2bin3	contig_43	Sugar-specific transcriptional regulator TrmB
GC_00001686	S1bin1	contig_13	2-hydroxy-3-oxopropionate reductase (mmsB) [uncultured marine thaumarchaeote KM3_136_D12]
GC_00001686	S2bin1	contig_3	2-hydroxy-3-oxopropionate reductase (mmsB) [uncultured marine thaumarchaeote KM3_136_D12]
GC_00001686	S3bin1	contig_5	2-hydroxy-3-oxopropionate reductase (mmsB) [uncultured marine thaumarchaeote KM3_136_D12]
GC_00001686	W1bin1	scaffold_878	2-hydroxy-3-oxopropionate reductase (mmsB) [uncultured marine thaumarchaeote KM3_136_D12]
GC_00001686	W2bin3	contig_440	2-hydroxy-3-oxopropionate reductase (mmsB) [uncultured marine thaumarchaeote KM3_136_D12]
GC_00001687	S1bin1	contig_187	zinc ribbon domain-containing protein [Clostridium neonatale]
GC_00001687	S2bin1	contig_14	zinc ribbon domain-containing protein [Clostridium neonatale]
GC_00001687	S3bin1	contig_194	zinc ribbon domain-containing protein [Clostridium neonatale]
GC_00001687	W1bin1	scaffold_4666	zinc ribbon domain-containing protein [Clostridium neonatale]
GC_00001687	W2bin3	contig_334	zinc ribbon domain-containing protein [Clostridium neonatale]
GC_00001689	S1bin1	contig_1	Signal-transduction protein containing cAMP-binding, CBS, and nucleotidyltransferase domains
GC_00001689	S2bin1	contig_27	Signal-transduction protein containing cAMP-binding, CBS, and nucleotidyltransferase domains
GC_00001689	S3bin1	contig_23	Signal-transduction protein containing cAMP-binding, CBS, and nucleotidyltransferase domains
GC_00001689	W1bin1	scaffold_12156	Signal-transduction protein containing cAMP-binding, CBS, and nucleotidyltransferase domains
GC_00001689	W2bin3	contig_223	Signal-transduction protein containing cAMP-binding, CBS, and nucleotidyltransferase domains
GC_00001692	S1bin1	contig_117	Uncharacterized conserved protein YbbK, DUF523 family|Uncharacterized conserved protein YbgA, DUF1722 family

GC_00001692	S2bin1	contig_31	Uncharacterized conserved protein YbbK, DUF523 family|Uncharacterized conserved protein YbgA, DUF1722 family
GC_00001692	S3bin1	contig_28	Uncharacterized conserved protein YbbK, DUF523 family|Uncharacterized conserved protein YbgA, DUF1722 family
GC_00001692	W1bin1	scaffold_24263	Uncharacterized conserved protein YbbK, DUF523 family|Uncharacterized conserved protein YbgA, DUF1722 family
GC_00001692	W2bin3	contig_43	Uncharacterized conserved protein YbbK, DUF523 family|Uncharacterized conserved protein YbgA, DUF1722 family
GC_00001714	S1bin1	contig_3_partial_0_26617	Archaellum biogenesis protein FlaJ, TadC family
GC_00001714	S2bin1	contig_59	Archaellum biogenesis protein FlaJ, TadC family
GC_00001714	W1bin1	scaffold_13376	Archaellum biogenesis protein FlaJ, TadC family
GC_00001714	W1bin1	scaffold_147792	Archaellum biogenesis protein FlaJ, TadC family
GC_00001714	W2bin3	contig_723	Archaellum biogenesis protein FlaJ, TadC family
GC_00001716	S1bin1	contig_13	Rad51 [Candidatus Nitrososphaera evergladensis SR1]
GC_00001716	S2bin1	contig_3	Rad51 [Candidatus Nitrososphaera evergladensis SR1]
GC_00001716	W1bin1	scaffold_878	Rad51 [Candidatus Nitrososphaera evergladensis SR1]
GC_00001716	W2bin3	contig_280	Rad51 [Candidatus Nitrososphaera evergladensis SR1]
GC_00001716	S3bin1	contig_264	Rad51 [Candidatus Nitrososphaera evergladensis SR1]
GC_00001718	S1bin1	contig_1	K+-sensing histidine kinase KdpD
GC_00001718	S2bin1	contig_6	K+-sensing histidine kinase KdpD
GC_00001718	S3bin1	contig_38	K+-sensing histidine kinase KdpD
GC_00001718	W1bin1	scaffold_12156	K+-sensing histidine kinase KdpD
GC_00001718	W2bin3	contig_175	K+-sensing histidine kinase KdpD
GC_00001719	S1bin1	contig_204	DNA repair exonuclease SbcCD ATPase subunit
GC_00001719	S2bin1	contig_56	DNA repair exonuclease SbcCD ATPase subunit
GC_00001719	S3bin1	contig_193	DNA repair exonuclease SbcCD ATPase subunit
GC_00001719	W1bin1	scaffold_21308	DNA repair exonuclease SbcCD ATPase subunit
GC_00001719	W2bin3	contig_449	DNA repair exonuclease SbcCD ATPase subunit
GC_00001722	S1bin1	contig_40	Small nuclear ribonucleoprotein (snRNP) homolog
GC_00001722	S1bin1	contig_130	Small nuclear ribonucleoprotein (snRNP) homolog
GC_00001722	S2bin1	contig_96	Small nuclear ribonucleoprotein (snRNP) homolog

GC_00001722	S3bin1	contig_276	Small nuclear ribonucleoprotein (snRNP) homolog
GC_00001722	W1bin1	scaffold_127581	Small nuclear ribonucleoprotein (snRNP) homolog
GC_00001727	S1bin1	contig_18	Uncharacterized conserved protein YbgA, DUF1722 family
GC_00001727	S2bin1	contig_18	Uncharacterized conserved protein YbgA, DUF1722 family
GC_00001727	S3bin1	contig_1	Uncharacterized conserved protein YbgA, DUF1722 family
GC_00001727	W1bin1	scaffold_3608	Uncharacterized conserved protein YbgA, DUF1722 family
GC_00001727	W2bin3	contig_185	Uncharacterized conserved protein YbgA, DUF1722 family
GC_00001773	S1bin1	contig_40	Ubiquinone/menaquinone biosynthesis C-methylase UbiE
GC_00001773	S1bin1	contig_130	Ubiquinone/menaquinone biosynthesis C-methylase UbiE
GC_00001773	S2bin1	contig_96	Ubiquinone/menaquinone biosynthesis C-methylase UbiE
GC_00001773	S3bin1	contig_276	Ubiquinone/menaquinone biosynthesis C-methylase UbiE
GC_00001773	W1bin1	scaffold_127581	Ubiquinone/menaquinone biosynthesis C-methylase UbiE
GC_00001820	S1bin1	contig_1	histidine kinase [Marine Group I thaumarchaeote SCGC AAA799-P11]
GC_00001820	S2bin1	contig_31	histidine kinase [Marine Group I thaumarchaeote SCGC AAA799-P11]
GC_00001820	S3bin1	contig_28	histidine kinase [Marine Group I thaumarchaeote SCGC AAA799-P11]
GC_00001820	W2bin3	contig_43	histidine kinase [Marine Group I thaumarchaeote SCGC AAA799-P11]
GC_00001821	S1bin1	contig_18	Multimeric flavodoxin WrbA
GC_00001821	S2bin1	contig_18	Multimeric flavodoxin WrbA
GC_00001821	S3bin1	contig_1	Multimeric flavodoxin WrbA
GC_00001821	W1bin1	scaffold_7106	Multimeric flavodoxin WrbA
GC_00001823	S1bin1	contig_136	Ubiquinone/menaquinone biosynthesis C-methylase UbiE
GC_00001823	S2bin1	contig_61	Ubiquinone/menaquinone biosynthesis C-methylase UbiE
GC_00001823	W1bin1	scaffold_38193	Ubiquinone/menaquinone biosynthesis C-methylase UbiE
GC_00001823	W2bin3	contig_316	Ubiquinone/menaquinone biosynthesis C-methylase UbiE

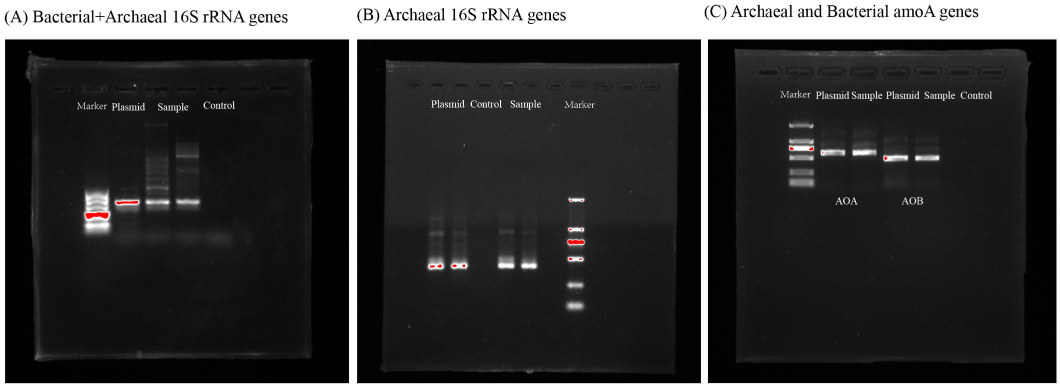
GC_00001849	W1bin1	scaffold_9271	Chemotaxis response regulator CheB, contains REC and protein-glutamate methylesterase domains
GC_00001849	S1bin1	contig_21	DNA-binding response regulator, NarL/FixJ family, contains REC and HTH domains
GC_00001849	S2bin1	contig_86	DNA-binding response regulator, NarL/FixJ family, contains REC and HTH domains
GC_00001849	S3bin1	contig_3	DNA-binding response regulator, NarL/FixJ family, contains REC and HTH domains
GC_00001863	S1bin1	contig_78	Phosphoglycolate phosphatase, HAD superfamily
GC_00001863	S2bin1	contig_34	Phosphoglycolate phosphatase, HAD superfamily
GC_00001863	W1bin1	scaffold_7654	Phosphoglycolate phosphatase, HAD superfamily
GC_00001863	W2bin3	contig_731	Phosphoglycolate phosphatase, HAD superfamily
GC_00001867	S1bin1	contig_35	CheY chemotaxis protein or a CheY-like REC (receiver) domain|CheY chemotaxis protein or a CheY-like REC (receiver) domain|HPt (histidine-containing phosphotransfer) domain|PAS domain|K+-sensing histidine kinase KdpD
GC_00001867	S2bin1	contig_19	CheY chemotaxis protein or a CheY-like REC (receiver) domain|CheY chemotaxis protein or a CheY-like REC (receiver) domain|HPt (histidine-containing phosphotransfer) domain|PAS domain|K+-sensing histidine kinase KdpD
GC_00001867	W1bin1	scaffold_2754	CheY chemotaxis protein or a CheY-like REC (receiver) domain|CheY chemotaxis protein or a CheY-like REC (receiver) domain|HPt (histidine-containing phosphotransfer) domain|PAS domain|K+-sensing histidine kinase KdpD
GC_00001867	W2bin3	contig_408	CheY chemotaxis protein or a CheY-like REC (receiver) domain|CheY chemotaxis protein or a CheY-like REC (receiver) domain|HPt (histidine-containing phosphotransfer) domain|PAS domain|K+-sensing histidine kinase KdpD
GC_00001874	S1bin1	contig_255	DNA-binding beta-propeller fold protein YncE|DNA-binding beta-propeller fold protein YncE
GC_00001874	S2bin1	contig_110	DNA-binding beta-propeller fold protein YncE|DNA-binding beta-propeller fold protein YncE
GC_00001874	S3bin1	contig_93	DNA-binding beta-propeller fold protein

GC_00001874	W1bin1	scaffold_2754	YncE|DNA-binding beta-propeller fold protein YncE DNA-binding beta-propeller fold protein YncE|DNA-binding beta-propeller fold protein YncE
GC_00001880	S1bin1	contig_3_partial_0_26617	Sugar-specific transcriptional regulator TrmB
GC_00001880	S2bin1	contig_59	Sugar-specific transcriptional regulator TrmB
GC_00001880	W1bin1	scaffold_13376	Sugar-specific transcriptional regulator TrmB
GC_00001880	W2bin3	contig_978	Sugar-specific transcriptional regulator TrmB
GC_00001881	S1bin1	contig_38	Uncharacterized conserved protein YeaO, DUF488 family
GC_00001881	S3bin1	contig_21	Uncharacterized conserved protein YeaO, DUF488 family
GC_00001881	W1bin1	scaffold_29150	Uncharacterized conserved protein YeaO, DUF488 family
GC_00001881	W2bin3	contig_128	Uncharacterized conserved protein YeaO, DUF488 family
GC_00001900	S1bin1	contig_18	UV DNA damage repair endonuclease
GC_00001900	S2bin1	contig_18	UV DNA damage repair endonuclease
GC_00001900	S3bin1	contig_1	UV DNA damage repair endonuclease
GC_00001900	W1bin1	scaffold_7106	UV DNA damage repair endonuclease
GC_00001902	S1bin1	contig_26	Peptidoglycan/xylan/chitin deacetylase, PgdA/CDA1 family
GC_00001902	S2bin1	contig_78	Peptidoglycan/xylan/chitin deacetylase, PgdA/CDA1 family
GC_00001902	W1bin1	scaffold_84949	Peptidoglycan/xylan/chitin deacetylase, PgdA/CDA1 family
GC_00001902	W2bin3	contig_177	Peptidoglycan/xylan/chitin deacetylase, PgdA/CDA1 family
GC_00002005	S1bin1	contig_79	Endonuclease YncB, thermonuclease family
GC_00002005	S3bin1	contig_217	Endonuclease YncB, thermonuclease family
GC_00002005	W1bin1	scaffold_54873	Endonuclease YncB, thermonuclease family
GC_00002009	S1bin1	contig_113	K+-sensing histidine kinase KdpD
GC_00002009	S2bin1	contig_44	K+-sensing histidine kinase KdpD
GC_00002009	W1bin1	scaffold_60906	K+-sensing histidine kinase KdpD
GC_00002014	S2bin1	contig_53	Predicted Zn-dependent protease
GC_00002014	W1bin1	scaffold_26013	Predicted Zn-dependent protease
GC_00002014	W2bin3	contig_107	Predicted Zn-dependent protease
GC_00002018	S2bin1	contig_79	PAS domain|PAS domain|Signal transduction histidine kinase regulating C4-dicarboxylate transport system|Signal transduction histidine kinase involved in nitrogen fixation and metabolism regulation

GC_00002018	W1bin1	scaffold_3608	PAS domain|PAS domain|Signal transduction histidine kinase regulating C4-dicarboxylate transport system|Signal transduction histidine kinase involved in nitrogen fixation and metabolism regulation
GC_00002018	W2bin3	contig_829	PAS domain|PAS domain|Signal transduction histidine kinase regulating C4-dicarboxylate transport system|Signal transduction histidine kinase involved in nitrogen fixation and metabolism regulation
GC_00002027	S1bin1	contig_51	Small nuclear ribonucleoprotein (snRNP) homolog
GC_00002027	S2bin1	contig_32	Small nuclear ribonucleoprotein (snRNP) homolog
GC_00002027	W2bin3	contig_814	Small nuclear ribonucleoprotein (snRNP) homolog
GC_00002032	S1bin1	contig_40	Small-conductance mechanosensitive channel
GC_00002032	S2bin1	contig_61	Small-conductance mechanosensitive channel
GC_00002032	W1bin1	scaffold_38193	Small-conductance mechanosensitive channel
GC_00002048	S2bin1	contig_199	signal recognition particle-docking protein FtsY [Candidatus Nitrosoarchaeum sp.]
GC_00002048	W1bin1	scaffold_24263	signal recognition particle-docking protein FtsY [Candidatus Nitrosoarchaeum sp.]
GC_00002048	S1bin1	contig_494	signal recognition particle-docking protein FtsY [Candidatus Nitrosopumilus adriaticus]
GC_00002083	S2bin1	contig_111	beta-lactamase domain protein [Nitrosopumilus maritimus SCM1]
GC_00002083	S1bin1	contig_211	Glyoxylase or a related metal-dependent hydrolase, beta-lactamase superfamily II
GC_00002083	W2bin3	contig_213	Glyoxylase or a related metal-dependent hydrolase, beta-lactamase superfamily II
GC_00002088	S1bin1	contig_30	Predicted dehydrogenase
GC_00002088	S2bin1	contig_67	Predicted dehydrogenase
GC_00002088	W1bin1	scaffold_8859	Predicted dehydrogenase
GC_00002093	S2bin1	contig_44	Secreted protein containing bacterial Ig-like domain and vWFA domain
GC_00002093	W1bin1	scaffold_60906	Secreted protein containing bacterial Ig-like domain and vWFA domain
GC_00002093	W2bin3	contig_970	Secreted protein containing bacterial Ig-like domain and vWFA domain
GC_00002102	S2bin1	contig_53	Arylsulfatase A or related enzyme
GC_00002102	W1bin1	scaffold_13204	Arylsulfatase A or related enzyme
GC_00002102	W1bin1	scaffold_26013	Arylsulfatase A or related enzyme
GC_00002143	S1bin1	contig_207	Uncharacterized conserved protein YfaS,

			alpha-2-macroglobulin family
GC_00002143	S2bin1	contig_91	Uncharacterized conserved protein YfaS, alpha-2-macroglobulin family
GC_00002143	S3bin1	contig_148	Uncharacterized conserved protein YfaS, alpha-2-macroglobulin family
GC_00002164	W2bin3	contig_21	hypothetical protein NMY3_00230 [Candidatus Nitrocosmicus oleophilus]
GC_00002164	S2bin1	contig_73	MEDS: MEthanogen/methylotroph, DcmR Sensory domain [Candidatus Nitrososphaera evergladensis SR1]
GC_00002164	W1bin1	scaffold_7346	MEDS: MEthanogen/methylotroph, DcmR Sensory domain [Candidatus Nitrososphaera evergladensis SR1]
GC_00002448	S2bin1	contig_105	carboxypeptidase regulatory-like domain-containing protein [Candidatus Nitrosoarchaeum koreensis]
GC_00002448	W1bin1	scaffold_3608	carboxypeptidase regulatory-like domain-containing protein [Candidatus Nitrosoarchaeum koreensis]
GC_00002610	S2bin1	contig_53	Phosphoserine phosphatase
GC_00002610	W1bin1	scaffold_26013	Phosphoserine phosphatase
GC_00002654	S1bin1	contig_219	Methyl-accepting chemotaxis protein
GC_00002654	S2bin1	contig_137	Methyl-accepting chemotaxis protein

Supplementary Figures

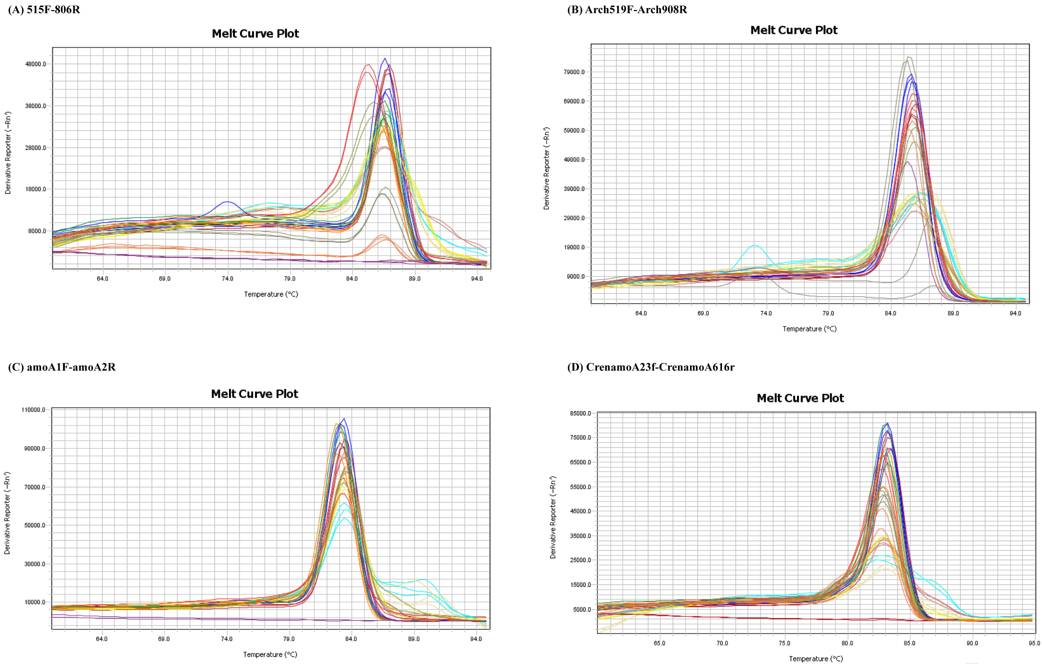


(A) Bacterial+Archaeal 16S rRNA genes

(B) Archaeal 16S rRNA genes

(C) Archaeal and Bacterial amoA genes

Figure S1 Gel electrophoretograms for all primer pairs used in the qPCR experiments, including Uni515F-Uni806R (A), Arch519F-Arch808R (B), amoA1F-amoA2R (AOB) and crenamoA23f-crenamoA616r (AOA) (C).



(A) 515F-806R

(B) Arch519F-Arch908R

(C) amoA1F-amoA2R

(D) CrenamoA23f-CrenamoA616r

Figure S2 Melt curve plots for all qPCR experiments conducted in this study, including total 16S rRNA genes (A), archaeal 16S rRNA genes (B), bacterial *amoA* genes (C) and archaeal *amoA* genes (D).

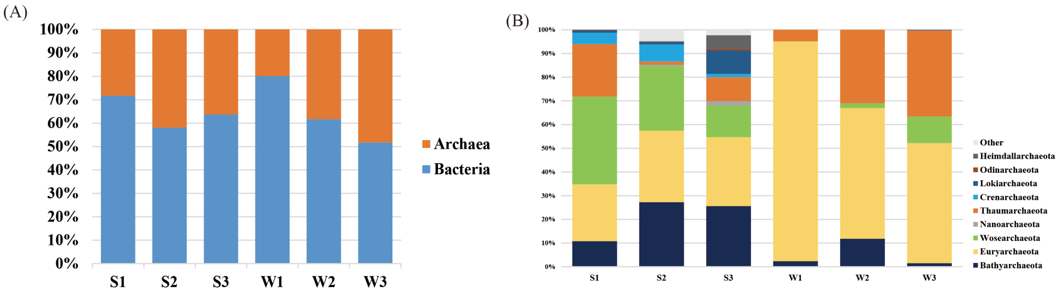


Figure S3 Prokaryotes community composition (A) and detailed archaeal community composition (B) of samples based on metagenomics. All bacterial and archaeal 16S rRNA genes were extracted from assembled data, and then normalized to calculate the relative abundance. Referring the method part for details.

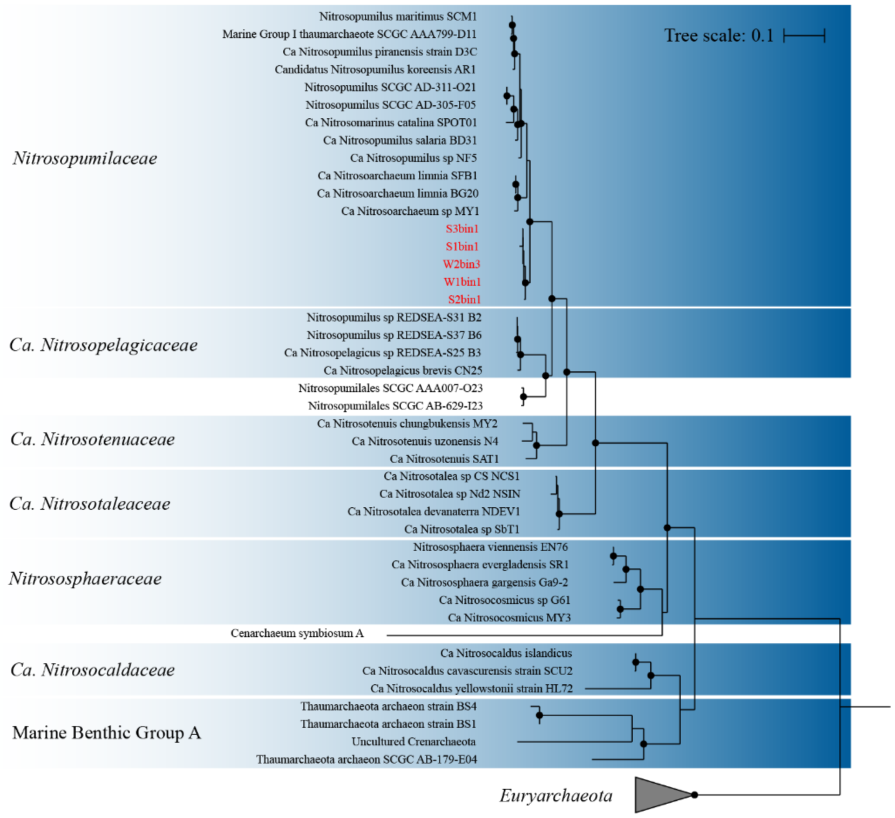


Figure S4 The phylogenetic tree of concatenated 16S-23S rRNA genes.

Euryarchaeotal genes were set as the outer group, using GTR+F+R6 model with 8489 aliened base sequences. The new AOA species in this study was in red. Bootstraps values > 90 are illustrated by solid circles.

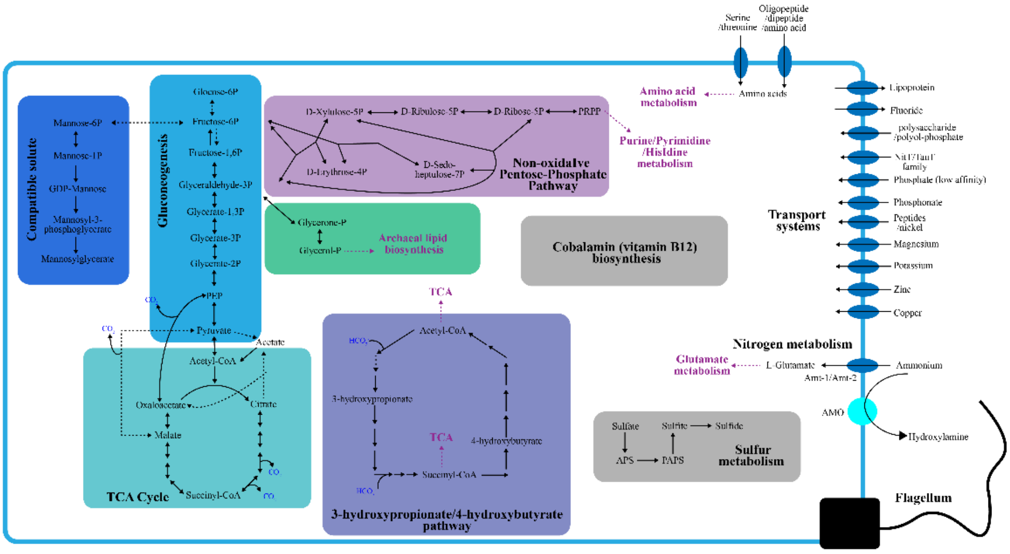


Figure S5 Major metabolic pathways in the new AOA species in this study. Dashed lines indicate the gene was absent.

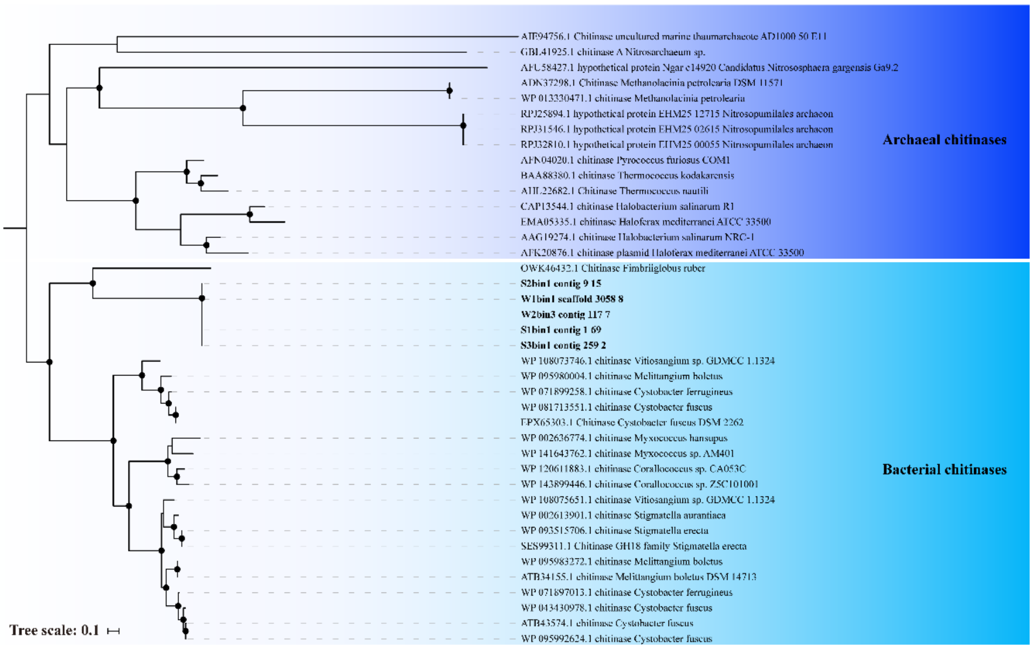


Figure S6 The phylogenetic tree of chitinases extracted from MAGs in this study (in black bold) and other refences sequences (WAG+F+G4 as best model with 2055 aligned amino acid residues). Bootstraps values > 90 are illustrated by solid circles.

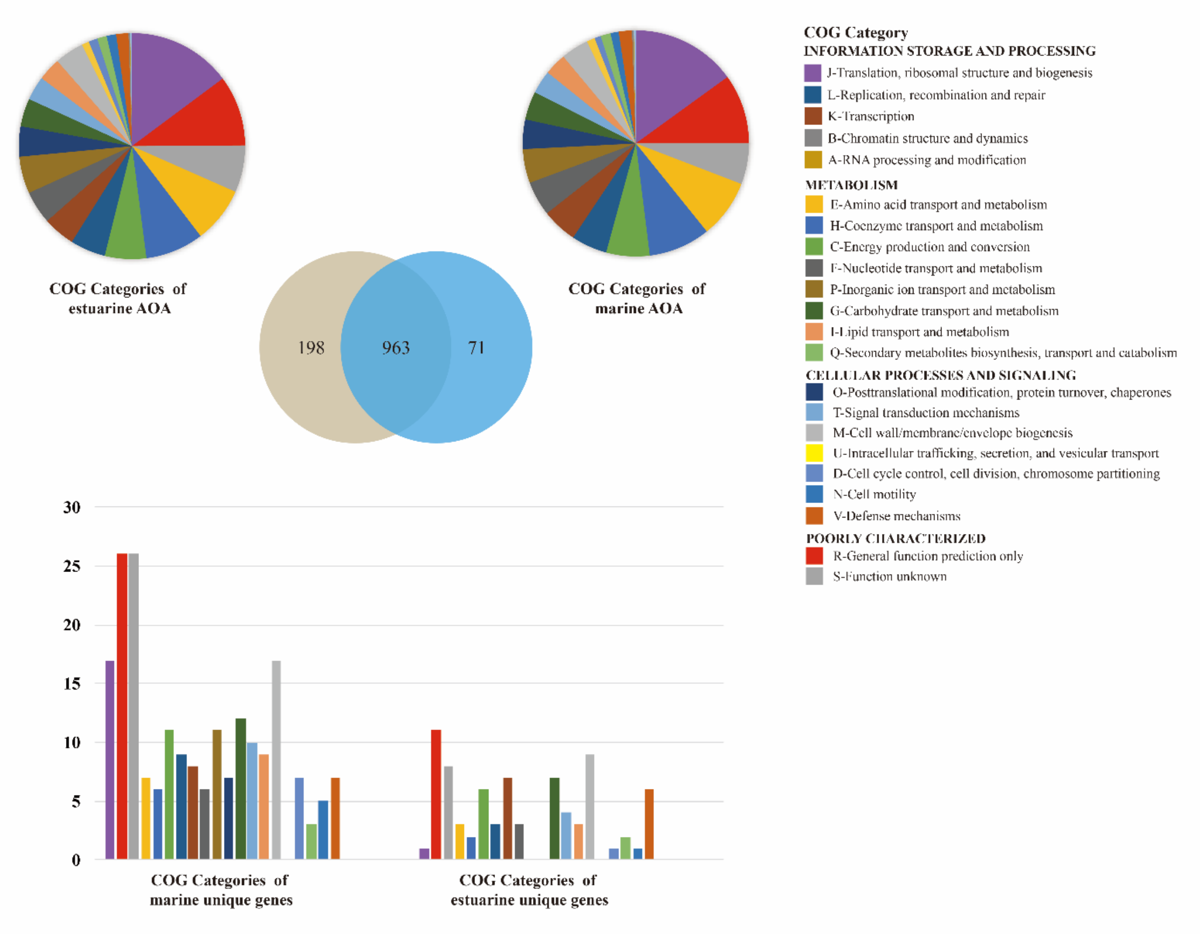


Figure S7 The comparison about COG function categories between estuarine and marine AOA (detailed definition see method part). Pie charts indicate the function composition, veen chart indicates unique and common gene categories, and bar charts indicate the detailed number of unique genes.

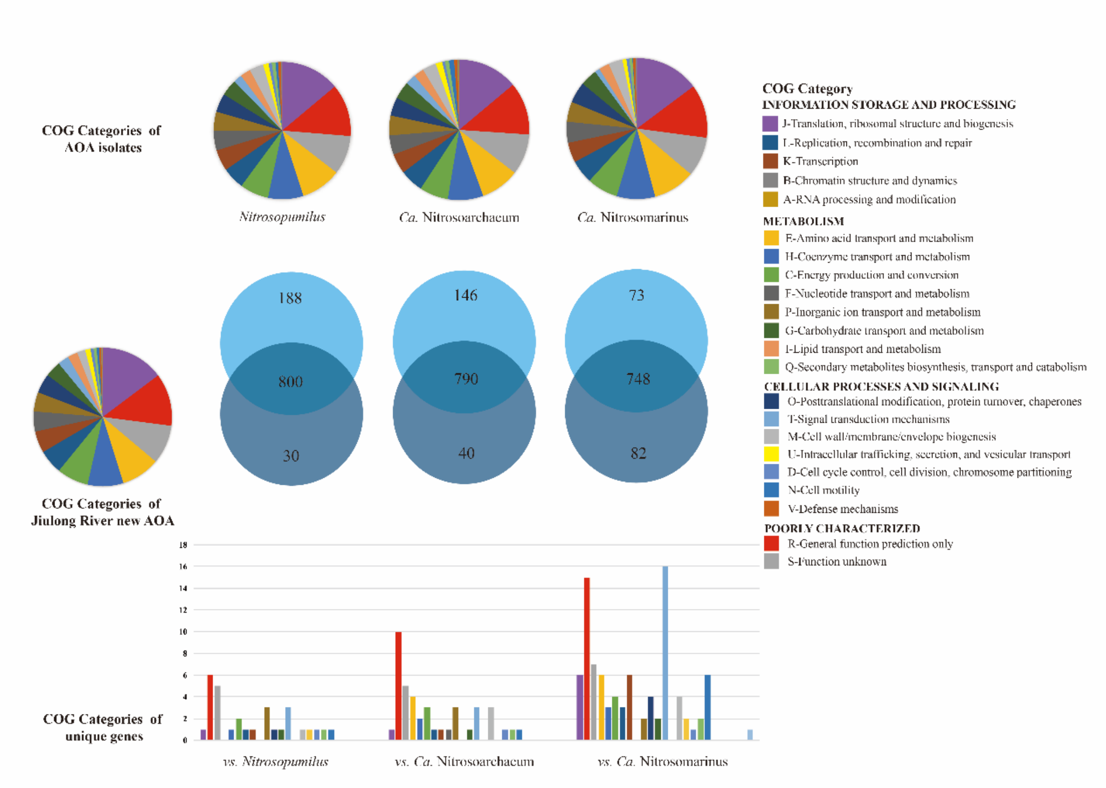


Figure S8 COG function categories comparison between the new AOA species in this study and other AOA species in family *Nitrosopumilaceae*. Pie charts indicate the function composition, veen charts indicate unique and common gene categories, and bar charts indicate the detailed number of unique genes.

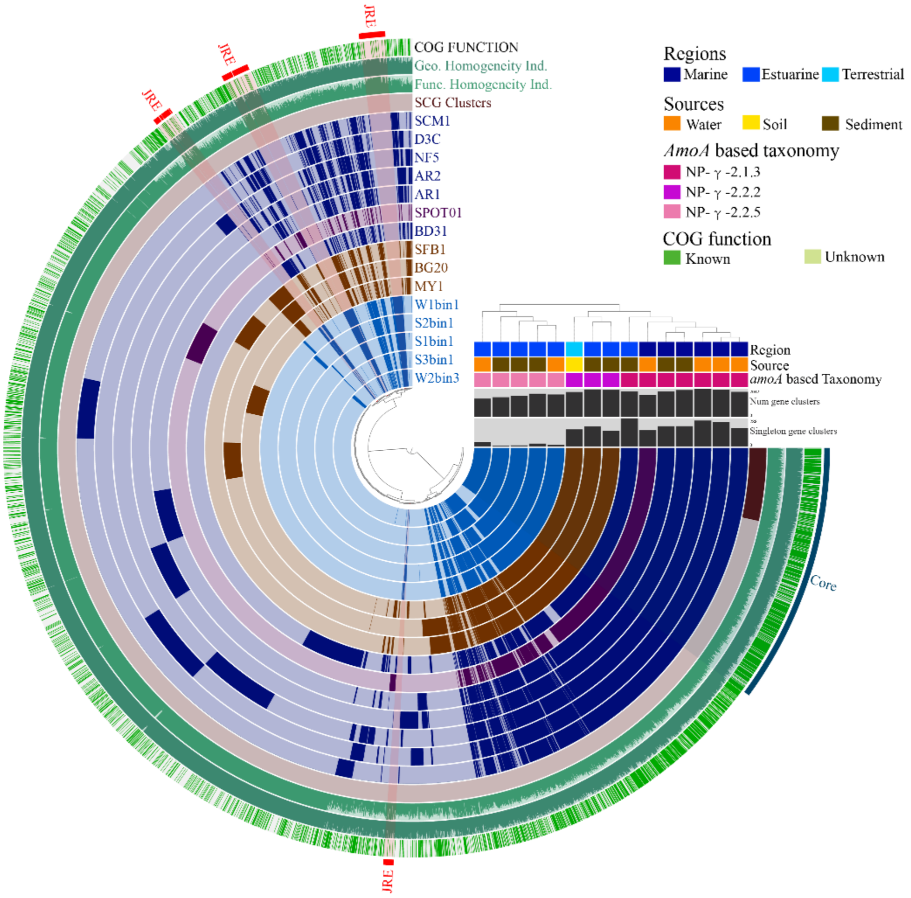


Figure S9 The pan-genomic analysis of AOA MAGs in this study (in light blue) and other AOA strains in the family *Nitrosopumilaceae*. The red bars (JRE) stands for genes only encoded in this study, and the dark green bar (Core) stands for genes shared by all genomes in this analysis. The isolation regions and sources, as well as the *amoA* genotypes, are presented by color of squares.