Supplementary material

Table S1 Comparison of different clustering algorithms using publicly available AOA *amoA* gene sequences

Threshold	Methods and OTU numbers						
	Furthest	Nearest	Average	Weight	Optimistic MCC	Vsearch AGC	Vsearch DGC
97%	3188	1933	2789	2843	2755	2857	2857
90%	619	267	392	469	418	445	445
85%	253	177	114	165	132	146	146

Table S2 Mean inter-cluster distance of the reconstructed phylogenetic tree

Clusters	*Nitrosopumilus* cluster	*Nitrosotalea* cluster	*Nitrososphaera* cluster	Nitrosocaldus cluster
Nitrosopumilus cluster				
Nitrosotalea cluster	23.9			
Nitrososphaera cluster	28.5	29.5		
Nitrosocaldus cluster	30.3	30.1	27.3	
Bacterial *amoA* (out group)	65.5	64	64.5	65.2

Table S3 Mean in-cluster distance of the *Nitrososphaera* cluster

Clusters	*Nitrososphaera* subcluster 1	*Nitrososphaera* subcluster 2	*Nitrososphaera* subcluster 3
Nitrososphaera subcluster 1			
Nitrososphaera subcluster 2	22.9		
Nitrososphaera subcluster 3	21.7	21.6	
Nitrososphaera subcluster 4	23.2	23.6	21.1

Table S4 Mean in-cluster distance of the *Nitrosopumilus* cluster

Clusters	*Nitrosopumilus* cubcluster 1	*Nitrosopumilus* cubcluster 2
Nitrosopumilus cubcluster 1		
Nitrosopumilus cubcluster 2	0.184452	
Nitrosopumilus cubcluster 3	0.225306	0.205400

Table S5 Sequence quantity comparison between raw sequencing data and high-quality sequences

Forward	Raw data	High quality		Reverse	Raw data	High quality	
	Sequence number	Sequence number	Percent		Sequence number	Sequence number	Percent
Hca_1F	22007	18503	84.08%	Hca_1R	21677	18149	83.72%
Hca_2F	17063	14941	87.56%	Hca_2R	17238	15422	89.47%
Hca_3F	8750	2985	34.11%	Hca_3R	9119	3056	33.51%
Hde_1F	26446	24310	91.92%	Hde_1R	26319	24905	94.63%
Hde_2F	15707	14657	93.32%	Hde_2R	15829	14852	93.83%
Hde_3F	18585	16955	91.23%	Hde_3R	18226	17058	93.59%
Qde_1F	12625	11320	89.66%	Qde_1R	11543	10118	87.65%
Qde_2F	25206	23260	92.28%	Qde_2R	24209	21913	90.52%
Qde_3F	16670	15389	92.32%	Qde_3R	16313	15070	92.38%

Table S6 The α-diversity indices of the *amoA* gene calculated at 85% similarity

Sequences	Sampling Sets	Chao1	Shannon	Simpson	Sobs
Forward	Hca_1F	1.15	0.00044	1.00	1.15
	Hca_2F	4.65	0.06	0.98	3.95
	Hca_3F	3.00	0.07	0.98	3.00
	Hde_1F	2.81	0.69	0.50	2.81
	Hde_2F	3.24	0.44	0.73	3.11
	Hde_3F	2.85	0.08	0.97	2.67
	Qde_1F	7.27	0.92	0.48	6.60
	Qde_2F	6.04	0.74	0.60	5.86
	Qde_3F	5.42	0.77	0.58	5.36
Reverse	Hca_1R	3.80	0.09	0.97	3.64
	Hca_2R	8.34	0.14	0.95	6.65
	Hca_3R	4.00	0.09	0.97	4.00
	Hde_1R	7.80	0.84	0.47	6.96
	Hde_2R	7.68	0.52	0.71	7.12
	Hde_3R	4.26	0.10	0.96	3.98
	Qde_1R	9.10	0.89	0.48	8.16
	Qde_2R	7.67	0.64	0.66	7.25
	Qde_3R	7.74	0.61	0.68	7.30

Table S7 Community composition of each sample based on 91% similarity of the reverse sequences

AOA Reverse Groups	Hca_1R	Hca_2R	Hca_3R	Hde_1R	Hde_2R	Hde_3R	Qde_1R	Qde_2R	Qde_3R
Nitrosopumilus subcluster 2.1	0	0	0	0	0	0	0.61000198	0.17812158	0.1530856
Sum of *Nitrosopumilus* cluster	**0**	**0**	**0**	**0**	**0**	**0**	**0.61000198**	**0.17812158**	**0.1530856**
Nitrososphaera subcluster 1.1	0.99994489	0.98981974	0.98789267	0.53932945	0.84399408	0.98493376	0.00148251	0.00730194	0.00623756
Nitrososphaera subcluster 1.2	0	0.00019453	0	0	0	0	0	0	0
Nitrososphaera subcluster 1.3	0	0.000195	0.00327225	0	0	5.8624E-05	0.18817948	0.762	0.752
Nitrososphaera subcluster 1.4	0	0.00038906	0	0	0.00013466	5.8624E-05	0.199	0.04654984	0.08639682
Nitrososphaera subcluster 2.1	-	-	-	-	-	-	-	-	-
Nitrososphaera subcluster 2.2	0	6.4842E-05	0	0	0	0	0	0	0
Nitrososphaera subcluster 3.1	0	0	0	0	0	0	0.0008895	0.00566	0.00179
Sum of *Nitrososphaera* cluster	**0.99994489**	**0.99066316**	**0.99116492**	**0.53932945**	**0.84412874**	**0.985051**	**0.38955149**	**0.82151177**	**0.84642437**
Nitrosotalea subcluster 1.1	5.5106E-05	0.00155622	0.00196335	0.00044168	0.00127929	5.8624E-05	0	0	0
Nitrosotalea subcluster 1.2	0	0.00778109	0.00687173	0.46022887	0.15459197	0.01489037	0	0	0
Sum of *Nitrosotalea* cluster	**5.5106E-05**	**0.00933731**	**0.00883508**	**0.46067055**	**0.15587126**	**0.014949**	**0**	**0**	**0**

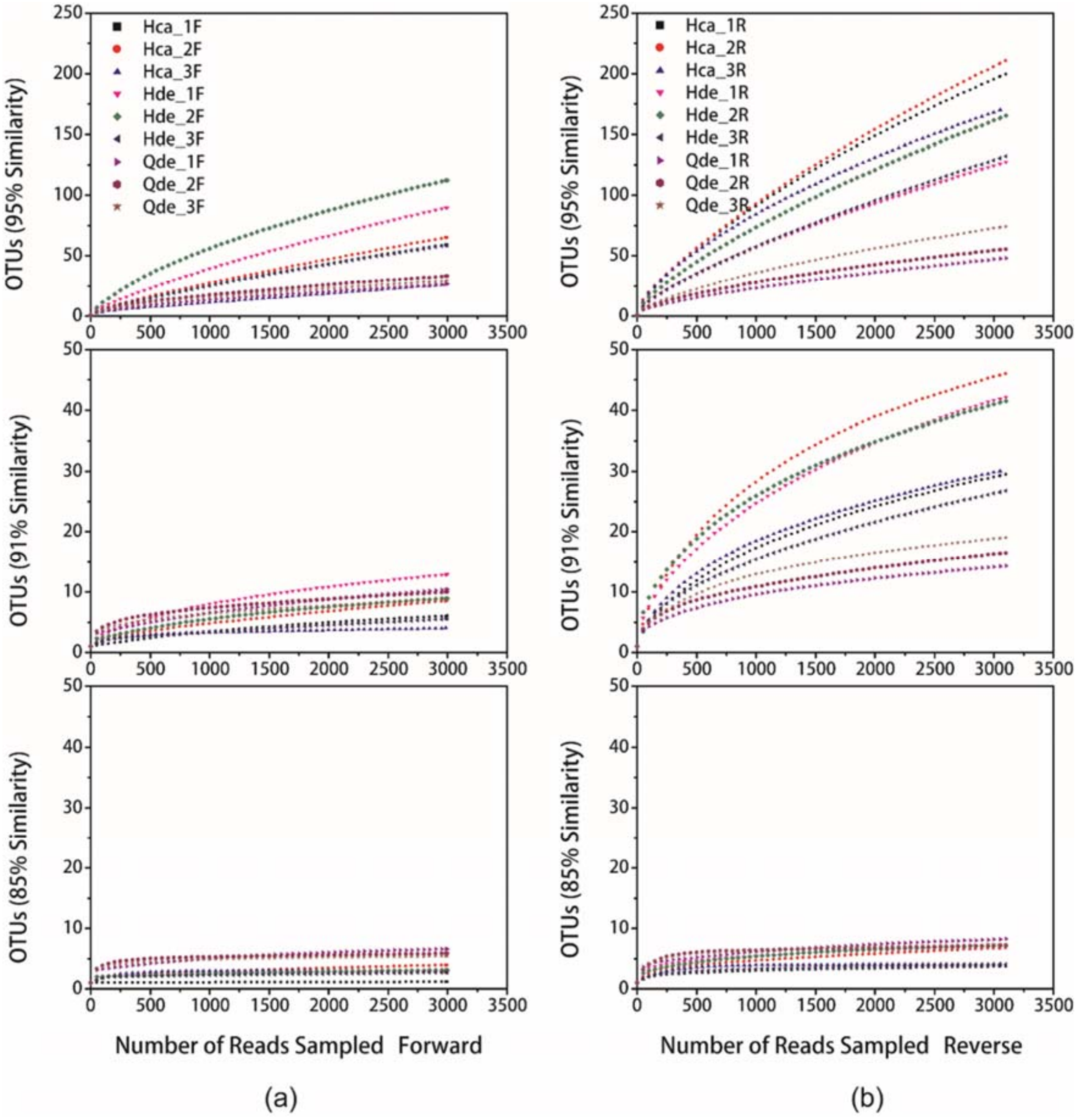

Fig. S1 Rarefraction curve of the *amoA* gene at 95%, 91%, and 85% similarities. (a) Forward sequences. (b) Reverse sequences.

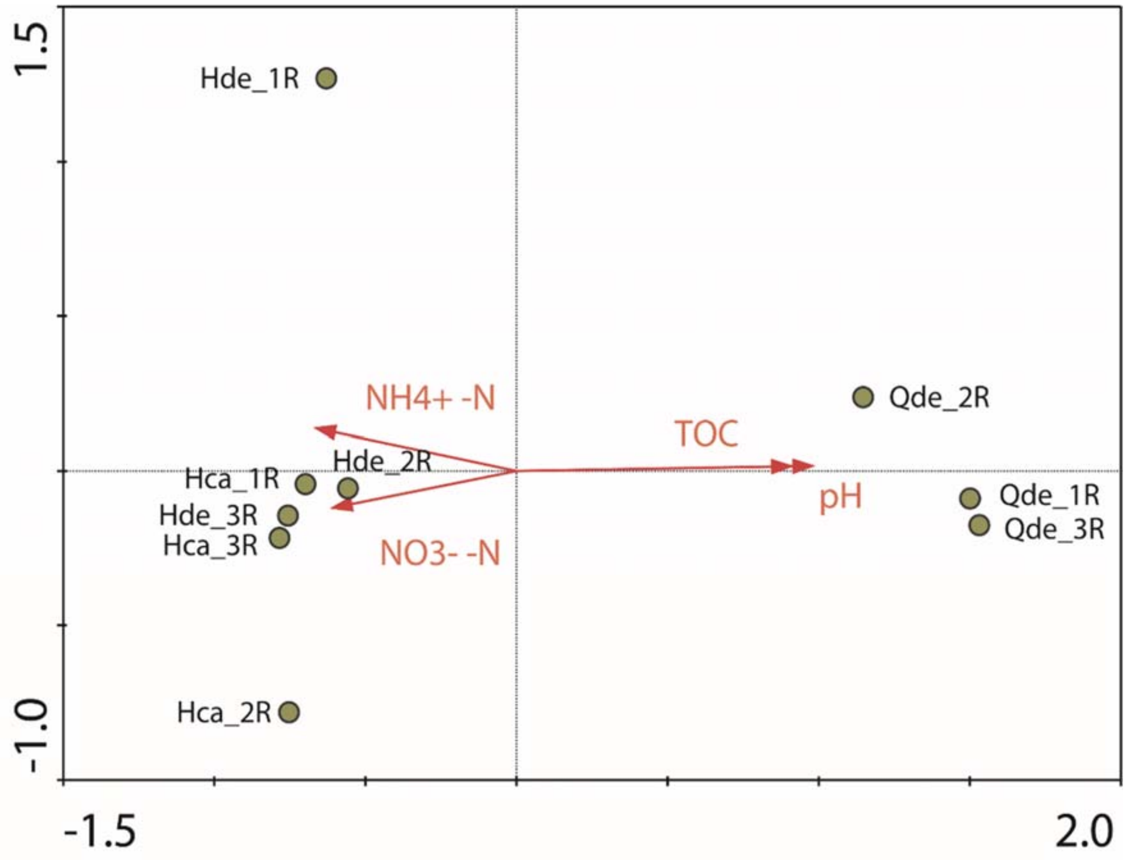

Fig. S2 CCA of the AOA communities in marsh wetlands and environmental factors (analyzed with reverse sequences).