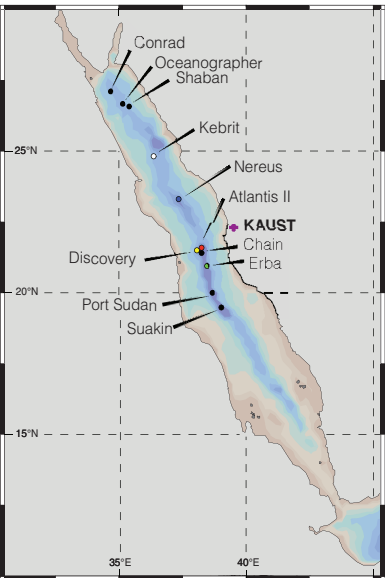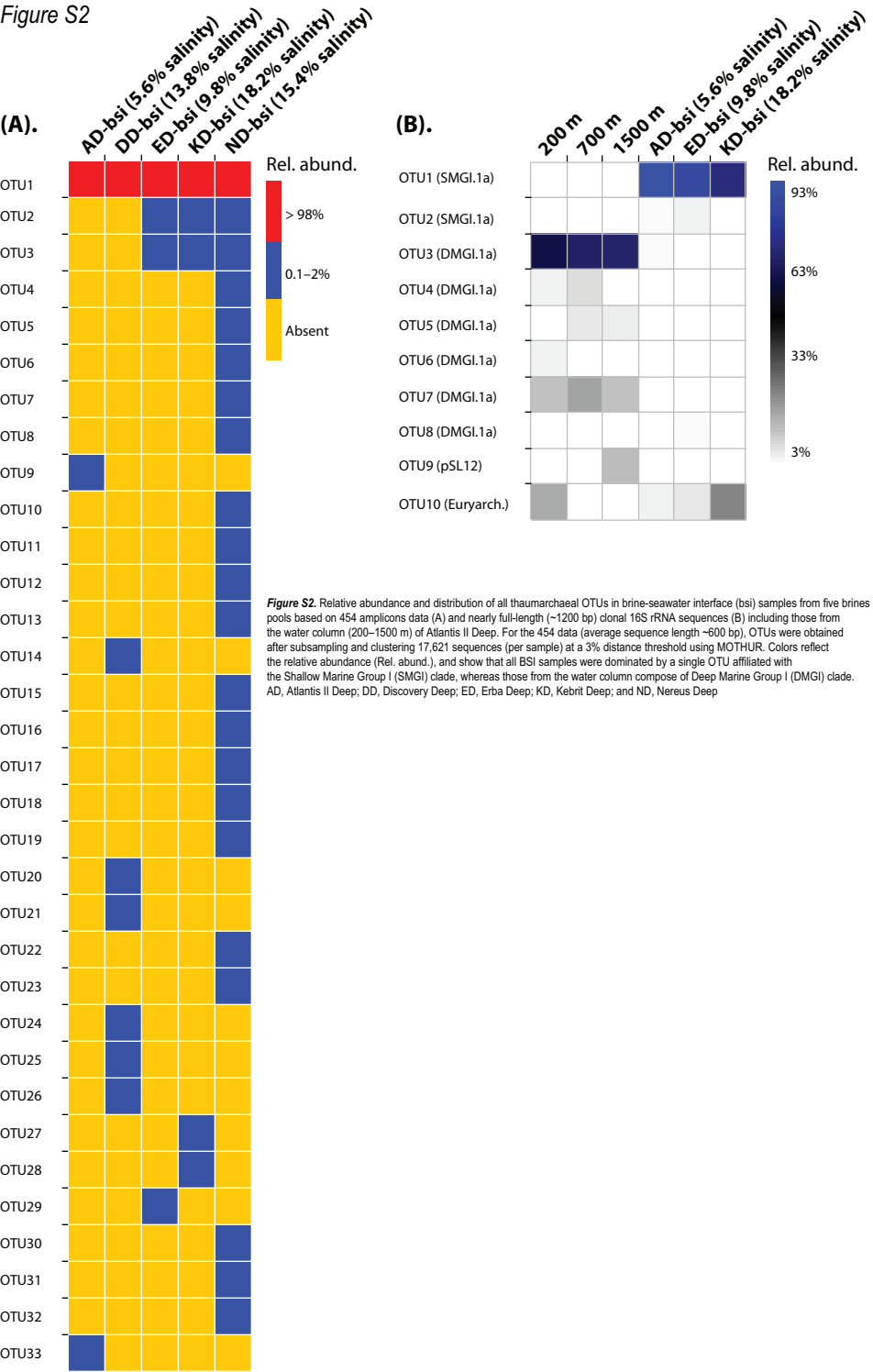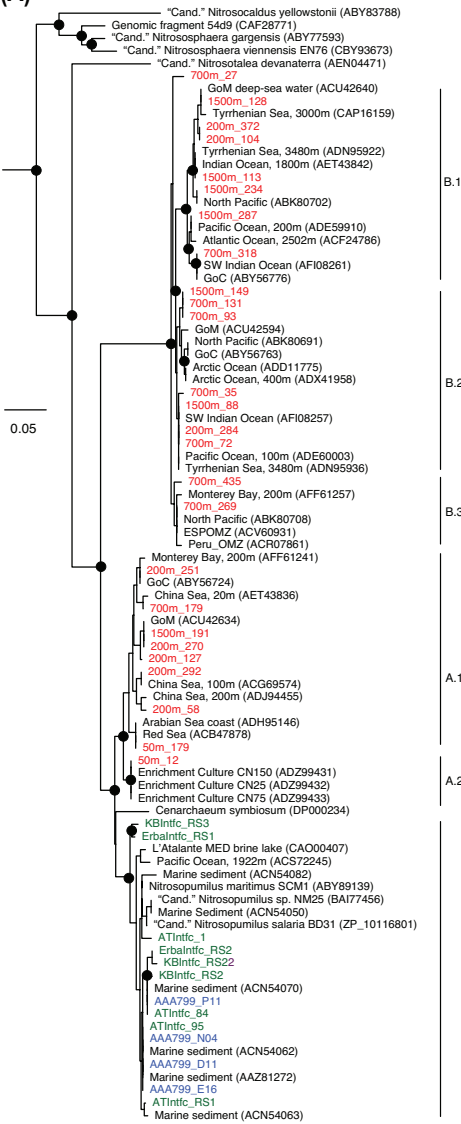*Figure S1*

Figure S1. Map depicting the geographic position of some of Red Sea's deep hypersaline anoxic brines, five of which were part of this study:
Atlantis II Deep, Discovery Deep, Erba Deep, Nereus Deep, and Kebrit Deep.

Figure S2

Figure S2. Relative abundance and distribution of all thaumarchaeal OTUs in brine-seawater interface (bsi) samples from five brines pools based on 454 amplicons data (A) and nearly full-length (~1200 bp) clonal 16S rRNA sequences (B) including those from the water column (200–1500 m) of Atlantis II Deep. For the 454 data (average sequence length ~600 bp), OTUs were obtained after subsampling and clustering 17,621 sequences (per sample) at a 3% distance threshold using MOTHUR. Colors reflect the relative abundance (Rel. abund.), and show that all BSI samples were dominated by a single OTU affiliated with the Shallow Marine Group I (SMGI) clade, whereas those from the water column compose of Deep Marine Group I (DMGI) clade. AD, Atlantis II Deep; DD, Discovery Deep; ED, Erba Deep; KD, Kebrit Deep; and ND, Nereus Deep

Figure S3

(A)

"Cand." Nitrosocaldus yellowstonii (ABY83788)
Genomic fragment 54d9 (CAF28771)
"Cand." Nitrososphaera gargensis (ABY77593)
"Cand." Nitrososphaera viennensis EN76 (CBY93673)
"Cand." Nitrosotalea devanaterra (AEN04471)

700m_27
GoM deep-sea water (ACU42640)
1500m_128
Tyrrhenian Sea, 3000m (CAP16159)
200m_372
200m_104
Tyrrhenian Sea, 3480m (ADN95922)
Indian Ocean, 1800m (AET43842)
1500m_113
1500m_234
North Pacific (ABK80702)
1500m_287
Pacific Ocean, 200m (ADE59910)
Atlantic Ocean, 2502m (ACF24786)
700m_318
SW Indian Ocean (AFI08261)
GoC (ABY56776)   B.1

1500m_149
700m_131
700m_93
GoM (ACU42594)
North Pacific (ABK80691)
GoC (ABY56763)
Arctic Ocean (ADD11775)
Arctic Ocean, 400m (ADX41958)
700m_35
1500m_88
SW Indian Ocean (AFI08257)
200m_284
700m_72
Pacific Ocean, 100m (ADE60003)
Tyrrhenian Sea, 3480m (ADN95936)   B.2

700m_435
Monterey Bay, 200m (AFF61257)
700m_269
North Pacific (ABK80708)
ESPOMZ (ACV60931)
Peru_OMZ (ACR07861)   B.3

Monterey Bay, 200m (AFF61241)
200m_251
GoC (ABY56724)
China Sea, 20m (AET43836)
700m_179
GoM (ACU42634)
1500m_191
200m_270
200m_127
200m_292
China Sea, 100m (ACG69574)
China Sea, 200m (ADJ94455)
200m_58
Arabian Sea coast (ADH95146)
Red Sea (ACB47878)
50m_179
50m_12   A.1

Enrichment Culture CN150 (ADZ99431)
Enrichment Culture CN25 (ADZ99432)
Enrichment Culture CN75 (ADZ99433)   A.2
Cenarchaeum symbiosum (DP000234)

KBIntfc_RS3
ErbaIntfc_RS1
L'Atalante MED brine lake (CAO00407)
Pacific Ocean, 1922m (ACS72245)
Marine sediment (ACN54082)
Nitrosopumilus maritimus SCM1 (ABY89139)
"Cand." Nitrosopumilus sp. NM25 (BAI77456)
Marine Sediment (ACN54050)
"Cand." Nitrosopumilus salaria BD31 (ZP_10116801)
ATIntfc_1
ErbaIntfc_RS2
KBIntfc_RS22
KBIntfc_RS2
Marine sediment (ACN54070)
AAA799_P11
ATIntfc_84
ATIntfc_95
AAA799_N04
AAA799_D11
Marine sediment (ACN54062)
AAA799_E16
ATIntfc_RS1
Marine sediment (ACN54063)

Water column B
Water column A
Nitrosopumilus-like Cluster

0.05

(B)

Crenarchaea type-1 (Hyperthermophilic)
"Anaerobe Cluster"

Cenarchaeum symbiosum (AABK76673)
BrackishWater (ADO79705)
BrackishWater_ADO79691
"Cand." Nitrosopumilus salaria BD31 (EIJ66608)
Brackish water sediment (ADO79702)
AAA799_P11
AAA799_B03
AAA799_E16
AAA799_D11
AAA799_N04
AAA799_G16
Nitrosopumilus maritimus SCM1 (ABX12103)
Brackish water sediment (ADO79693)
Brackish water sediment (ADO79707)
"Cand." Nitrosoarchaeum limnia SFB1 (ZP_08257474)
"Cand." Nitrosoarchaeum koreensis MY1 (ZP_08667209)

Mesophilic "Marine Cluster 1"

AAA799_D07
1500m_RS4
GOS_9176226 (EBH94044)
200m_RS10
200m_RS7
200m_RS9
200m_RS2   DSC I

1500m_RS7
Tyrrhenian Sea, 3480 m (ADV17457)
Tyrrhenian Sea, 3480 m (ADV17453)
1500m_RS10
1500m_RS8
200m_RS4
200m_RS5
1500m_RS12
200m_RS15
1500m_RS6
1500m_RS6
200m_RS6
200m_RS8
200m_RS13
1500m_RS5
Tyrrhenian Sea, 3480 m (ADV17449)
200m_RS3
200m_RS18
1500m_RS3
1500m_RS13
200m_RS1
200m_RS12
1500m_RS14
1500m_RS9
1500m_RS15
South Atlantic, 800m (AEI17237)
Tyrrhenian Sea, 3480 m (ADV17458)
Tyrrhenian Sea, 3480 m (ADV17452)
Tyrrhenian Sea, 3480 m (ADV17450)
HOT, 4000 m (ABZ08580)
Tyrrhenian Sea, 3480m (ADV17448)
Tyrrhenian Sea, 3480m (ADV17447)
Tyrrhenian Sea, 3480m (ADV17459)
Tyrrhenian Sea, 3480m (ADV17451)

Deep-sea Cluster (DSC)
DSC II

0.02
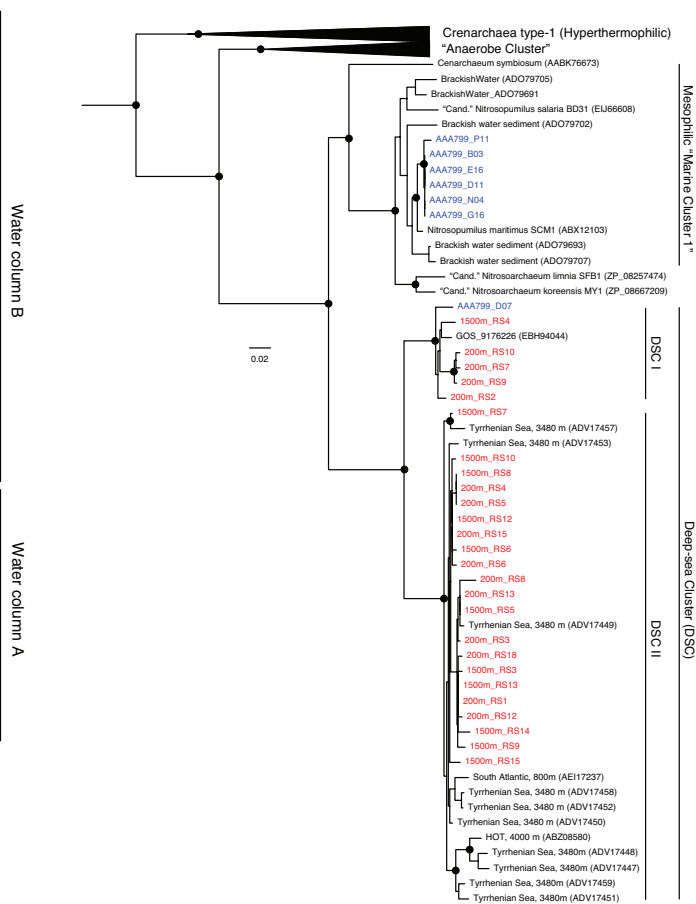
Figure S3. Phylogeny of putative AOA present in the water column overlying Atlantis II Deep (50–1500 m; in red)
and the brine-seawater interface of different brine pools (in green), based on the amoA (A) and 4-hydroxybutyryl-CoA dehydratase (B)
gene sequences. Sequences obtained from single-cell amplified genomes (SAGs) are shown in blue.
The phylogenetic trees were inferred using the neighbor-joining approach with a Jukes-Cantor distance correction matrix based
on 1000 replications, with bootstrap support values ≥ 70% displayed as a black-filled circle at the nodes. The nomenclature for the
amoA and 4-*hbdA* follows that from Francis et al. (2005) and Yakimov et al. (2011), respectively.
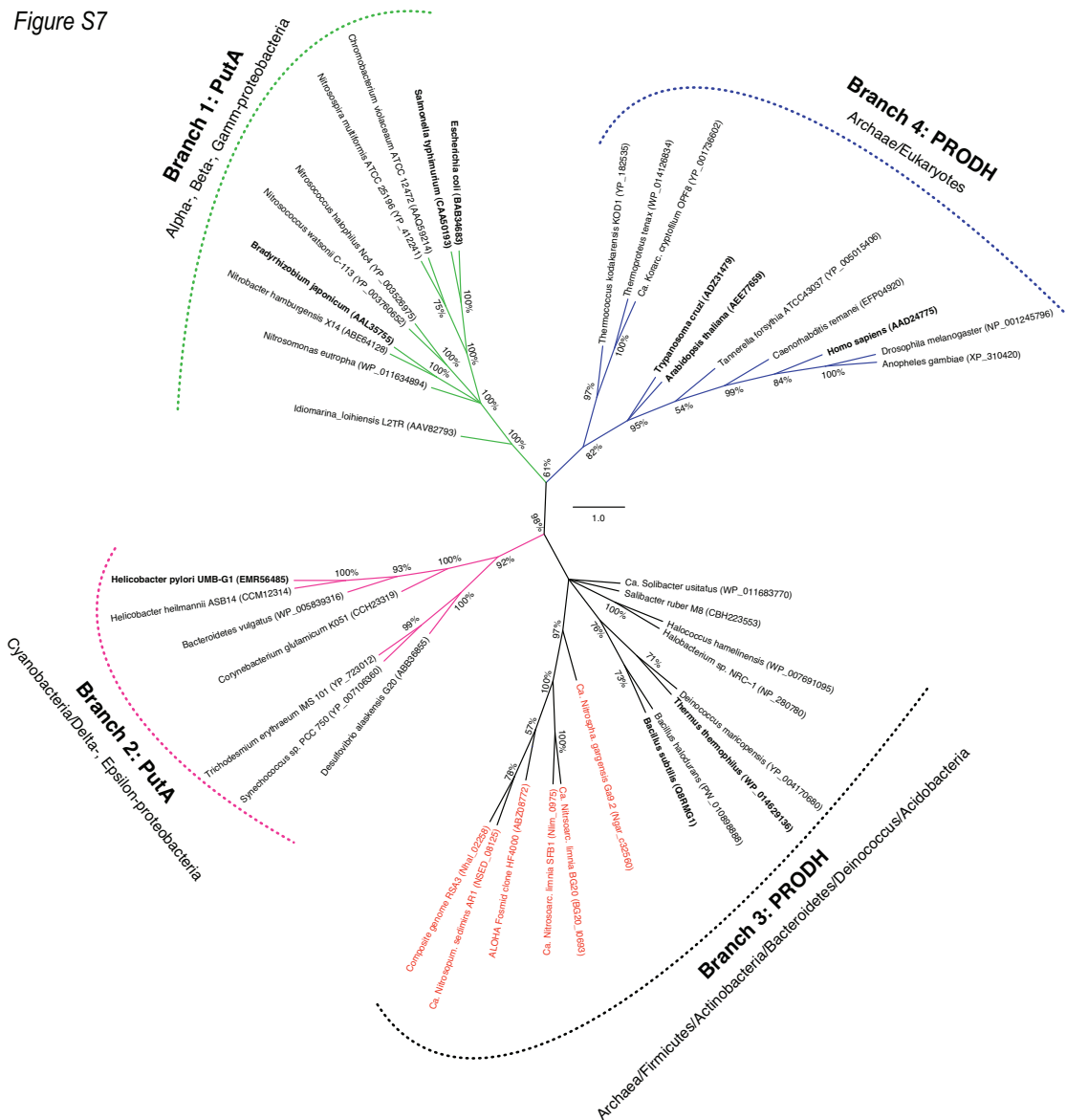
Figure S4
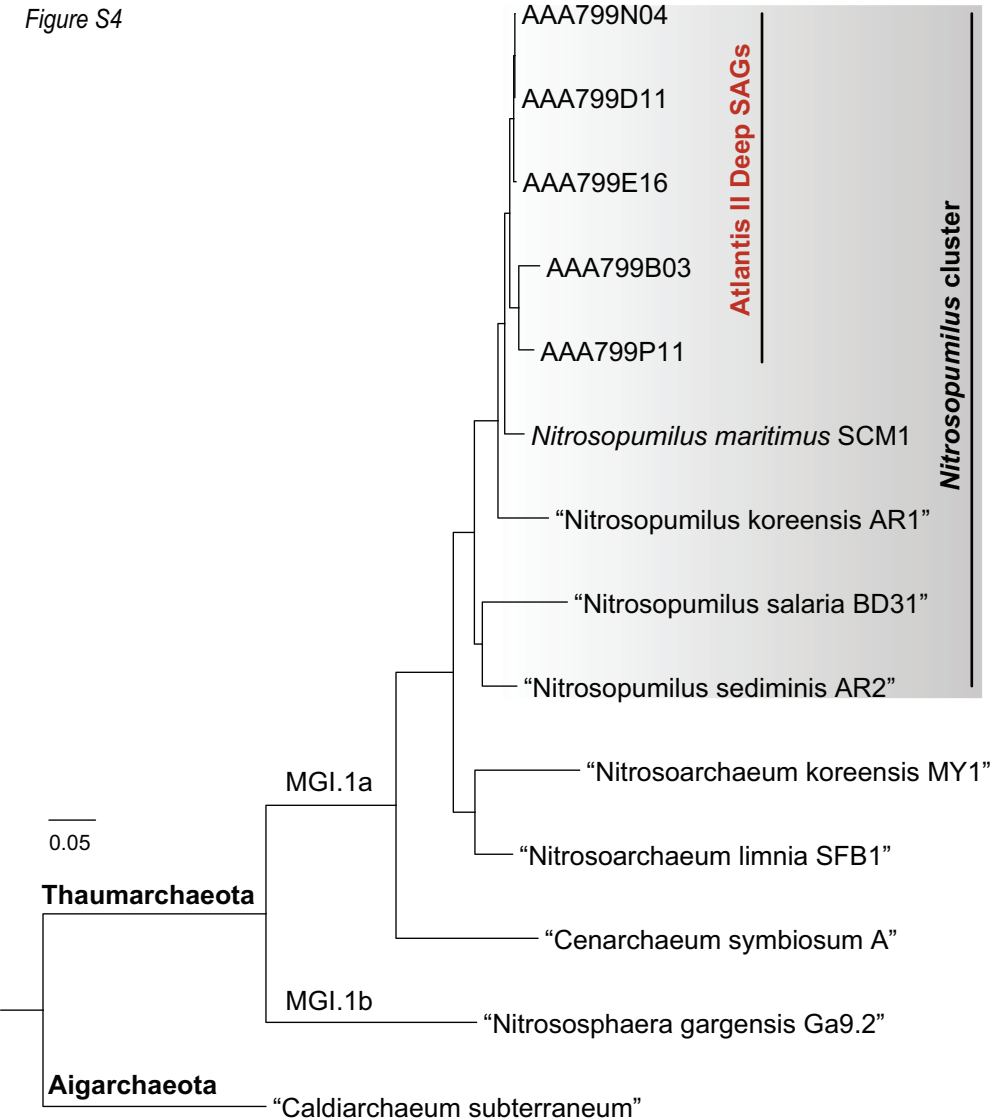
Figure S4. Core genes-based phylogenetic tree showing the tight clustering of our single-cell amplified genomes (SAGs), in particular E16, D11, and N04, whose reads were chosen for a combined genome assembly. The maximum-likelihood tree is inferred from a concatenation of 131 shared, single-copy genes present in these SAGs and other Thaumarchaea using RAxML under the PROTGAMMAI+WAG amino acid substitution model. All nodes are well supported with 100% bootstraps values (not shown) after 1000 replications.

Figure S5
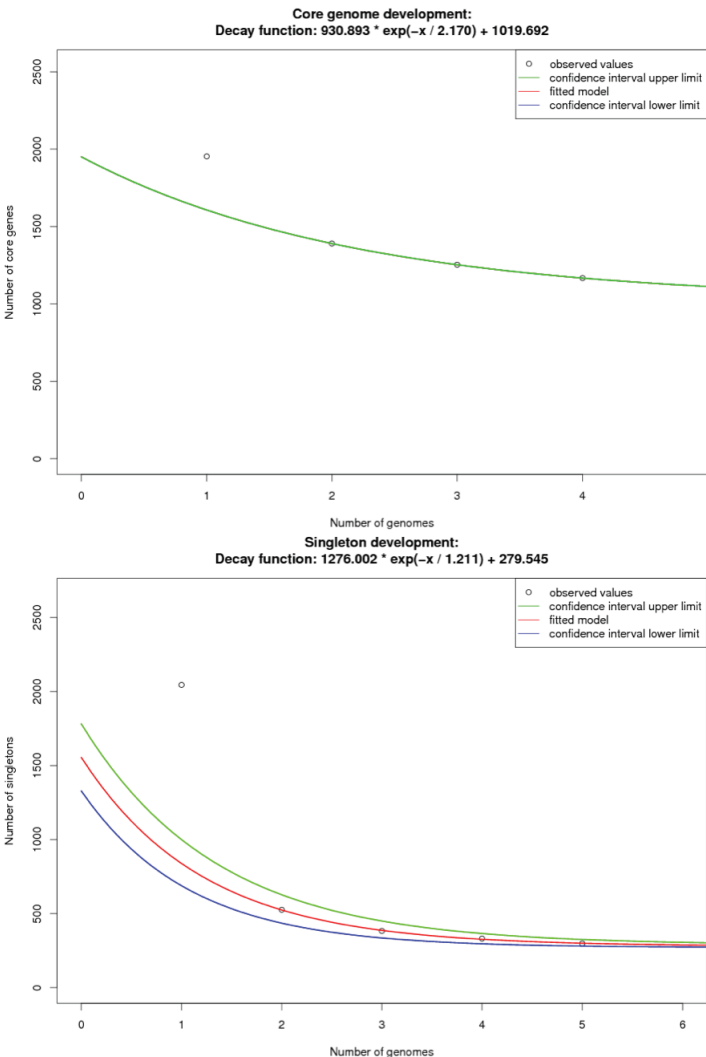
Figure S5. Core and singletons development plots predicting the number of core or unique genes that would be uncovered with addition of more *Nitrosopumilus* genomes. The curves are fitted with a nonlinear decay function. The asymptotic nature of the singletons plot (lower panel), with unique gene counts not approaching zero, means that we still have an open pan-genome.

Figure S6. An unrooted maximum-likelihood tree, showing the phylogeny of proline utilization (PutA) and proline dehydrogenase (PRODH) gene clusters. The nomenclature follows that of White et al. (2007; see Figure S1 in this publication) but now includes archaeal and eukaryotic sequences (this study). Each branch contains at least a single species with characterized proteins (in bold), while PRODH-containing thaumarchaea are shown in red. Methods are provided in supplementary methods.