Phylogenomics Suggests the Origin of Obligately Anaerobic Anammox Bacteria Correlates with the Great Oxidation Event

Tianhua Liao[^], Sishuo Wang[^], Haiwei Luo[*]

Simon F. S. Li Marine Science Laboratory, School of Life Sciences and State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, Hong Kong SAR

[^]These authors contribute equally to this work.

[*]**Corresponding author:**
Haiwei Luo
The Chinese University of Hong Kong
Shatin, Hong Kong SAR
Phone: (+852) 39436121
E-mail: hluo2006@gmail.com

Running title: The origin time and phylogeny of anammox bacteria

Keywords: anammox bacteria, molecular dating analysis, Planctomycetes

Abstract

The anaerobic ammonium oxidation (anammox) bacteria transform ammonium and nitrite to dinitrogen gas, and this obligate anaerobic process accounts for nearly half of global nitrogen loss. Yet its origin and evolution, which may give important insights into the biogeochemistry of early Earth, remains enigmatic. Here, we compile a comprehensive sequence data set of anammox bacteria, and confirm their single origin within the phylum Planctomycetes. After accommodating the uncertainties and factors influencing time estimates with different statistical methods, we estimate that anammox bacteria originated at around the so-called Great Oxidation Event (GOE; 2.32 to 2.5 billion years ago [Gya]) which is thought to have fundamentally changed global biogeochemical cycles. We further show that during the origin of anammox bacteria, genes involved in oxidative stress, bioenergetics and anammox granules formation were recruited, which may have contributed to their survival in an increasingly oxic Earth. Our findings suggest the rising level of atmospheric oxygen, which made nitrite increasingly available, as a potential driving force for the emergence of anammox bacteria. This is one of the first studies that link GOE to the evolution of obligate anaerobic bacteria.

Introduction

Accounting for up to 50% of the removal of fixed nitrogen (N) in nature, anaerobic ammonium oxidation (anammox, $NH_4^+ + NO_2^- \rightarrow N_2 + 2H_2O$) is recognized as an important biological process that leads to N loss, the other being denitrification [1]. Compared with denitrification, anammox is cost-effective and environmentally friendly thanks to its negligible emissions of greenhouse gases like N_2O, and consequently, anammox bacteria, the organisms that perform anammox, have been widely used in wastewater treatments [2]. Previous studies [3, 4] investigated the roles of key genes driving anammox, including *hzsABC* (hydrazine synthase), *hdh* (hydrazine dehydrogenase) and *hao* (hydroxylamine oxidoreductase). Further, all members of anammox bacteria are found within the deeply-branching phylum Planctomycetes. Six candidate genera of anammox bacteria, namely '*Candidatus* Brocadia', '*Candidatus* Kuenenia', '*Candidatus* Jettenia', '*Candidatus* Scalindua', '*Candidatus* Anammoximicrobium' and '*Candidatus* Anammoxoglobus', have been proposed based on 16S ribosomal RNA (rRNA) gene sequences [5], but none of them were obtained as typical pure cultures. Apparently, the habitat of anammox bacteria requires simultaneous presence of reduced (ammonia) and oxidized (nitrite) inorganic N compounds. Such habitats are often found at the aerobic-anaerobic interface in aquatic ecosystems including the margins of oxygen minimum zones (OMZs) in the ocean and sediment-water interfaces where ammonium originates from the anaerobic degradation of organic matter and nitrite could be produced by aerobic ammonia oxidation [6]. It is believed

that the early Earth was largely anaerobic and deficient in suitable niches for aerobic nitrogen cycles generating nitrite/nitrate, until around 2.4 billion years ago (Gya), the so-called Great Oxidation Event (GOE), plausibly as a result of the emergence of oxygenic cyanobacteria [7]. On the other side, ammonium, the other substrate of anammox metabolism, might have been accumulated during the mid-to-late Archaean [8]. It is therefore tempting to hypothesize that the origin of anammox coincided with GOE, a revolutionary event that makes molecular oxygen (O_2) to a concentration that is biologically meaningful, thereby creating the (micro)environments satisfying anammox metabolism.

The link between the evolution of metabolic pathways and associated geological factors could be implied by investigating their timelines. One way to explore the timescale of a special metabolic pathway is based on biogeochemical evidence, which, however, is often affected by the poor conservation of the biomarker during the long period of time [9]. For instance, ladderanes, a type of special lipids unique to anammox bacteria [10], are rarely preserved to a level that can be used to date their origin time. Further, the nitrogen isotope fractionations from sedimentary records are usually taken as evidence to determine the presence or absence of N cycling metabolism in the past [8, 11, 12]. However, the vague differences of isotopic fractionations between different metabolic pathways make it difficult to single out and elucidate the evolution of the different redox reactions within the N cycle [13]. Alternatively, molecular dating, which estimates the age of the last common ancestor (LCA) of analyzed lineages by

comparing their sequences based on the molecular clock theory [14], provides a powerful strategy to investigate this issue. Here, we hypothesize the rising level of O_2 during GOE, which made nitrite available to power anammox [15], as a driving force of the origin of anammox bacteria. To test this hypothesis and to obtain more insights into anammox bacteria evolution, we compiled an up-to-date genomic data set of Planctomycetes (see Supplementary Text section 1; Dataset S1.1), placed the evolution of anammox bacteria in the context of geological events, and investigated genomic changes underlying the origin of this ecologically important bacterial lineage.

Results and discussion

Monophyletic origins of anammox bacteria and anammox genes

The anammox bacteria form a monophyletic group embedded in other lineages of Planctomycetes in a comprehensive phylogenomic tree with 881 Planctomycetes genomes (Fig. S1). The anammox bacteria clade in this phylogenomic tree (Fig. S1) comprises four known genera including the early-branching *Ca.* Scalindua and *Ca.* Kuenenia, and the late-branching *Ca.* Jettenia and *Ca.* Brocadia. It also includes two understudied lineages which we named 'basal lineage' and '*hzsCBA*-less lineage' (Fig. S1), both of which have relatively low genome completeness compared to the other anammox bacteria and are represented by metagenome-assembled genomes (MAGs) sampled from groundwater (see Supplementary Text

section 2.1). Two described genera (*Ca*. Anammoximicrobium and *Ca.* Anammoxoglobus) do not have genomes available by the time of the present study (last accessed: April 2021), and are therefore not included in the phylogenomic analysis. Furthermore, using the last updated set of 2,077 Planctomycetes genomes (retrieved in April 2021) from NCBI, we obtained a consistent topology of anammox bacteria (Fig. S2).

Besides, to check whether important groups, particularly early-branching lineages, have been encompassed in our analyzed genome data sets, we built 16S rRNA gene tree (Fig. S3) using the identified 16S rRNA genes from genomic sequences of anammox bacteria and the deposited 16S rRNA gene amplicons in SILVA from the class Brocadiae where both anammox and non-anammox bacteria (see Supplementary Texts section 2.2). Since the anammox bacteria with genome sequences in this 16S rRNA gene tree (Fig. S3) showed a branching order congruent with the topology in the phylogenomic tree (Fig. S1), and since the 16S rRNA gene tree included 913 (clustered from 20,142) sequences sampled from a wide array of habitats (marine, sediments, man-made reactor, freshwater and terrestrial ecosystems), the 16S rRNA gene phylogeny likely encompasses the early-branching lineages of anammox bacteria and thus is proper to be used to study the evolutionary origin of anammox bacteria (Dataset S1.2).

We further analyzed the evolution (Fig. S4) of the genes (*hzsCBA*) critical to the anammox reaction [4, 16]. Reconciliation (see Supplementary Texts section 2.3) of the gene and the genome-based species phylogenies pointed to a single origin of anammox metabolism at the

LCA of anammox bacteria (Fig. S4). Taken together, the above analyses suggest that presumably our genome data sets have encompassed the earliest-branching lineages and deep phylogenetic diversity of anammox bacteria, and indicate its monophyletic origin, thereby providing a foundation for estimating the origin of anammox metabolism by estimating the origin time of anammox bacteria.

The evolutionary origin of anammox bacteria coincided with the rising O_2

We employed MCMCTree to estimate the divergence times of the anammox bacteria with i) soft bounds (time constraint allowing a small probability of violation) based on cyanobacteria fossils, and ii) the relaxed molecular clock approaches allowing substitution rates to vary among branches. Using a best-practiced dating scheme (Fig. 1; see Supplementary Text section 3), we dated the origin of anammox bacteria to 2,117 million years ago (Mya) (95% highest posterior density [HPD] interval, 2,002 - 2,226 Mya). The analysis with the expanded Planctomycetes genomes (see Supplementary Texts section 3) also shows similar time estimate at 2,105 Mya (95% HPD: 1,961 - 2,235 Mya).

Analyses with different combinations of calibrations and parameters broadly converge to similar time estimates (Fig. S5; Dataset S2.1). Specifically, assigning a more ancient time constraint to the total group of oxygenic cyanobacteria, which is based on the biogeochemical evidence of the presence of O_2 near 3.0 billion years ago (Gya) [17] instead of GOE (Dataset

S2.1), shifted the posterior dates of the LCA of anammox bacteria to the past by ~300 million years (Myr) (e.g., C1 versus C10; Fig. S5). Varying the maximum bounds of the root with either 4.5, 3.8 or 3.5 Gyr had a minor impact (around 100 Myr) on the time estimates of anammox bacteria (e.g., C1 versus C3; Fig. S5). Though the auto-correlated rates (AR) model was less favored than the independent rates (IR) model based on the model comparison (see Supplementary Text section 3.3; Dataset S2.2), it is worth noting that the posterior ages estimated by the AR model were 1-12% younger than estimated by the IR model (Dataset S2.1). Besides, there remains a possibility that there are unsampled or even extinct lineages of anammox bacteria that diverged earlier than all anammox bacteria analyzed in the present study. This scenario, if true, hints that the first anammox bacterium could have originated before the occurrence of the LCA of sequenced anammox bacteria but later than their split from the sister lineage (Node X in Fig. 1; up to 2,600 Mya). Accommodating the above uncertainties, our analyses suggest that the origin of anammox bacteria, hence the origin of anammox, was roughly around GOE and most likely ranges from 2,000 to 2,600 Mya.

<u>The geochemical context of the origin of anammox bacteria</u>

We speculate that the coincident time between the origin of an obligate anaerobe and the GOE might be ascribed to the increasing availability of nitrite, whose accumulation in turn depends on the availability of O_2 for biological nitrification [15, 18, 19]. It was shown that a low

concentration of nitrite significantly decreased the N removal rate of anammox in reactors [20]. In modern OMZs which presumably resemble the primordial Earth environment, the concentration of nitrite is the rate-limiting factor to the anammox metabolism [21]. Besides, it was suggested that nitrite might not reach a biologically meaningful level (micromolar level) on Earth until GOE, likely due to a few nitrite sinks related to the pervasive UV photolysis and Fe(II) reduction on early Earth [18]. Note that a prior study suggested that the seawater nitrite was presumably accumulated up to micromolar concentrations via lightning and photochemical processes [22], which implies the possibility of an earlier origin of anammox bacteria. However, this argument was made under a model assuming high amounts of CO_2, which was argued against because the concentration of CO_2 in late Archean was likely hundredth of that estimate [23]. The above arguments suggest that the concentration of nitrite before GOE was likely too low to support the origin of anammox bacteria.

In modern environments, the majority of nitrite used in anammox is derived from nitrate reduction or aerobic ammonia oxidation [24], the latter performed by ammonia-oxidizing archaea (AOA) or bacteria (AOB). Multiple studies [15, 25, 26] proposed that AOA and/or AOB likely appeared on Earth at about 2.7 Gya according to elevated $\delta^{15}N$ values (Fig. 1). In support of this idea, a recent study dated the LCA of AOA to be ~2,300 Myr driven by the increase of O_2 concentration from terrestrial ecosystems and that the expansion of AOA from land to the ocean did not occur until nearly 1.0 Gya [27]. This hints at a dominant role of AOB in early ocean.

Consistent with this idea, the vast majority of the early-branching lineages of anammox bacteria (*Ca*. Scalindua) and the sister Planctomycetes lineages of the anammox clade in the phylogenomic tree (Fig. 1) are found in marine environments. Furthermore, in the 16S rRNA gene tree which represents a greater diversity than the phylogenomic tree, marine lineages still account for the majority of the earliest-branching lineages of anammox bacteria (Fig. S3). Thus, it seems likely that the LCA of anammox bacteria might originate from the early ocean where ammonia-oxidizing organisms particularly AOB had likely inhabited. The above arguments, although speculative, tentatively suggest that the nitrite demand for anammox had been readily met by GOE by taking advantage of the significant amount of O_2 newly available, which is broadly consistent with our molecular dating results suggesting an origin of anammox around the GOE. Note that there are recently reported anammox bacteria lineages showing less dependence on nitrite, which may use nitric oxide or hydroxylamine as the electron receptor in anammox [5]. These lineages are affiliated with *Ca*. Kuenenia and *Ca.* Brocadia, which evolved recently, implying that the ability to use alternative electron receptors are likely not ancestral to all anammox bacteria.

<u>Genomic changes related to anammox metabolism upon the origin of anammox bacteria</u>

We further explored the genomic changes characterizing the origin of anammox bacteria (see Supplementary Text section 4). This includes the gains of the aforementioned genes that

directly participate in anammox, viz., *hzsABC*, *hdh*, and *hao*. The current view of the anammox metabolism posits that electrons consumed by anammox for carbon fixation are replenished by the oxidation of nitrite to nitrate by nitrite oxidoreductase (NXR) [28], which was inferred to be acquired upon the origin of anammox bacteria (Fig. 2). Moreover, multiple auxiliary genes for N assimilatory pathways including N regulatory protein (*glnB*), assimilatory nitrite reductase large subunit (*nirB*), transporters like *amt* (ammonium), *nirC* (nitrite) and NRT family proteins (nitrate/nitrite), were also gained at the origin of anammox bacteria (Fig. 2). However, genes for nitrogen related dissimilatory pathways encoding nitrite reduction to nitric oxide (*nirK* or *nirS*) and to ammonium (*nrfAH*), respectively, were likely acquired after the origin of anammox bacteria (Fig. 2). Anammox occurs at the membrane of anammoxosome, an organelle predominantly composed of a special type of lipid called ladderane. The ladderane membranes show low proton permeability, which helps maintain the proton motive force during the anammox metabolism [10]. Although the biosynthetic pathway of ladderane is yet to be characterized, a previous study [29] predicted 34 candidate genes responsible for the synthesis of ladderane lipids, four of which were potentially acquired during the origin of anammox bacteria (see Supplementary Text section 4.2; Fig. 2), hinting at their potential roles in ladderane synthesis and providing clues to future experimental investigations.

Anammox bacteria are detected at anoxic habitats and exhibit low oxygen tolerance [5]. Specifically, the O_2-sensitive intermediates generated by anammox, e.g., hydrazine, a powerful

reductant, calls for maintaining the anaerobic environments. Accordingly, the likely acquired peroxidases include cytochrome c peroxidase (Ccp) which could scavenge hydrogen peroxide in periplasm [30] and the most prevalent peroxidase [31], thioredoxin-dependent peroxiredoxin (AhpC). Besides peroxide scavengers, the desulfoferrodoxin (Dfx) which functions as superoxide reductase (SOR) to reduce superoxide to hydrogen peroxide and which is broadly distributed among anaerobic bacteria [31], were likely acquired during the origin of anammox bacteria (Fig. 2). Another acquired gene is *fprA* encoding flavo-diiron proteins which could scavenge O_2. Besides, *cbi* genes which are exclusively involved in the anaerobic biosynthesis of vitamin B_{12} [32], are widely present in the genomes of anammox bacteria, suggesting their acquisition during the early evolution of anammox bacteria (Fig. 2).

Investigating other metabolisms which were also acquired upon the origin of anammox bacteria could help us understand the coeval ecology. Here, we inferred that *sat* (Fig. 2) which encodes sulfate adenylyltransferase for assimilatory incorporating sulfate into bioavailable adenylyl sulfate, and *aprAB* and *fsr* (Fig. 2) encoding dissimilatory sulfite reductase, were acquired upon the origin of anammox bacteria (Fig. 2). The likely acquired sulfur metabolism is coincident with the increased concentrations of seawater sulfate from ~100 μM throughout much of the Paleoproterozoic to over 1mM after the rise of atmospheric O_2 [33]. Besides, unlike other Planctomycetes which are heterotrophs, anammox bacteria are generally autotrophs that use Wood-Ljungdahl pathway for carbon assimilation [34]. Consequently, they potentially lost

many genes involved in carbohydrate utilization (Fig. 2). The metabolic loss is further strengthened by the enrichment analysis where the genes predicted to be lost by the LCA of anammox bacteria are enriched in pathways involving hydrolase activity, intramolecular oxidoreductase activity and carbohydrate kinase activity (Fig. 2; see Data and Code availability), hinting at a decreased ability of anammox bacteria to degrade organics. Also lost are genes (*mnhABCDEG*) encoding Na^+/H^+ antiporter (Fig. 2) which usually provides sodium tolerance and pH homeostasis abilities in marine environments [35]. Only members of the dominant marine genus *Ca.* Scalindua keep this gene cluster. Further, though the way to build up proton motive force is vital to energy harvesting, it remains unexplored in anammox bacteria [36]. Here, we identified that *nuo* genes encoding NADH-quinone oxidoreductase which could build up proton motive force was likely acquired by the LCA of anammox bacteria (Fig. 2). Besides, the presence of genes coding for Na^+ coupled respiratory enzymes, for example *rnf* (ferredoxin-NADH oxidoreductase) and *nqr* (ubiquinone-NADH oxidoreductase), suggests that anammox bacteria could also use the sodium motive force for ATP synthesis (Fig. 2).

Additionally, iron is a vital element for HZS [3] and HDH [37], and increased iron concentrations could promote the growth of anammox bacteria, *in vitro* [38]. The ancient ocean is thought to be ferruginous (high iron availability) according to iron isotope and other sedimentary records across several coeval marine settings [39], which could provide large amounts of soluble iron for anammox bacteria. Interestingly, a series of iron related genes to

make use of the abundant iron from the environment were likely acquired upon the origin of anammox bacteria. First, anammox bacteria likely acquired gene *fur* encoding ferric uptake regulator for iron assimilation. Moreover, the iron-containing proteins, cytochrome, is encoded by the acquired genes *petBC* potentially encoding Rieske-heme b [28] and *CcsAB* encoding a protein involved in cytochrome c maturation systems (Fig. 2). Notably, *ahbABCD*, which encodes proteins synthesizing *b*-type hemes in anammox bacteria [40], were likely acquired before the origin of anammox bacteria (Fig. 2). Furthermore, we identified that the gene *Bfr* encoding oligomeric protein, bacterioferritin, involved in the uptake and storage of iron [41] was likely acquired during the origin of anammox bacteria (Fig. 2), which may help to hoard iron in the expanding euxinic (low soluble iron availability) ocean. Our results tentatively support the hypothesis that the microenvironments those primordial anammox bacteria colonized, potentially near the anoxic-oxic interface where nitrite was more available, could be iron limited.

Taken together, the timelines and corresponding genomic changes of anammox bacteria provide implications for the physiological characteristics of descendant anammox bacteria and the origin of other nitrogen transformation pathways in the context of an estimated evolution timeline of anammox bacteria. For example, since anammox bacteria emerged at the time when the nitrite concentration was merely available, the earlier branching lineage *Ca.* Scalindua exhibit high nitrite affinity (0.45 μM) compared to other anammox genera (up to 370 μM) [5, 42]. Likewise, the gains (Fig. S7; see Supplementary Text section 4.4) of genes *nrfAH* for the

dissimilatory nitrate reduction to ammonium (DNRA) at the late branching lineages (*Ca.* Brocadia and *Ca*. Kuenenia) around 800 Ma are consistent with the time when the nitrite/nitrate availability was increased by the Neoproterozoic Oxygenation Event [43]. The loss of a large number of genes involved in carbohydrate metabolism and the iron dependence upon their origin might limit the distribution of anammox bacteria to oligotrophic environments. The time estimate and genomic changes of the anammox bacteria have gone some way towards enhancing the understanding of the physiological characteristics of modern anammox bacteria and the historical N cycle.

Concluding remarks

Traditionally, the metabolic evolution during the GOE was mainly investigated in the phylum Cyanobacteria [44], which is likely because oxygenic photosynthesis is believed as a major driver of the GOE. Although the cyanobacteria-based strategy might vary a lot using different maximum constraints at the root in the time estimation of Alphaproteobacteria according to recent study [45], the relatively consistent time estimates shown here (Fig. S5) implied that this uncertainty was not significant for anammox bacteria in our study, which might be related to the different phylogenetic relationships with cyanobacteria between Alphaproteobacteria and Planctomycetes. More recently, scientists have increasingly realized that the impact of the GOE on the bacterial world may not be limited to cyanobacteria. This is

well illustrated in Ren et al [27] where the authors showed that the GOE is likely a driving force of the origin of the aerobic AOA and thus an aerobic nitrogen cycle on Earth. Furthermore, another molecular dating study suggested the expanded arsenic resistance systems as biological adaptations to the increased toxicity of oxidized arsenic species after the GOE [46]. However, to our knowledge, little is known about the relationships between the GOE and obligate anaerobic metabolism.

Here, using molecular dating and comparative genomics approaches, we link the emergence of an obligate anaerobic bacterial group, which drives the loss of fixed N in many environments, to the rise of O_2, and highlight their evolutionary responses to major environmental disturbances. Apparently, the GOE might open novel niches for the origin and subsequent expansion of diverse aerobic prokaryotes such as AOA and AOB, which might in turn facilitate the origin of other bacterial lineages, including some anaerobes like the anammox bacteria, by providing the resources for their energy conservation. Hence, our study suggests that the impact of GOE can well go beyond aerobic prokaryotes. Thus, the origin of anammox bacteria which is proposed driven by the prevalence of aerobic nitrogen transformation provided biological supports to isotopic evidence from sedimentary records. In particular, we show molecular dating analysis as a powerful approach to complement isotopic evidence to solve the origin time of anammox, which has the potential to be extended to the investigation of other ecologically important metabolisms.

Materials and Methods

A phylogenomic tree (Fig. S1) was generated based on the concatenated alignment of 120 ubiquitous proteins (bac120; Dataset S1.3) proposed for tree inference by the Genome Taxonomy Database [47]. All published genomic sequences (952 in total) affiliated with the phylum Planctomycetes in NCBI Genbank up to December 2019 were retrieved (Dataset S1.1). To further examine whether the patterns obtained with this Planctomycetes dataset hold the same, we further constructed an expanded set of Planctomycetes genomes by retrieving a total of 2,077 Planctomycetes genomes released by April 2021 at Genbank (Fig. S2; see Data and Code availability). Key anammox genes *hzsCBA* and *hdh* were identified against manually curated reference sequences by BLASTP. For each gene, identified protein sequences were aligned using MAFFT (v7.222) [48] and the alignments were refined by trimAl (v1.4) [49]. All phylogenies were constructed by IQ-tree (v1.6.2) [50] with substitution models automatically selected by ModelFinder [51] and branch support assessed with 1,000 ultrafast bootstrap replicates. Following manual curation, four well-recognized genera and two separate lineages of anammox bacteria were highlighted with different colours, and the presence of key genes were annotated with filled symbols beside the labels (Fig. S1). Furthermore, the phylogenomic tree was generated with similar methods using 16S rRNA genes identified from downloaded genomes and those retrieved from the SILVA database (Fig. S3; Dataset S1.2). All trees (including

phylogenomic and 16S trees) in our study were visualized with iTOL [52]. More details are shown in supplementary text, section 2.

Molecular dating analysis was carried out using the program MCMCTree from the PAML package (4.9j) [53]. In our study, 13 calibration sets constructed with different time constraints of the root and three calibration nodes within the cyanobacteria lineage were used (see Supplemental Texts section 3.2). The topology constraint for dating analysis was generated by 85 genomes which were sampled from the phylogenomic tree of phylum Planctomycetes (see Supplemental Texts section 3.1). To perform dating analysis, clock models and different calibration sets were compared in 26 schemes (see Supplemental Texts section 3; Dataset S2.1) [54, 55]. For each scheme, the approximate likelihood method [56] of MCMCTree were conducted in duplicate with identical iteration parameters (burn-in: 10,000; sample frequency: 20; number of sample: 20,000). The convergence of each scheme was evaluated by comparing the posterior dates of two independent runs. The repeat analysis using latest Planctomycetes genomes was conducted with the best practice dating strategy (IR model and calibration C1).

The phylogenetic differences between gene trees and the species tree were reconciliated by GeneRax [57] with unrooted gene tree as input and automatically optimized duplication, transfer, loss (DTL) rates. For each gene, we used a species tree comprised by all anammox bacteria pruned from the phylogenomic tree as the reference. We used recommended parameters including SPR strategy, undated DTL model and a maximum radius of five.

For comparative genomics analysis, the protein-coding sequences were annotated against KEGG [58], CDD [59], InterPro [60], Pfam [61], TIGRFAM [62] and TCDB [63], individually (see Supplemental Texts section 4). Following annotations, the potentially gained/lost genes between groups (anammox bacteria versus non-anammox bacteria; an anammox bacterial genus versus all other anammox bacteria) were statistically tested using two-sided fisher's exact tests. Finally, the resulting *p*-values were corrected with the Benjamini-Hochberg FDR (false discovery rate) procedure. Moreover, genes with corrected *p*-values smaller than 0.05 and with larger ratio in the study group were defined as gained genes, and genes with corrected *p*-values smaller than 0.05 and with smaller ratio in the study group were defined as potentially lost genes. Those potentially gained/lost genes were summarized for their detailed functions (see Data and Code availability).

Acknowledgements

We thank Bite Pei for building the basic dataset for this project, Yang Qian for his help in editing the earlier version of the manuscript, Jinjin Tao for her helpful discussion, and Mario dos Reis for his suggestion on dating analysis. This work is funded by the National Science Foundation of China (92051113), the Hong Kong Research Grants Council Area of Excellence Scheme (AoE/M-403/16), the Direct Grant of CUHK (4053495), and The CUHK Impact Postdoctoral Fellowship Scheme to (S.W.).

Data and Code availability

All used genomic sequences, generated phylogenetic trees, estimated divergence times and the used python codes are deposited in the online repository https://github.com/luolab-cuhk/anammox-origin.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Oshiki M, Satoh H, Okabe S. Ecology and physiology of anaerobic ammonium oxidizing bacteria. *Environ Microbiol* 2016; **18:** 2784-2796.
2. Li L, Ling Y, Wang HY, Chu ZS, Yan GK, Li ZW, et al. N_2O emission in partial nitritation-anammox process. *Chinese Chemical Letters* 2020; **31:** 28-38.
3. Dietl A, Ferousi C, Maalcke WJ, Menzel A, de Vries S, Keltjens JT, et al. The inner workings of the hydrazine synthase multiprotein complex. *Nature* 2015; **527:** 394-397.
4. Kartal B, Maalcke WJ, de Almeida NM, Cirpus I, Gloerich J, Geerts W, et al. Molecular mechanism of anaerobic ammonium oxidation. *Nature* 2011; **479:** 127-130.
5. Zhang L, Okabe S. Ecological niche differentiation among anammox bacteria. *Water Res* 2020; **171:** 115468.
6. Sonthiphand P, Hall MW, Neufeld JD. Biogeography of anaerobic ammonia-oxidizing (anammox) bacteria. *Front Microbiol* 2014; **5:** 399.
7. Lyons TW, Reinhard CT, Planavsky NJ. The rise of oxygen in Earth's early ocean and atmosphere. *Nature* 2014; **506:** 307-315.
8. Yang J, Junium CK, Grassineau NV, Nisbet EG, Izon G, Mettam C, et al. Ammonium availability in the Late Archaean nitrogen cycle. *Nature Geoscience* 2019; **12:** 553-557.

9. Hallmann C, Kelly AE, Gupta SN, Summons RE. Reconstructing Deep-Time Biology with Molecular Fossils. *Quantifying the Evolution of Early Life : Numerical Approaches to the Evaluation of Fossils and Ancient Ecosystems* 2011; **36:** 355-401.

10. Moss FR, Shuken SR, Mercer JAM, Cohen CM, Weiss TM, Boxer SG, et al. Ladderane phospholipids form a densely packed membrane with normal hydrazine and anomalously low proton/hydroxide permeability. *Proceedings of the National Academy of Sciences of the United States of America* 2018; **115:** 9098-9103.

11. Mettam C, Zerkle AL, Claire MW, Prave AR, Poulton SW, Junium CK. Anaerobic nitrogen cycling on a Neoarchaean ocean margin. *Earth and Planetary Science Letters* 2019; **527**.

12. Ader M, Thomazo C, Sansjofre P, Busigny V, Papineau D, Laffont R, et al. Interpretation of the nitrogen isotopic composition of Precambrian sedimentary rocks: Assumptions and perspectives. *Chemical Geology* 2016; **429:** 93-110.

13. Fuchsman CA, Stueken EE. Using modern low-oxygen marine ecosystems to understand the nitrogen cycle of the Paleo- and Mesoproterozoic oceans. *Environ Microbiol* 2021; **23:** 2801-2822.

14. Yang Z, Rannala B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol* 2006; **23:** 212-226.

15. Stueken EE, Kipp MA, Koehler MC, Buick R. The evolution of Earth's biogeochemical nitrogen cycle. *Earth-Science Reviews* 2016; **160:** 220-239.

16. Harhangi HR, Le Roy M, van Alen T, Hu BL, Groen J, Kartal B, et al. Hydrazine synthase, a unique phylomarker with which to study the presence and biodiversity of anammox bacteria. *Appl Environ Microbiol* 2012; **78:** 752-758.

17. Betts HC, Puttick MN, Clark JW, Williams TA, Donoghue PCJ, Pisani D. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat Ecol Evol* 2018; **2:** 1556-1562.

18. Ranjan S, Todd ZR, Rimmer PB, Sasselov DD, Babbin AR. Nitrogen Oxide Concentrations in Natural Waters on Early Earth. *Geochemistry Geophysics Geosystems* 2019; **20:** 2021-2039.

19. Cheng C, Busigny V, Ader M, Thomazo C, Chaduteau C, Philippot P. Nitrogen isotope evidence for stepwise oxygenation of the ocean during the Great Oxidation Event. *Geochimica Et Cosmochimica Acta* 2019; **261:** 224-247.

20. Li W, Zhuang JL, Zhou YY, Meng FG, Kang D, Zheng P, et al. Metagenomics reveals microbial community differences lead to differential nitrate production in anammox reactors with differing nitrogen loading rates. *Water Res* 2020; **169:** 115279.

21. Zhang L, Narita Y, Gao L, Ali M, Oshiki M, Ishii S, et al. Microbial competition among anammox bacteria in nitrite-limited bioreactors. *Water Res* 2017; **125:** 249-258.

22. Wong ML, Charnay BD, Gao P, Yung YL, Russell MJ. Nitrogen Oxides in Early Earth's Atmosphere as Electron Acceptors for Life's Emergence. *Astrobiology* 2017; **17:** 975-983.

23. Charnay B, Le Hir G, Fluteau F, Forget F, Catling DC. A warm or a cold early Earth? New insights from a 3-D climate-carbon model. *Earth and Planetary Science Letters* 2017; **474:** 97-109.

24. Dalsgaard T, Thamdrup B, Farias L, Revsbech NP. Anammox and denitrification in the oxygen minimum zone of the eastern South Pacific. *Limnology and Oceanography* 2012; **57:** 1331-1346.

25. Canfield DE, Glazer AN, Falkowski PG. The evolution and future of Earth's nitrogen cycle. *Science* 2010; **330:** 192-196.

26. Zerkle AL, Poulton SW, Newton RJ, Mettam C, Claire MW, Bekker A, et al. Onset of the aerobic nitrogen cycle during the Great Oxidation Event. *Nature* 2017; **542:** 465-467.

27. Ren M, Feng X, Huang Y, Wang H, Hu Z, Clingenpeel S, et al. Phylogenomics suggests oxygen availability as a driving force in Thaumarchaeota evolution. *ISME J* 2019; **13:** 2150-2161.

28. Kartal B, Keltjens JT. Anammox Biochemistry: a Tale of Heme c Proteins. *Trends Biochem Sci* 2016; **41:** 998-1011.

29. Rattray JE, Strous M, Op den Camp HJ, Schouten S, Jetten MS, Damste JS. A comparative genomics study of genetic products potentially encoding ladderane lipid biosynthesis. *Biol Direct* 2009; **4:** 8.

30. Van Vliet AH, Ketley JM, Park SF, Penn CW. The role of iron in Campylobacter gene regulation, metabolism and oxidative stress defense. *FEMS Microbiol Rev* 2002; **26:** 173-186.

31. Johnson LA, Hug LA. Distribution of reactive oxygen species defense mechanisms across domain bacteria. *Free Radical Biology and Medicine* 2019; **140:** 93-102.

32. Moore SJ, Warren MJ. The anaerobic biosynthesis of vitamin B12. *Biochem Soc Trans* 2012; **40:** 581-586.

33. Fakhraee M, Hancisse O, Canfield DE, Crowe SA, Katsev S. Proterozoic seawater sulfate scarcity and the evolution of ocean-atmosphere chemistry. *Nature Geoscience* 2019; **12:** 375-380.

34. Kartal B, Keltjens JT, Jetten MSM. Metabolism and Genomics of Anammox Bacteria. *Nitrification* 2011**:** 181-200.

35. Ito M, Morino M, Krulwich TA. Mrp Antiporters Have Important Roles in Diverse Bacteria and Archaea. *Front Microbiol* 2017; **8:** 2325.

36. Peeters SH, van Niftrik L. Trending topics and open questions in anaerobic ammonium oxidation. *Curr Opin Chem Biol* 2019; **49:** 45-52.

37. Akram M, Dietl A, Mersdorf U, Prinz S, Maalcke W, Keltjens J, et al. A 192-heme electron transfer network in the hydrazine dehydrogenase complex. *Sci Adv* 2019; **5:** eaav4310.

38. Bi Z, Qiao S, Zhou J, Tang X, Zhang J. Fast start-up of Anammox process with appropriate ferrous iron concentration. *Bioresour Technol* 2014; **170:** 506-512.

39. Poulton SW, Canfield DE. Ferruginous Conditions: A Dominant Feature of the Ocean through Earth's History. *Elements* 2011; **7:** 107-112.

40. Ferousi C, Lindhoud S, Baymann F, Kartal B, Jetten MS, Reimann J. Iron assimilation and utilization in anaerobic ammonium oxidizing bacteria. *Curr Opin Chem Biol* 2017; **37:** 129-136.

41. Keren N, Aurora R, Pakrasi HB. Critical roles of bacterioferritins in iron storage and proliferation of cyanobacteria. *Plant Physiol* 2004; **135:** 1666-1673.

42. Wu Z, Meng H, Huang X, Wang Q, Chen WH, Gu JD, et al. Salinity-driven heterogeneity toward anammox distribution and growth kinetics. *Appl Microbiol Biotechnol* 2019; **103:** 1953-1960.

43. Ader M, Sansjofre P, Halverson GP, Busigny V, Trindade RIF, Kunzmann M, et al. Ocean redox structure across the Late Neoproterozoic Oxygenation Event: A nitrogen isotope perspective. *Earth and Planetary Science Letters* 2014; **396:** 1-13.

44. Schopf JW. Geological evidence of oxygenic photosynthesis and the biotic response to the 2400-2200 ma "great oxidation event". *Biochemistry (Mosc)* 2014; **79:** 165-177.

45. Wang S, Meade A, Lam HM, Luo H. Evolutionary Timeline and Genomic Plasticity Underlying the Lifestyle Diversity in Rhizobiales. *mSystems* 2020; **5**.

46. Chen SC, Sun GX, Yan Y, Konstantinidis KT, Zhang SY, Deng Y, et al. The Great Oxidation Event expanded the genetic repertoire of arsenic metabolism and cycling. *Proc Natl Acad Sci U S A* 2020; **117:** 10414-10421.

47. Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2017; **2:** 1533-1542.

48. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013; **30:** 772-780.

49. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009; **25:** 1972-1973.

50. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015; **32:** 268-274.

51. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017; **14:** 587-589.

52. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* 2019; **47:** W256-W259.

53. Yang ZH. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 2007; **24:** 1586-1591.

54. Reis MD, Gunnell GF, Barba-Montoya J, Wilkins A, Yang Z, Yoder AD. Using Phylogenomic Data to Explore the Effects of Relaxed Clocks and Calibration Strategies on Divergence Time Estimation: Primates as a Test Case. *Syst Biol* 2018; **67:** 594-615.

55. Dos Reis M, Donoghue PC, Yang Z. Bayesian molecular clock dating of species divergences in the genomics era. *Nat Rev Genet* 2016; **17:** 71-80.

56. Dos Reis M, Yang Z. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol Biol Evol* 2011; **28:** 2161-2172.

57. Morel B, Kozlov AM, Stamatakis A, Szollosi GJ. GeneRax: A Tool for Species-Tree-Aware Maximum Likelihood-Based Gene Family Tree Inference under Gene Duplication, Transfer, and Loss. *Mol Biol Evol* 2020; **37:** 2763-2774.

58. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 2000; **28:** 27-30.

59. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* 2020; **48:** D265-D268.

60. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* 2019; **47:** D351-D360.

61. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res* 2019; **47:** D427-D432.

62. Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, et al. TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res* 2001; **29:** 41-43.

63. Saier MH, Jr., Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res* 2016; **44:** D372-D379.

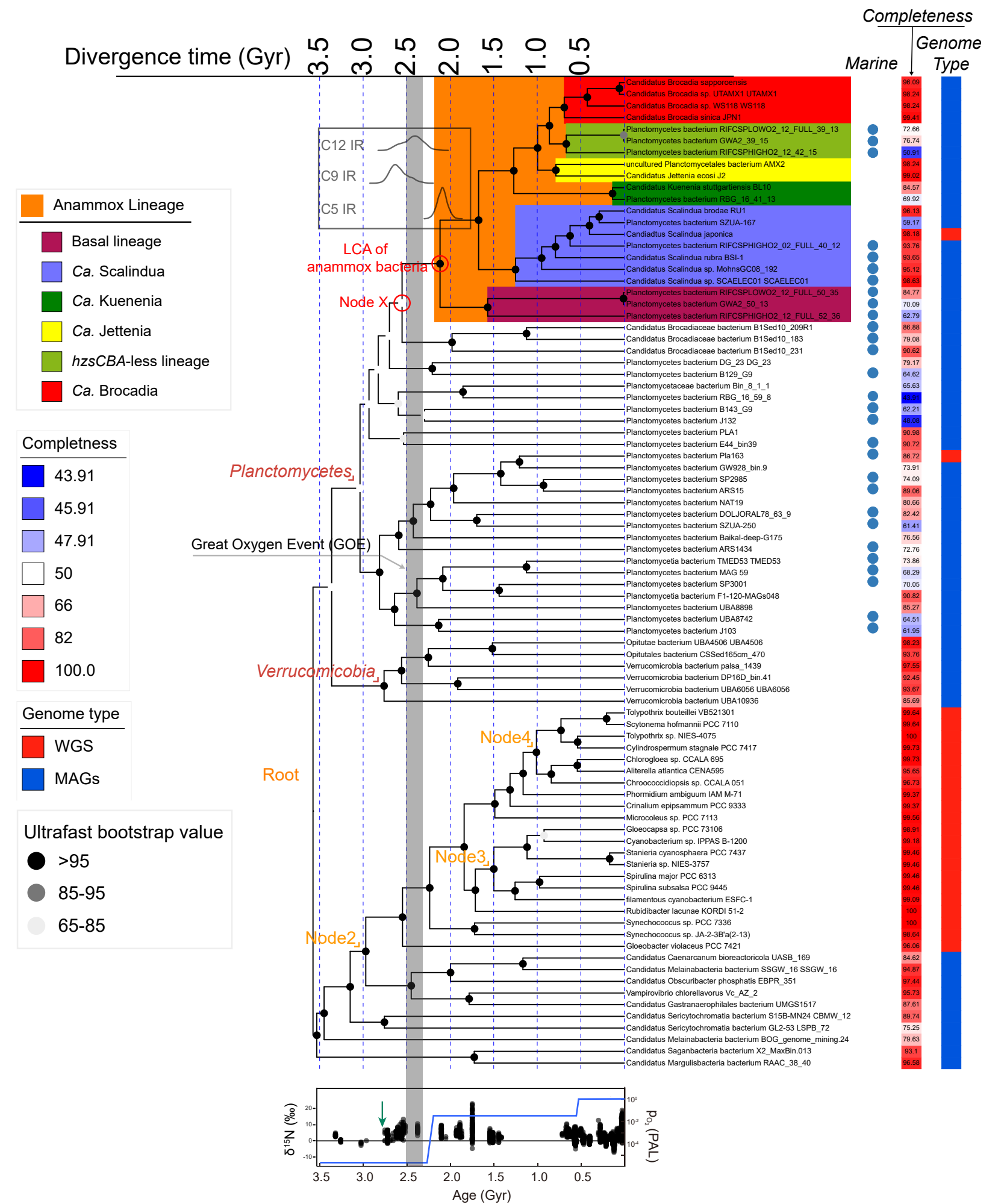
64. Kipp MA, Stueken EE, Yun M, Bekker A, Buick R. Pervasive aerobic nitrogen cycling in the surface ocean across the Paleoproterozoic Era. *Earth and Planetary Science Letters* 2018; **500:** 117-126.

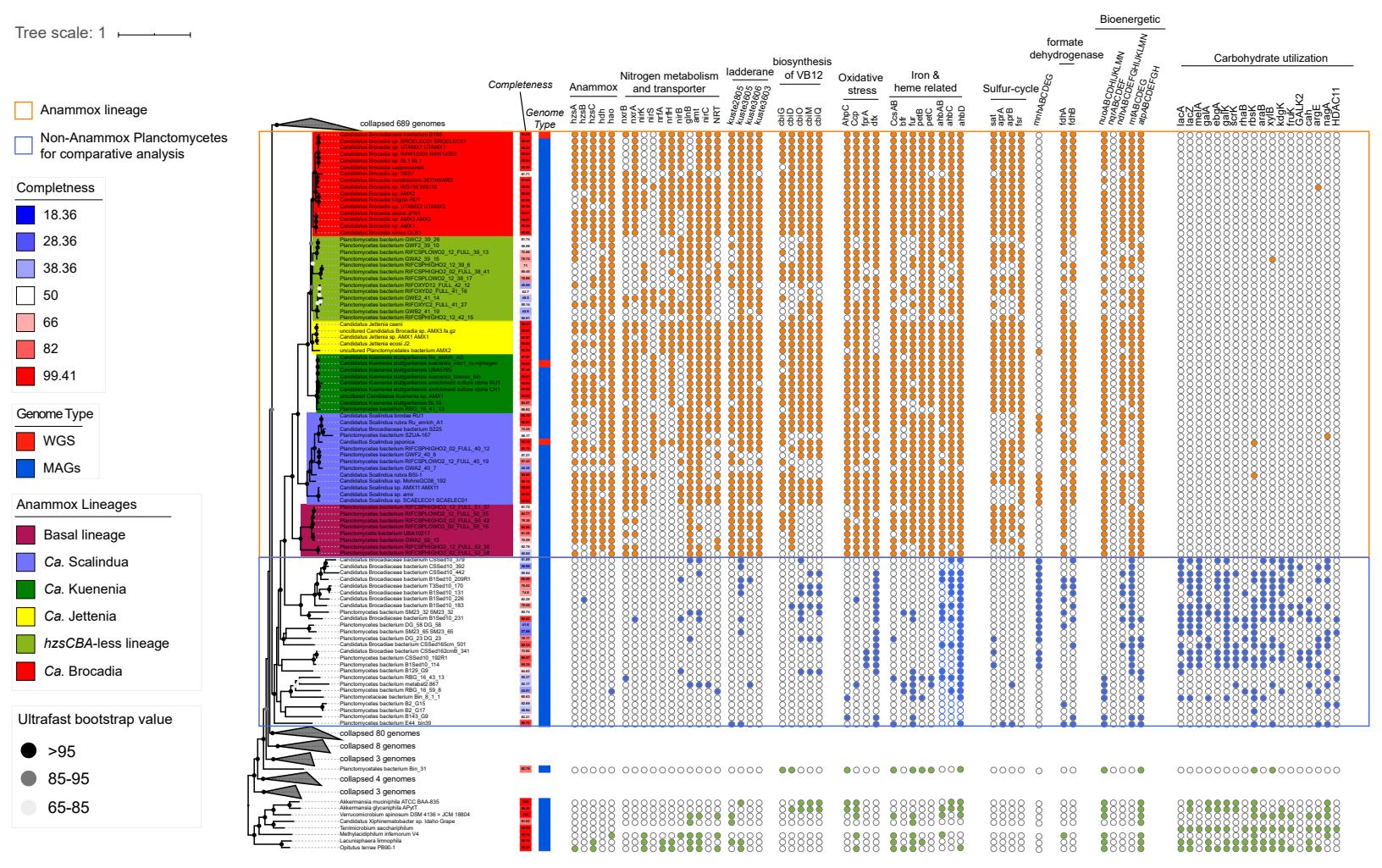
Figure legends

Figure 1. The evolutionary timeline of anammox bacteria using MCMCTree. The chronogram was estimated based on the calibration set C1 (see Supplementary Text section 3). The distributions of the posterior dates estimated using other calibration sets are shown above the last common ancestor (LCA) of anammox bacteria. These alternative calibration sets were selected by choosing the three that accommodate different calibration constraints (see Supplementary Texts section 3.3). More alternative time estimates are provided in Figure S4. The vertical grey bar represents the period of the GOE from 2,500 to 2,320 Mya. The calibration constraints used within the phylum Cyanobacteria are marked with orange texts: the LCA of Planctomycetes and Cyanobacteria (Root), the total group of oxygenic Cyanobacteria (Node 2), the total group of *Nostocales* (Node 3), and the total group of *Pleurocapsales* (Node 4). The genome completeness estimated by checkM is visualized with a gradient color strip. The right next color strip indicates the genome type of genomic sequences used in our study including metagenome-assembled genomes (MAGs) and whole-genome sequencing (WGS) from either enriched culture sample or isolate. The diagram below the dated tree illustrates the change of atmospheric partial pressure of O_2 (P_{O_2}) and nitrogen isotope fractionations. The blue line shows the proposed model according to Lyons et al [7]. The green arrow suggests the earliest evidence for aerobic nitrogen cycle at around 2.7 Ga. The PAL in right axis means P_{O_2} relative to the present atmospheric level. The black dots denote the change of nitrogen ^{15}N isotope values according to Kipp *et al* [64].

Figure 2. The phyletic pattern of ecologically relevant genes in the comparison between anammox bacteria and non-anammox bacteria. The phylogenomic tree on the left was constructed with 887 genomic sequences described in Supplementary Texts section 2.1. Solid circles at the nodes indicate the ultrafast bootstrap values in 1,000 bootstrapped replicates. Those Planctomycetes genomes not used for comparative genomics analyses are collapsed into grey triangles, and the numbers of collapsed genomes are labelled next to the triangles. The target group and reference group for comparative genomic analysis are within an orange or blue box, separately. For each genome, the genome completeness estimated by CheckM is visualized with a color strip and labelled besides leaf nodes. The right next color strip represents the type of genomic sequences used in our study including metagenome-assembled genomes (MAGs) and whole-genome sequencing (WGS) from either enriched culture sample or isolate. The filled and empty circles, respectively, represent the presence and absence of particular genes in corresponding genomes. For gene clusters, only genomes with at least half of the members of the gene cluster are indicated by a filled circle. The classifications of annotated genes are labelled above the gene names. *hzsCBA*, hydrazine synthase subunit C, B and A; *hdh*, hydrazine dehydrogenase; *hao*, hydroxylamine dehydrogenase; *nxrAB*, nitrite oxidoreductase subunit A and B; *nirK*, copper-containing and NO-forming nitrite reductase; *nirS*, cytochrome NO-forming nitrite reductase; *nrfAH*, ammonia-forming nitrite reductase subunit A and H; *glnB*, nitrogen

regulatory protein P-II; *nirC*, nitrite transporter; NRT, nitrate/nitrite transporter; *amt*, ammonium transporter; kuste2805, 3603, 3605-3606, proposed genes relative to the synthetic pathways for ladderane at Rattray et al [29]; *cbiG*, cobalt-precorrin 5A hydrolase; *cbiD*, cobalt-precorrin-5B(C1)-methyltransferase; *cbiOMQ*, cobalt/nickel transport system; *AhpC*, peroxiredoxin; *Ccp*, cytochrome *c* peroxidase; *dfx*, superoxide reductase; *fprA*, H_2O-forming enzyme flavoprotein; *CcsAB*, cytochrome c maturation systems; *petB*, Cytochrome b subunit of the bc complex; *petC*, Rieske Fe-S protein; *ahbABCD*, heme biosynthesis; *sat*, sulfate adenylyltransferase; *aprAB*, adenylylsulfate reductase, subunit A and B; *fsr*, sulfite reductase (coenzyme F420); *higB-1*, toxin; *higA-1*, antitoxin; *mnhABCDEG*, multicomponent Na^+/H^+ antiporter; *fdhAB*, formate dehydrogenase subunit A and B; *nuo*(A-N), NADH-quinone oxidoreductase; *ndh*(A-N), NAD(P)H-quinone oxidoreductase; *nqrABCDEF*, Na^+-transporting NADH:ubiquinone oxidoreductase; *rnfABCDEG*, Na^+-translocating ferredoxin:NAD^+ oxidoreductase; *atp*(A-H), F-type H^+-transporting ATPase; *lacAZ*, beta-galactosidase; *melA*, galA, alpha-galactosidase; *ebgA*, beta-galactosidase; *galK*, galactokinase; , fructokinase; *fruK*, 1-phosphofructokinase; *rhaB*, rhamnulokinase; rbsK, ribokinase; *araB*, L-ribulokinase; *xylB*, xylulokinase; *kdgK*, 2-dehydro-3-deoxygluconokinase; GALK2, N-acetylgalactosamine kinase; *cah*, cephalosporin-C deacetylase; *argE*, acetylornithine deacetylase; *nagA*; N-acetylglucosamine-6-phosphate deacetylase; *HDAC11*, histone deacetylase 11.


Divergence time (Gyr)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
Completeness
Marine
Genome Type
C12 IR
C9 IR
C5 IR
LCA of anammox bacteria
Node X
Anammox Lineage
Basal lineage
Ca. Scalindua
Ca. Kuenenia
Ca. Jettenia
hzsCBA-less lineage
Ca. Brocadia
Completness
43.91
45.91
47.91
50
66
82
100.0
Genome type
WGS
MAGs
Ultrafast bootstrap value
>95
85-95
65-85
Planctomycetes
Great Oxygen Event (GOE)
Verrucomicobia
Root
Node4
Node3
Node2
Candidatus Brocadia sapporoensis
Candidatus Brocadia sp. UTAMX1 UTAMX1
Candidatus Brocadia sp. WS118 WS118
Candidatus Brocadia sinica JPN1
Planctomycetes bacterium RIFCSPLOWO2_12_FULL_39_13
Planctomycetes bacterium GWA2_39_15
Planctomycetes bacterium RIFCSPHIGHO2_12_42_15
uncultured Planctomycetales bacterium AMX2
Candidatus Jettenia ecosi J2
Candidatus Kuenenia stuttgartiensis BL10
Planctomycetes bacterium RBG_16_41_13
Candidatus Scalindua brodae RU1
Planctomycetes bacterium SZUA-167
Candiadtus Scalindua japonica
Planctomycetes bacterium RIFCSPHIGHO2_02_FULL_40_12
Candidatus Scalindua rubra BSI-1
Candidatus Scalindua sp. MohnsGC08_192
Candidatus Scalindua sp. SCAELEC01 SCAELEC01
Planctomycetes bacterium RIFCSPLOWO2_12_FULL_50_35
Planctomycetes bacterium GWA2_50_13
Planctomycetes bacterium RIFCSPHIGHO2_12_FULL_52_36
Candidatus Brocadiaceae bacterium B1Sed10_209R1
Candidatus Brocadiaceae bacterium B1Sed10_183
Candidatus Brocadiaceae bacterium B1Sed10_231
Planctomycetes bacterium DG_23 DG_23
Planctomycetes bacterium B129_G9
Planctomycetaceae bacterium Bin_8_1_1
Planctomycetes bacterium RBG_16_59_8
Planctomycetes bacterium B143_G9
Planctomycetes bacterium J132
Planctomycetes bacterium PLA1
Planctomycetes bacterium E44_bin39
Planctomycetes bacterium Pla163
Planctomycetes bacterium GW928_bin.9
Planctomycetes bacterium SP2985
Planctomycetes bacterium ARS15
Planctomycetes bacterium NAT19
Planctomycetes bacterium DOLJORAL78_63_9
Planctomycetes bacterium SZUA-250
Planctomycetes bacterium Baikal-deep-G175
Planctomycetes bacterium ARS1434
Planctomycetia bacterium TMED53 TMED53
Planctomycetes bacterium MAG 59
Planctomycetes bacterium SP3001
Planctomycetia bacterium F1-120-MAGs048
Planctomycetes bacterium UBA8898
Planctomycetes bacterium UBA8742
Planctomycetes bacterium J103
Opitutae bacterium UBA4506 UBA4506
Opitutales bacterium CSSed165cm_470
Verrucomicrobia bacterium palsa_1439
Verrucomicrobia bacterium DP16D_bin.41
Verrucomicrobia bacterium UBA6056 UBA6056
Verrucomicrobia bacterium UBA10936
Tolypothrix bouteillei VB521301
Scytonema hofmannii PCC 7110
Tolypothrix sp. NIES-4075
Cylindrospermum stagnale PCC 7417
Chlorogloea sp. CCALA 695
Aliterella atlantica CENA595
Chroococcidiopsis sp. CCALA 051
Phormidium ambiguum IAM M-71
Crinalium epipsammum PCC 9333
Microcoleus sp. PCC 7113
Gloeocapsa sp. PCC 73106
Cyanobacterium sp. IPPAS B-1200
Stanieria cyanosphaera PCC 7437
Stanieria sp. NIES-3757
Spirulina major PCC 6313
Spirulina subsalsa PCC 9445
filamentous cyanobacterium ESFC-1
Rubidibacter lacunae KORDI 51-2
Synechococcus sp. PCC 7336
Synechococcus sp. JA-2-3B'a(2-13)
Gloeobacter violaceus PCC 7421
Candidatus Caenarcanum bioreactoricola UASB_169
Candidatus Melainabacteria bacterium SSGW_16 SSGW_16
Candidatus Obscuribacter phosphatis EBPR_351
Vampirovibrio chlorellavorus Vc_AZ_2
Candidatus Gastranaerophilales bacterium UMGS1517
Candidatus Sericytochromatia bacterium S15B-MN24 CBMW_12
Candidatus Sericytochromatia bacterium GL2-53 LSPB_72
Candidatus Melainabacteria bacterium BOG_genome_mining.24
Candidatus Saganbacteria bacterium X2_MaxBin.013
Candidatus Margulisbacteria bacterium RAAC_38_40
96.09
98.24
98.24
99.41
72.66
76.74
50.91
98.24
99.02
84.57
69.92
96.13
59.17
98.18
93.76
93.65
95.12
98.63
84.77
70.09
62.79
86.88
79.08
90.62
79.17
64.62
65.63
43.91
62.21
48.08
90.98
90.72
86.72
73.91
74.09
89.06
80.66
82.42
61.41
76.56
72.76
73.86
68.29
70.05
90.82
85.27
64.51
61.95
98.23
93.76
97.55
92.45
93.67
85.69
99.64
99.64
100
99.73
99.73
95.65
96.73
99.37
99.37
99.56
98.91
99.18
99.46
99.46
99.46
99.46
99.09
100
100
98.64
96.06
84.62
94.87
97.44
95.73
87.61
89.74
75.25
79.63
93.1
96.58
δ15N (‰)
20
10
0
-10
10⁰
10⁻²
10⁻⁴
PO2 (PAL)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
Age (Gyr)



Tree scale: 1
Anammox lineage
Non-Anammox Planctomycetes for comparative analysis
Completness
18.36
28.36
38.36
50
66
82
99.41
Genome Type
WGS
MAGs
Anammox Lineages
Basal lineage
Ca. Scalindua
Ca. Kuenenia
Ca. Jettenia
hzsCBA-less lineage
Ca. Brocadia
Ultrafast bootstrap value
>95
85-95
65-85
Completeness
Genome Type
Anammox
Nitrogen metabolism and transporter
ladderane
biosynthesis of VB12
Oxidative stress
Iron & heme related
Sulfur-cycle
formate dehydrogenase
Bioenergetic
Carbohydrate utilization
collapsed 689 genomes
collapsed 80 genomes
collapsed 8 genomes
collapsed 3 genomes
collapsed 4 genomes
collapsed 3 genomes