Phylogenomic evidence for the Origin of Obligately Anaerobic Anammox Bacteria around the Great Oxidation Event

Tianhua Liao[1^], Sishuo Wang[1^], Eva E. Stüeken[2], Haiwei Luo[1*]

[1]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

[2]School of Earth and Environmental Sciences and Centre for Exoplanet Science, University of St Andrews, Bute Building, Queen's Terrace, KY16 9TS, UK

[^]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 the global nitrogen loss for surface environments. 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 [Ga]) 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 on an increasingly oxic Earth. Our findings suggest the rising levels of atmospheric oxygen, which made nitrite increasingly available, was a potential driving force for the emergence of anammox bacteria. This is one of the first studies that link the GOE to the evolution of obligate anaerobic bacteria.

Introduction

Anaerobic ammonium oxidation (anammox, $NH_4^+ + NO_2^- \rightarrow N_2 + 2H_2O$), which usually occurs in anoxic marine, freshwater and wetland settings, accounts for up to 50% of the removal of fixed nitrogen (N) in nature. Along with denitrification, it is recognized as an important biological process that leads to N loss from the environment [1,2]. In contrast to the nitrification-denitrification process for wastewater treatment, coupling partial nitrification and anammox is cost-effective and environmentally friendly due to its lower oxygen requirement for aeration (nitrite [N(+III)], instead of nitrate [N(+V)] is sufficient for the anammox metabolism), carbon free cultivation (some denitrifying bacteria are heterotrophic) and its negligible emissions of greenhouse gases like N_2O. Consequently, anammox bacteria, the organisms that perform anammox, have been widely used in wastewater treatments [2-5]. Nevertheless, their evolutionary history and antiquity are poorly known, which hinders accurate reconstructions of the biogeochemical nitrogen cycle over geologic time.

Previous studies investigated the roles of key genes driving anammox, including *hzsABC* (hydrazine synthase)[6], *hdh* (hydrazine dehydrogenase), *hao* (hydroxylamine oxidoreductase) [7] and *nxr* (nitrite oxidoreductase) [8]. In general, the vital enzymes encoded by these genes either are directly involved in anammox or participate in replenishing electrons to the cyclic electron flow which is consumed by the reduction of NAD^+ during the ATP synthesis [9,10]. Further, all members of anammox bacteria are found within the phylum Planctomycetes [11]. 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 [12], but none of them have been successfully isolated into 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 can be produced by aerobic ammonia oxidation [13]. Although a recent study [14] revealed the potential of anammox bacteria to utilize metal oxides as the electron acceptor instead of nitrite, the feasibility of this metabolism in nature remains largely unresolved. Thus, the availability of nitrite is presumably the determinant factor to the origin of anammox. It is believed that the early Earth (before 3 Ga) was fully anoxic and deficient in aerobic ecosystems that were able to generate nitrite/nitrate [15-17]. The first transient and/or localized occurrences of aerobic nitrogen cycling appear in the Mid- to Neoarchean (2.7 Ga [18-21]), but evidence of widespread nitrate availability does not appear until around 2.4 billion years ago (Ga), the so-called Great Oxidation Event (GOE) [22-24]. The GOE marks the rise of free O_2 in the atmosphere above 10^{-5} times modern levels, and was ultimately a result of the emergence of oxygenic cyanobacteria [25]. In contrast to nitrite, ammonium, the other substrate of anammox

metabolism, was thermodynamically stable in the deep ocean throughout the Archean [26] and probably extending well into the Proterozoic [27]. It is therefore tempting to hypothesize that the origin of anammox coincided with Neoarchean oxygen oases or the GOE itself, when molecular oxygen (O_2) rose permanently to a concentration that is biologically meaningful [28], thereby creating the (micro-)environments satisfying anammox metabolism by providing enough nitrite.

Several tools exist for investigating the link between the evolution of metabolic pathways and geo-environmental transformations. One way to explore the evolutionary history of a specific metabolic pathway is based on organic biomarkers, which, however, are often affected by the poor preservation over geologic timescales [29]. For instance, ladderanes, a type of lipids that is unique to anammox bacteria [30], are rarely preserved to a level that can be used to date their evolutionary origin. Another approach is based on nitrogen isotope ratios of sedimentary records [26,31,32]. However, the isotopic fractionation factors for different metabolic pathways overlap widely (denitrification: -5 to -30‰; anammox: -16 or -24‰ [27]), making it difficult to single out and elucidate the evolution of the different redox reactions within the N cycle by this method [33]. 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 [34], provides a powerful strategy to investigate this issue. Here, we hypothesize that the rising level of O_2 during GOE, which made nitrite available to power anammox [27], was a driving force of the origin of anammox bacteria. To test this hypothesis and to obtain more insights into the evolution

of anammox bacteria, 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 into 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 within the phylum 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). Note that the absence of *hzsCBA* in several genomes within the '*hzsCBA*-less lineage' might be ascribed to the loss of these genes in evolution or their low genome completeness (Fig. S1). Because of their shallow (late-branching) phylogenetic position, this uncertainty should not affect the inference of the origin of anammox bacteria. 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).

To check whether important groups, particularly early-branching lineages, have been encompassed in our analyzed genome data sets, we built a 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, which comprises 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, which could better capture the diversity of anammox bacteria by including uncultured samples, likely encompasses the early-branching lineages of anammox bacteria. Taken together, our genome set, which encompasses the genomes of the early-branching lineages, is appropriate to be used to study the evolutionary origin of anammox bacteria (Dataset S1.3).

We further analyzed the evolution (Fig. S4) of the genes (*hzsCBA*) critical to the anammox reaction [7,35]. 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 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 (Ma) (95% highest posterior density [HPD] interval, 2,002 - 2,226 Ma). Further, we performed repeat analysis based on the expanded Planctomycetes genomes (Genome set 2; Dataset S1.1) or based on the alternative constraint topology inferred with a better profile mixture model (LG+C60+G; Genome set 1; Dataset S1.1). In general, both analyses resolved consistent topologies (Fig. S5) as shown in Figure 1. Overall, they gave similar posterior times of the LCA of the anammox bacteria at 2,105 Ma (95% HPD: 1,961 - 2,235 Ma) or 2,005 Ma (95% HPD: 1,869 - 2,146 Ma).

Analyses with different combinations of calibrations and parameters broadly converge to similar time estimates (Fig. S6; 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 at nearly 3.0 Ga [36] 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. S6). Varying the maximum bounds of the root of the tree of life between 4.5 Ga, 3.8 Ga or 3.5 Ga had a minor impact (around 100 Myr) on the time estimates of anammox bacteria (e.g., C1 versus C3; Fig. S6). 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 Ma). Accommodating the above uncertainties, our analyses suggest that the origin of anammox bacteria, hence the origin of anammox occurred most likely falls into the 2.6-2.0 Ga interval. Running MCMCTree analysis with no sequence data showed different distributions of time estimates for both the clade of anammox bacteria and

the four calibration points (Fig. S7), suggesting that sequence data are informative for our molecular clock analysis [37].

The geochemical context of the origin of anammox bacteria

We speculate that the timing of the origin of the anammox metabolism is linked to the increasing availability of nitrite in surface environments, because ammonium was likely present in the deep anoxic ocean throughout the Precambrian [27] and therefore probably not a limiting substrate. The first abiotic source of nitrite on the early Earth would have been lightning reactions between N_2 and CO_2 in the Archean atmosphere [38,39]. A prior study suggested that this process led to micromolar levels of nitrite in seawater [38], which implies the possibility of an earlier origin of anammox bacteria. However, this estimate was based on perhaps unrealistically high amounts of CO_2 [40]. Furthermore, there is no isotopic evidence in the early Archean rock record prior to 3 Ga for the presence of a significant nitrite or nitrate reservoir in the early ocean [15-17], and it is conceivable that any nitrite that was supplied to the ocean by lightning was rapidly reduced to ammonium or N_2 by ferrous iron, possibly even abiotically [41,42]. Hence, although a background lightning flux of nitrite almost certainly existed, there is no evidence that it supplied a reliable metabolic substrate for early life.

The first transient appearances of nitrite/nitrate-dependent metabolisms are captured by the sedimentary nitrogen isotope record in late Mesoarchean soils at 3.2 Ga [19] and in Neoarchean

shallow-marine settings at 2.7 Ga and 2.5 Ga [18,20,21]. These observations may reflect local and/or temporally restricted oxygen oases [43,44]. Widespread nitrite/nitrate availability is inferred for the Paleoproterozoic (2.4-1.8 Ga), i.e. in the immediate aftermath of the GOE [22-24]. Importantly, O_2 is necessary, although a trace amount is feasible, to extant nitrifying organisms for the oxidation of ammonium to nitrite and nitrate [45-47]. Hence the rise of oxygen almost certainly triggered the growth of the nitrite/nitrate reservoir, and therefore provided one of the key substrates for anammox bacteria. It has been shown that a low concentration of nitrite significantly decreases the N removal rate by anammox bacteria in reactors [48]. In modern OMZs, which may to some extent serve as analogues for the Paleoproterozoic redox-stratified ocean, the concentration of nitrite is the rate-limiting factor to the anammox metabolism [49]. Hence it was most likely that the rise of nitrite, driven by the rise of O_2, facilitated the appearance of anammox bacteria at 2.6-2.0 Ga.

In modern environments, the majority of nitrite used in anammox is derived from nitrate reduction or aerobic ammonia oxidation [50], the latter performed by ammonia-oxidizing archaea (AOA) or bacteria (AOB). As noted above, the nitrification likely appeared by at least 2.7 Ga, as indicated by nitrogen-isotope evidence for local accumulation of nitrate in surface ocean waters [27,51,52]. A recent study inferred that the LCA of AOA dates back to ~2.3 Ga and first appeared on land, driven by increasing O_2 concentrations in the atmosphere at that time, and that the expansion of AOA from land to the ocean did not occur until nearly 1.0 Ga [53]. This hints at a

dominant role of AOB in early ocean, before 2.3 Ga. 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 originated in the marine realm where AOB thrived first. 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 shortly before or after 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 [12]. 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. This aspect further indicates that lightning was likely not an important source of metabolic substrates in the early Archean, because nitric oxide is the major product of the lightning reaction between N_2 and CO_2 [38,39].

<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) [10], 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 [30]. Although the biosynthetic pathway of ladderane is yet to be characterized, a previous study [54] 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 occur in anoxic habitats and exhibit low oxygen tolerance [12]. Specifically, the O_2-sensitive intermediates generated by anammox, e.g., hydrazine, a powerful reductant, requires anoxic conditions. Accordingly, the likely acquired peroxidases include cytochrome c peroxidase (Ccp), which could scavenge hydrogen peroxide in the periplasm [55], and the most prevalent peroxidase [56], 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 [56], were likely acquired during the origin of anammox bacteria (Fig. 2). Another acquired gene is *fprA*, which encodes flavo-diiron proteins that scavenge O_2. Furthermore, *cbi* genes, which are exclusively involved in the anaerobic biosynthesis of vitamin B_{12} [57], are widely present in the genomes of anammox bacteria, suggesting their acquisition during the early evolution of anammox bacteria (Fig. 2).

Investigating other metabolisms that were also acquired upon the origin of anammox bacteria allow reconstructing the coeval ecology. Our data show that *sat* (Fig. 2), which encodes sulfate adenylyltransferase for assimilatory incorporation of sulfate into bioavailable adenylyl sulfate, as well as *aprAB* and *fsr* (Fig. 2) for encoding dissimilatory sulfite reductase, were acquired upon the origin of anammox bacteria (Fig. 2). This acquired capability of sulfate metabolism is coincident with the increased concentrations of seawater sulfate from ~100 μM or less throughout much of the Archean [58,59] to over 1mM after the rise of atmospheric O_2 [60,61].

Unlike other Planctomycetes, anammox bacteria are generally autotrophs that use the Wood-Ljungdahl pathway for carbon assimilation [62]. 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 the Na^+/H^+ antiporter (Fig. 2), which usually regulates sodium tolerance and pH homeostasis in marine environments [63]. Only members of the dominant marine genus *Ca.* Scalindua keep this gene cluster. Although building up a proton motive force is vital to energy harvesting, the mechanism for this process was far unexplored in anammox bacteria [64]. Our analysis shows that *nuo* genes, which encode NADH-quinone oxidoreductase that can build up a proton motive force, was likely acquired by the LCA of anammox bacteria (Fig. 2). Moreover, 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 [6] and HDH [65], and *in vitro* studies revealed that increased iron concentrations can promote the growth of anammox bacteria [66]. The ancient oceans are thought to have been ferruginous (high iron availability) according to sedimentary

records of iron speciation across several coeval marine settings from the Archean through most of the Proterozoic [67], meaning that large amounts of soluble iron would have been available 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 the gene *fur*, which encodes a ferric uptake regulator for iron assimilation. Moreover, the iron-containing proteins, specifically cytochrome, are encoded by the acquired genes *petBC*, which potentially encodes Rieske-heme b [10], and *CcsAB*, which encodes a protein involved in cytochrome c maturation systems (Fig. 2). Notably, *ahbABCD*, which encodes proteins-synthesizing *b*-type hemes in anammox bacteria [68], were likely acquired before the origin of anammox bacteria (Fig. 2). Furthermore, we identified that the gene *Bfr* that encodes the oligomeric protein, bacterioferritin involved in the uptake and storage of iron [69] was likely acquired during the origin of anammox bacteria (Fig. 2), which may help to hoard iron in settings where iron became locally sparse, such as along euxinic (sulfide-rich) or oxic marine margins in the Paleoproterozoic [70]. It is thought that the appearance of euxinic margins trapped iron that upwelled from the deep ocean, leading to the demise of iron oxide deposits (i.e., banded iron formations) [71]. Our results tentatively support this model, if the microenvironments that those primordial anammox bacteria colonized were located near the euxinic-oxic interface where nitrite and ammonium were available while upwelled iron was titrated out of the water column by freely dissolved hydrogen sulfide.

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-transforming pathways in the context of an estimated evolutionary timeline of anammox bacteria. For example, the earlier branching lineage *Ca.* Scalindua exhibit high nitrite affinity (0.45 μM) compared to other anammox genera (up to 370 μM) [12,72], which may reflect a progressive growth of the marine nitrite reservoir, linked to the protracted oxygenation of the surface environments over geologic timescales [25]. Likewise, genomic data reveal that also canonical denitrification radiated across the tree of life after the GOE [73], and the gains (Fig. S8; see Supplementary Text section 4.4) of genes *nrfAH* for dissimilatory nitrate reduction to ammonium (DNRA) at the late-branching lineages (*Ca.* Brocadia and *Ca.* Kuenenia) at around 800 Ma are consistent with the time when the nitrite/nitrate availability was increased by the Neoproterozoic Oxygenation Event [74]. 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 of the biosphere across the GOE has mainly been investigated in the phylum Cyanobacteria [75], likely because oxygenic photosynthesis is believed as a major driver of the GOE. More recently, scientists have increasingly realized that the impact

of the GOE on the bacterial world may not be limited to cyanobacteria. This was well illustrated by Ren, et al. [53], who showed that the GOE was 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 expanded arsenic resistance systems as biological adaptations to the increased toxicity of oxidized arsenic species after the GOE [76]. However, to our knowledge, little was so far known about the relationships between the GOE and obligate anaerobic metabolisms.

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 opened novel niches for the origin and subsequent expansion of diverse aerobic prokaryotes such as AOA and AOB, which in turn facilitated the origin of other bacterial lineages, including some anaerobes like the anammox bacteria, by providing the resources for their energy conservation. Our results open up the possibility that a significant proportion of the Proterozoic nitrite budget was consumed by anammox bacteria, and that the sedimentary nitrogen isotope record may be influenced by their activity. Given that anammox bacteria tend to produce less N_2O as a by-product compared to canonical nitrification/denitrification [4,5], our findings may have implications for Proterozoic climate, because N_2O has previously been invoked as an important greenhouse gas at that time [77]. Lastly, our study suggests that the impacts of the GOE go well beyond aerobic prokaryotes.

Specifically, the origin of anammox bacteria was likely driven by the expansion of nitrite/nitrate availability, as supported by isotopic evidence from sedimentary records. We therefore conclude that molecular dating is a powerful approach to complement isotopic evidence for resolving the timeline of biological evolution and for providing additional constraints on climate models of the distant past.

Materials and Methods

Overall, we compiled two genome sets and one 16S rRNA gene set in our study (Dataset S1.1). 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 [78]. All published genomic sequences (952 in total) affiliated with the phylum Planctomycetes in NCBI Genbank up to December 2019 were retrieved (Dataset S1.2). 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) [79] and the alignments were refined by trimAl (v1.4) [80]. All phylogenies were constructed by IQ-tree (v1.6.2) [81] with substitution models automatically selected by

ModelFinder [82] and branch support assessed with 1,000 ultrafast bootstrap replicates. Note that the constraint topologies for dating analysis (Fig. 1; Fig. S5) were inferred with the profile mixture model (LG+C60+G) which better accommodates across-site heterogeneity in deep-time evolution. 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 S1.3). All trees (including phylogenomic and 16S trees) in our study were visualized with iTOL [83]. 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) [84]. 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 using 85 genomes, which were sampled from the phylogenomic tree of phylum Planctomycetes (see Supplemental Texts section 3.1), and the model LG+C20+F+G under posterior mean site frequency (PMSF) approximation [85]. To perform dating analysis, clock models and different calibration sets were compared in 26 schemes (see Supplemental Texts section 3; Dataset S2.1) [86,87]. For each scheme, the approximate likelihood method [88] 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. With the updated genome set 2 (Dataset S1.1), we repeated the taxon sampling process and generated an alternative phylogeny for dating analysis (Fig. S5) using LG+C20+F+G model (C20: 20 classes of site-specific amino acid profiles) under PMSF approximations. Similarly, we also generated another phylogeny for dating analysis using the same genome set as used in Fig. 1 but with the profile mixture model (LG+C60+G) which uses 60 classes of amino acid profiles [89]. Note that the missing parameter +F would not significantly change the results since it only adds another profile amino acid calculated from the original data. These two repeat dating analyses were conducted with the best practice dating scheme (IR model and calibration C1).

The phylogenetic differences between gene trees and the species tree were reconciliated by GeneRax [90] with unrooted gene tree as input and automatically optimized duplication, transfer, loss (DTL) rates. For each gene identified from the genome set 2 (Dataset S1.1), we used a species tree comprised by all anammox bacteria pruned from the phylogenomic tree (Fig. S2) 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 [91], CDD [92], InterPro [93], Pfam [94], TIGRFAM [95] and TCDB [96], 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* **18**, 2784-2796, doi:10.1111/1462-2920.13134 (2016).

2 Stein, L. Y. & Klotz, M. G. The nitrogen cycle. *Current Biology* **26**, R94-R98, doi:DOI 10.1016/j.cub.2015.12.021 (2016).

3 Li, L. *et al.* N_2O emission in partial nitritation-anammox process. *Chinese Chemical Letters* **31**, 28-38, doi:10.1016/j.cclet.2019.06.035 (2020).

4 Okabe, S., Oshiki, M., Takahashi, Y. & Satoh, H. N2O emission from a partial nitrification-anammox process and identification of a key biological process of N2O emission from anammox granules. *Water Research* **45**, 6461-6470, doi:10.1016/j.watres.2011.09.040 (2011).

5 Gao, D. W. & Tao, Y. Versatility and application of anaerobic ammonium-oxidizing bacteria. *Applied Microbiology and Biotechnology* **91**, 887-894, doi:10.1007/s00253-011-3411-3 (2011).

6 Dietl, A. *et al.* The inner workings of the hydrazine synthase multiprotein complex. *Nature* **527**, 394-397, doi:10.1038/nature15517 (2015).

7 Kartal, B. *et al.* Molecular mechanism of anaerobic ammonium oxidation. *Nature* **479**, 127-130, doi:10.1038/nature10453 (2011).

8 Chicano, T. M. *et al.* Structural and functional characterization of the intracellular filament-forming nitrite oxidoreductase multiprotein complex. *Nature Microbiology*, doi:10.1038/s41564-021-00934-8 (2021).

9 Hu, Z., Wessels, H., van Alen, T., Jetten, M. S. M. & Kartal, B. Nitric oxide-dependent anaerobic ammonium oxidation. *Nat Commun* **10**, 1244, doi:10.1038/s41467-019-09268-w (2019).

10 Kartal, B. & Keltjens, J. T. Anammox Biochemistry: a Tale of Heme c Proteins. *Trends Biochem Sci* **41**, 998-1011, doi:10.1016/j.tibs.2016.08.015 (2016).

11 Wang, Y., Xu, L. Y., Wang, S. Y., Ye, F. & Zhu, G. B. Global Distribution of Anaerobic Ammonia Oxidation (Anammox) Bacteria - Field Surveys in Wetland, Dryland, Groundwater Aquifer and Snow. *Frontiers in Microbiology* **10**, doi:10.3389/fmicb.2019.02583 (2019).

12 Zhang, L. & Okabe, S. Ecological niche differentiation among anammox bacteria. *Water Res* **171**, 115468, doi:10.1016/j.watres.2020.115468 (2020).

13 Sonthiphand, P., Hall, M. W. & Neufeld, J. D. Biogeography of anaerobic ammonia-oxidizing (anammox) bacteria. *Front Microbiol* **5**, 399, doi:10.3389/fmicb.2014.00399 (2014).

14 Shaw, D. R. *et al.* Extracellular electron transfer-dependent anaerobic oxidation of ammonium by anammox bacteria. *Nature Communications* **11**, doi:10.1038/s41467-020-16016-y (2020).

15 Ossa Ossa, F. *et al.* Limited oxygen production in the Mesoarchean ocean. *Proc Natl Acad Sci U S A* **116**, 6647-6652, doi:10.1073/pnas.1818762116 (2019).

16 Stueken, E. E., Buick, R., Guy, B. M. & Koehler, M. C. Isotopic evidence for biological nitrogen fixation by molybdenum-nitrogenase from 3.2 Gyr. *Nature* **520**, 666-669, doi:10.1038/nature14180 (2015).

17 Koehler, M. C., Buick, R. & Barley, M. E. Nitrogen isotope evidence for anoxic deep marine environments from the Mesoarchean Mosquito Creek Formation, Australia. *Precambrian Research* **320**, 281-290, doi:10.1016/j.precamres.2018.11.008 (2019).

18 Koehler, M. C., Buick, R., Kipp, M. A., Stueken, E. E. & Zaloumis, J. Transient surface ocean oxygenation recorded in the similar to 2.66-Ga Jeerinah Formation, Australia. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 7711-7716, doi:10.1073/pnas.1720820115 (2018).

19 Homann, M. *et al.* Microbial life and biogeochemical cycling on land 3,220 million years ago. *Nature Geoscience* **11**, 665-+, doi:10.1038/s41561-018-0190-9 (2018).

20 Godfrey, L. V. & Falkowski, P. G. The cycling and redox state of nitrogen in the Archaean ocean. *Nature Geoscience* **2**, 725-729, doi:10.1038/Ngeo633 (2009).

21 Garvin, J., Buick, R., Anbar, A. D., Arnold, G. L. & Kaufman, A. J. Isotopic Evidence for an Aerobic Nitrogen Cycle in the Latest Archean. *Science* **323**, 1045-1048, doi:10.1126/science.1165675 (2009).

22 Zerkle, A. L. *et al.* Onset of the aerobic nitrogen cycle during the Great Oxidation Event. *Nature* **542**, 465-+, doi:10.1038/nature20826 (2017).

23 Luo, G. M. *et al.* Nitrogen fixation sustained productivity in the wake of the Palaeoproterozoic Great Oxygenation Event. *Nature Communications* **9**, doi:10.1038/s41467-018-03361-2 (2018).

24 Kipp, M. A., Stueken, E. E., Yun, M., Bekker, A. & Buick, R. Pervasive aerobic nitrogen cycling in the surface ocean across the Paleoproterozoic Era. *Earth and Planetary Science Letters* **500**, 117-126, doi:10.1016/j.epsl.2018.08.007 (2018).

25 Lyons, T. W., Reinhard, C. T. & Planavsky, N. J. The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **506**, 307-315, doi:10.1038/nature13068 (2014).

26 Yang, J. *et al.* Ammonium availability in the Late Archaean nitrogen cycle. *Nature Geoscience* **12**, 553-557, doi:10.1038/s41561-019-0371-1 (2019).

27 Stueken, E. E., Kipp, M. A., Koehler, M. C. & Buick, R. The evolution of Earth's biogeochemical nitrogen cycle. *Earth-Science Reviews* **160**, 220-239, doi:10.1016/j.earscirev.2016.07.007 (2016).

28 Klotz, M. G. & Stein, L. Y. Nitrifier genomics and evolution of the nitrogen cycle. *Fems Microbiology Letters* **278**, 146-156, doi:10.1111/j.1574-6968.2007.00970.x (2008).

29 Hallmann, C., Kelly, A. E., Gupta, S. N. & Summons, R. E. Reconstructing Deep-Time Biology with Molecular Fossils. *Quantifying the Evolution of Early Life : Numerical Approaches to the Evaluation of Fossils and Ancient Ecosystems* **36**, 355-401, doi:10.1007/978-94-007-0680-4_14 (2011).

30 Moss, F. R. *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* **115**, 9098-9103, doi:10.1073/pnas.1810706115 (2018).

31 Mettam, C. *et al.* Anaerobic nitrogen cycling on a Neoarchaean ocean margin. *Earth and Planetary Science Letters* **527**, doi:10.1016/j.epsl.2019.115800 (2019).

32 Ader, M. *et al.* Interpretation of the nitrogen isotopic composition of Precambrian sedimentary rocks: Assumptions and perspectives. *Chemical Geology* **429**, 93-110, doi:10.1016/j.chemgeo.2016.02.010 (2016).

33 Fuchsman, C. A. & Stueken, E. E. Using modern low-oxygen marine ecosystems to understand the nitrogen cycle of the Paleo- and Mesoproterozoic oceans. *Environ Microbiol* **23**, 2801-2822, doi:10.1111/1462-2920.15220 (2021).

34 Yang, Z. & Rannala, B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol* **23**, 212-226, doi:10.1093/molbev/msj024 (2006).

35 Harhangi, H. R. *et al.* Hydrazine synthase, a unique phylomarker with which to study the presence and biodiversity of anammox bacteria. *Appl Environ Microbiol* **78**, 752-758, doi:10.1128/AEM.07113-11 (2012).

36 Betts, H. C. *et al.* Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat Ecol Evol* **2**, 1556-1562, doi:10.1038/s41559-018-0644-x (2018).

37 Roger, A. J. & Susko, E. Molecular clocks provide little information to date methanogenic Archaea. *Nat Ecol Evol* **2**, 1676-1677, doi:10.1038/s41559-018-0687-z (2018).

38 Wong, M. L., Charnay, B. D., Gao, P., Yung, Y. L. & Russell, M. J. Nitrogen Oxides in Early Earth's Atmosphere as Electron Acceptors for Life's Emergence. *Astrobiology* **17**, 975-983, doi:10.1089/ast.2016.1473 (2017).

39 Kasting, J. F. & Walker, J. C. G. Limits on Oxygen Concentration in the Prebiological Atmosphere and the Rate of Abiotic Fixation of Nitrogen. *Journal of Geophysical Research-Oceans* **86**, 1147-1158, doi:DOI 10.1029/JC086iC02p01147 (1981).

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

41 Brandes, J. A. *et al.* Abiotic nitrogen reduction on the early Earth. *Nature* **395**, 365-367, doi:Doi 10.1038/26450 (1998).

42 Summers, D. P. & Chang, S. Prebiotic Ammonia from Reduction of Nitrite by Iron(Ii) on the Early Earth. *Nature* **365**, 630-632, doi:DOI 10.1038/365630a0 (1993).

43 Lalonde, S. V. & Konhauser, K. O. Benthic perspective on Earth's oldest evidence for oxygenic photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 995-1000, doi:10.1073/pnas.1415718112 (2015).

44 Anbar, A. D. *et al.* A whiff of oxygen before the Great Oxidation Event? *Science* **317**, 1903-1906, doi:10.1126/science.1140325 (2007).

45 Ranjan, S., Todd, Z. R., Rimmer, P. B., Sasselov, D. D. & Babbin, A. R. Nitrogen Oxide Concentrations in Natural Waters on Early Earth. *Geochemistry Geophysics Geosystems* **20**, 2021-2039, doi:10.1029/2018gc008082 (2019).

46 Cheng, C. *et al.* Nitrogen isotope evidence for stepwise oxygenation of the ocean during the Great Oxidation Event. *Geochimica Et Cosmochimica Acta* **261**, 224-247, doi:10.1016/j.gca.2019.07.011 (2019).

47 Lehtovirta-Morley, L. E. Ammonia oxidation: Ecology, physiology, biochemistry and why they must all come together. *FEMS Microbiol Lett* **365**, doi:10.1093/femsle/fny058 (2018).

48 Li, W. *et al.* Metagenomics reveals microbial community differences lead to differential nitrate production in anammox reactors with differing nitrogen loading rates. *Water Res* **169**, 115279, doi:10.1016/j.watres.2019.115279 (2020).

49 Zhang, L. *et al.* Microbial competition among anammox bacteria in nitrite-limited bioreactors. *Water Res* **125**, 249-258, doi:10.1016/j.watres.2017.08.052 (2017).

50 Dalsgaard, T., Thamdrup, B., Farias, L. & Revsbech, N. P. Anammox and denitrification in the oxygen minimum zone of the eastern South Pacific. *Limnology and Oceanography* **57**, 1331-1346, doi:10.4319/lo.2012.57.5.1331 (2012).

51 Canfield, D. E., Glazer, A. N. & Falkowski, P. G. The evolution and future of Earth's nitrogen cycle. *Science* **330**, 192-196, doi:10.1126/science.1186120 (2010).

52 Zerkle, A. L. *et al.* Onset of the aerobic nitrogen cycle during the Great Oxidation Event. *Nature* **542**, 465-467, doi:10.1038/nature20826 (2017).

53 Ren, M. *et al.* Phylogenomics suggests oxygen availability as a driving force in Thaumarchaeota evolution. *ISME J* **13**, 2150-2161, doi:10.1038/s41396-019-0418-8 (2019).

54 Rattray, J. E. *et al.* A comparative genomics study of genetic products potentially encoding ladderane lipid biosynthesis. *Biol Direct* **4**, 8, doi:10.1186/1745-6150-4-8 (2009).

55 Van Vliet, A. H., Ketley, J. M., Park, S. F. & Penn, C. W. The role of iron in Campylobacter gene regulation, metabolism and oxidative stress defense. *FEMS Microbiol Rev* **26**, 173-186, doi:10.1016/s0168-6445(02)00095-5 (2002).

56 Johnson, L. A. & Hug, L. A. Distribution of reactive oxygen species defense mechanisms across domain bacteria. *Free Radical Biology and Medicine* **140**, 93-102, doi:10.1016/j.freeradbiomed.2019.03.032 (2019).

57 Moore, S. J. & Warren, M. J. The anaerobic biosynthesis of vitamin B12. *Biochem Soc Trans* **40**, 581-586, doi:10.1042/BST20120066 (2012).

58 Crowe, S. A. *et al.* Sulfate was a trace constituent of Archean seawater. *Science* **346**, 735-739, doi:10.1126/science.1258966 (2014).

59 Jamieson, J. W., Wing, B. A., Farquhar, J. & Hannington, M. D. Neoarchaean seawater sulphate concentrations from sulphur isotopes in massive sulphide ore. *Nature Geoscience* **6**, 61-64, doi:10.1038/Ngeo1647 (2013).

60 Blattler, C. L. *et al.* Two-billion-year-old evaporites capture Earth's great oxidation. *Science* **360**, 320-+, doi:10.1126/science.aar2687 (2018).

61 Fakhraee, M., Hancisse, O., Canfield, D. E., Crowe, S. A. & Katsev, S. Proterozoic seawater sulfate scarcity and the evolution of ocean-atmosphere chemistry. *Nature Geoscience* **12**, 375-380, doi:10.1038/s41561-019-0351-5 (2019).

62 Kartal, B., Keltjens, J. T. & Jetten, M. S. M. Metabolism and Genomics of Anammox Bacteria. *Nitrification*, 181-200 (2011).

63 Ito, M., Morino, M. & Krulwich, T. A. Mrp Antiporters Have Important Roles in Diverse Bacteria and Archaea. *Front Microbiol* **8**, 2325, doi:10.3389/fmicb.2017.02325 (2017).

64 Peeters, S. H. & van Niftrik, L. Trending topics and open questions in anaerobic ammonium oxidation. *Curr Opin Chem Biol* **49**, 45-52, doi:10.1016/j.cbpa.2018.09.022 (2019).

65 Akram, M. *et al.* A 192-heme electron transfer network in the hydrazine dehydrogenase complex. *Sci Adv* **5**, eaav4310, doi:10.1126/sciadv.aav4310 (2019).

66 Bi, Z., Qiao, S., Zhou, J., Tang, X. & Zhang, J. Fast start-up of Anammox process with appropriate ferrous iron concentration. *Bioresour Technol* **170**, 506-512, doi:10.1016/j.biortech.2014.07.106 (2014).

67 Poulton, S. W. & Canfield, D. E. Ferruginous Conditions: A Dominant Feature of the Ocean through Earth's History. *Elements* **7**, 107-112, doi:10.2113/gselements.7.2.107 (2011).

68 Ferousi, C. *et al.* Iron assimilation and utilization in anaerobic ammonium oxidizing bacteria. *Curr Opin Chem Biol* **37**, 129-136, doi:10.1016/j.cbpa.2017.03.009 (2017).

69 Keren, N., Aurora, R. & Pakrasi, H. B. Critical roles of bacterioferritins in iron storage and proliferation of cyanobacteria. *Plant Physiol* **135**, 1666-1673, doi:10.1104/pp.104.042770 (2004).

70 Poulton, S. W., Fralick, P. W. & Canfield, D. E. Spatial variability in oceanic redox structure 1.8 billion years ago. *Nature Geoscience* **3**, 486-490, doi:10.1038/Ngeo889 (2010).

71 Konhauser, K. O. *et al.* Iron formations: A global record of Neoarchaean to Palaeoproterozoic environmental history. *Earth-Science Reviews* **172**, 140-177, doi:10.1016/j.earscirev.2017.06.012 (2017).

72 Wu, Z. *et al.* Salinity-driven heterogeneity toward anammox distribution and growth kinetics. *Appl Microbiol Biotechnol* **103**, 1953-1960, doi:10.1007/s00253-018-9521-4 (2019).

73 Parsons, C., Stueken, E. E., Rosen, C. J., Mateos, K. & Anderson, R. E. Radiation of nitrogen-metabolizing enzymes across the tree of life tracks environmental transitions in Earth history. *Geobiology* **19**, 18-34, doi:10.1111/gbi.12419 (2021).

74 Ader, M. *et al.* Ocean redox structure across the Late Neoproterozoic Oxygenation Event: A nitrogen isotope perspective. *Earth and Planetary Science Letters* **396**, 1-13, doi:10.1016/j.epsl.2014.03.042 (2014).

75 Schopf, J. W. Geological evidence of oxygenic photosynthesis and the biotic response to the 2400-2200 ma "great oxidation event". *Biochemistry (Mosc)* **79**, 165-177, doi:10.1134/S0006297914030018 (2014).

76 Chen, S. C. *et al.* The Great Oxidation Event expanded the genetic repertoire of arsenic metabolism and cycling. *Proc Natl Acad Sci U S A* **117**, 10414-10421, doi:10.1073/pnas.2001063117 (2020).

77 Buick, R. Did the Proterozoic 'Canfield Ocean' cause a laughing gas greenhouse? *Geobiology* **5**, 97-100, doi:10.1111/j.1472-4669.2007.00110.x (2007).

78 Parks, D. H. *et al.* Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* **2**, 1533-1542, doi:10.1038/s41564-017-0012-7 (2017).

79 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).

80 Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973, doi:10.1093/bioinformatics/btp348 (2009).

81 Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268-274, doi:10.1093/molbev/msu300 (2015).

82 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587-589, doi:10.1038/nmeth.4285 (2017).

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

84 Yang, Z. H. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**, 1586-1591 (2007).

85 Wang, H.-C., Minh, B. Q., Susko, E. & Roger, A. J. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Systematic biology* **67**, 216-235 (2018).

86 Reis, M. D. *et al.* Using Phylogenomic Data to Explore the Effects of Relaxed Clocks and Calibration Strategies on Divergence Time Estimation: Primates as a Test Case. *Syst Biol* **67**, 594-615, doi:10.1093/sysbio/syy001 (2018).

87 Dos Reis, M., Donoghue, P. C. & Yang, Z. Bayesian molecular clock dating of species divergences in the genomics era. *Nat Rev Genet* **17**, 71-80, doi:10.1038/nrg.2015.8 (2016).

88 Dos Reis, M. & Yang, Z. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol Biol Evol* **28**, 2161-2172, doi:10.1093/molbev/msr045 (2011).

89 Quang, L. S., Gascuel, O. & Lartillot, N. Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* **24**, 2317-2323, doi:10.1093/bioinformatics/btn445 (2008).

90 Morel, B., Kozlov, A. M., Stamatakis, A. & Szollosi, G. J. GeneRax: A tool for Species-Tree-Aware Maximum Likelihood-based gene family tree inference under gene duplication, transfer, and loss. *Mol Biol Evol* **37**, 2763-2774, doi:10.1093/molbev/msaa141 (2020).

91 Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27-30 (2000).

92 Lu, S. *et al.* CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* **48**, D265-D268, doi:10.1093/nar/gkz991 (2020).

93 Mitchell, A. L. *et al.* InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* **47**, D351-D360, doi:10.1093/nar/gky1100 (2019).

94 El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res* **47**, D427-D432, doi:10.1093/nar/gky995 (2019).

95 Haft, D. H. *et al.* TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res* **29**, 41-43, doi:10.1093/nar/29.1.41 (2001).

96 Saier, M. H., Jr. *et al.* The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res* **44**, D372-D379, doi:10.1093/nar/gkv1103 (2016).

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 blue bars on the four calibration nodes and the last common ancestor (LCA) of anammox bacteria represent the posterior 95% highest probability density (HPD) interval of the posterior time estimates. 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 Ma. 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. [25]. The green arrow suggests the earliest evidence for aerobic nitrogen cycling at around 2.7

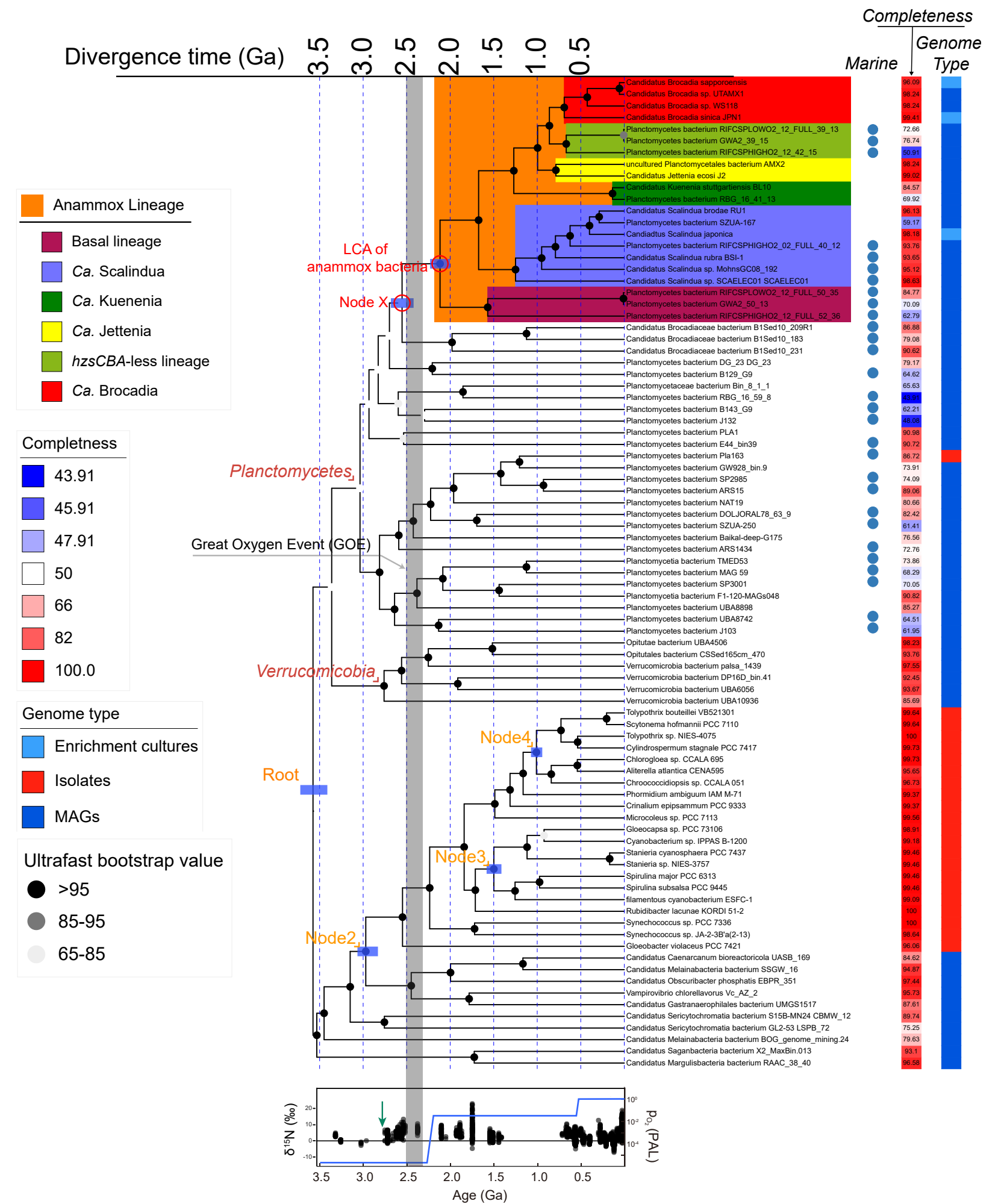
Ga. PAL on right axis means P_{O_2} relative to the present atmospheric level. The black dots denote the change of nitrogen $\delta^{15}N$ isotope values according to Kipp, et al. [24].

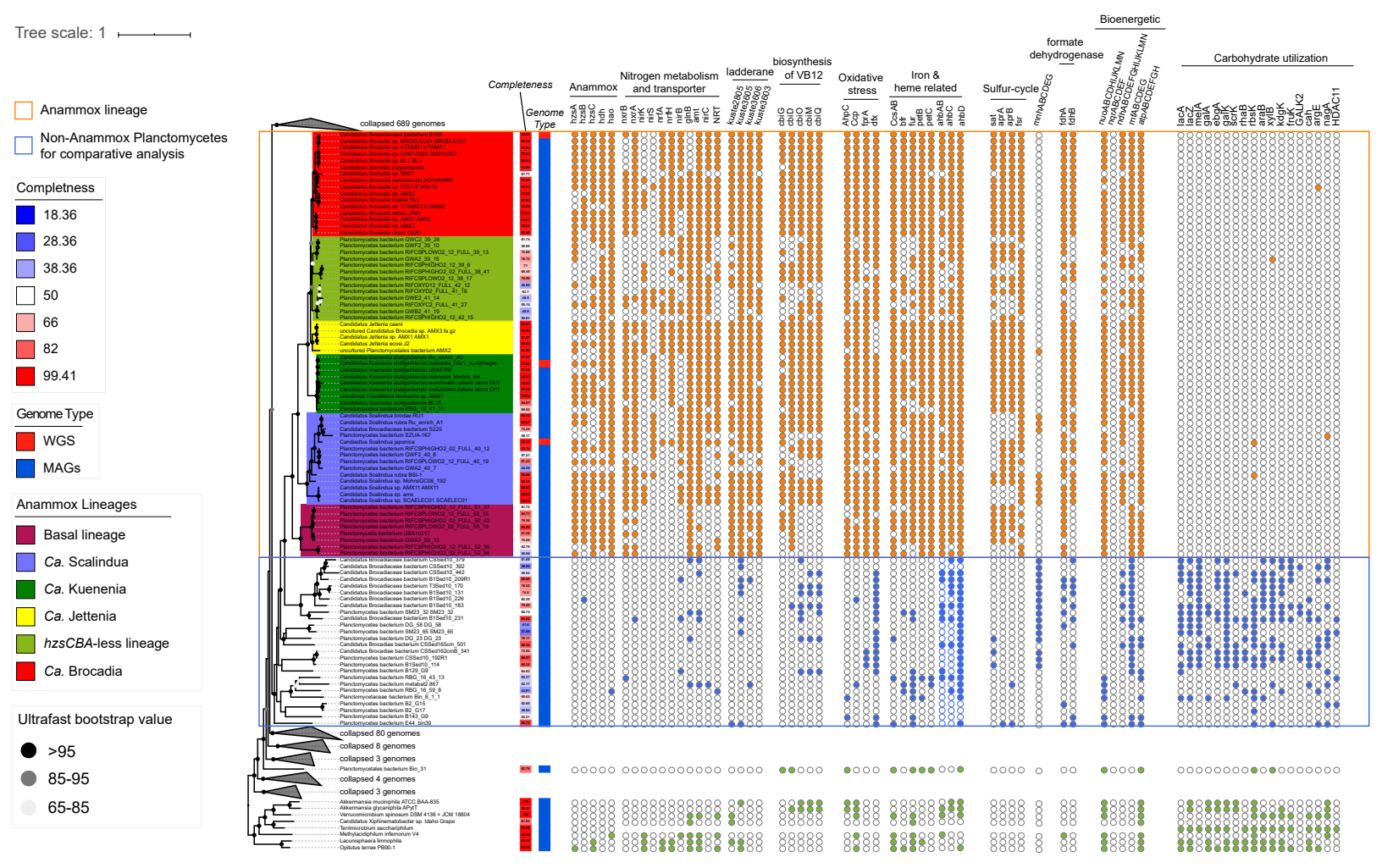
Figure 2. The phyletic pattern of ecologically relevant genes in the comparison between anammox bacteria and non-anammox bacteria. The phylogenomic tree on the left was constructed with 887 genomic sequences described in Supplementary Texts section 2.1. Solid circles at the nodes indicate the ultrafast 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. [54]; *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;

743 *cah*, cephalosporin-C deacetylase; *argE*, acetylornithine deacetylase; *nagA*;

744 N-acetylglucosamine-6-phosphate deacetylase; *HDAC11*, histone deacetylase 11.


Divergence time (Ga)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
Completeness
Marine
Genome Type
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
Enrichment cultures
Isolates
MAGs
Ultrafast bootstrap value
>95
85-95
65-85
LCA of anammox bacteria
Node X
Planctomycetes
Great Oxygen Event (GOE)
Verrucomicobia
Root
Node4
Node3
Node2
Candidatus Brocadia sapporoensis
Candidatus Brocadia sp. UTAMX1
Candidatus Brocadia sp. 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
Planctomycetes bacterium MAG 59
Planctomycetes bacterium SP3001
Planctomycetia bacterium F1-120-MAGs048
Planctomycetes bacterium UBA8898
Planctomycetes bacterium UBA8742
Planctomycetes bacterium J103
Opitutae bacterium UBA4506
Opitutales bacterium CSSed165cm_470
Verrucomicrobia bacterium palsa_1439
Verrucomicrobia bacterium DP16D_bin.41
Verrucomicrobia bacterium 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
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
PO2 (PAL)
10⁰
10⁻²
10⁻⁴
3.5
3.0
2.5
2.0
1.5
1.0
0.5
Age (Ga)



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