Title: Increased Replication Rates of Dissimilatory Nitrogen-Reducing Bacteria Leads to Decreased Anammox Reactor Performance

Authors: Ray Keren[1]*, Jennifer E. Lawrence[1]*, Weiqin Zhuang[1,2], David Jenkins[1], Jillian F. Banfield[3], Lisa Alvarez-Cohen[1], Lijie Zhou[1,4]+, Ke Yu[1,5]+

*These authors contributed equally to the work

+Corresponding authors

Affiliations

1- Department of Civil and Environmental Engineering, University of California Berkeley, Berkeley, CA, USA
2- Department of Civil and Environmental Engineering, The University of Auckland, Auckland, New Zealand
3- Earth and Planetary Sciences, University of California Berkeley, Berkeley, CA, USA
4- College of Chemistry and Environmental Engineering, Shenzhen University, Shenzhen, China
5- School of Environment & Energy, Peking University, Beijing, China

Abstract

Anaerobic ammonium oxidation (anammox) is a biological process employed to remove reactive nitrogen from wastewater. While a substantial body of literature describes the performance of anammox bioreactors under various operational conditions and perturbations, few studies have resolved the metabolic roles of community members. Here, we use metagenomics to study the microbial community within a laboratory-scale anammox bioreactor from inoculation, through performance destabilizations, to

stable steady-state. Metabolic analyses reveal that dissimilatory nitrogen reduction to ammonium (DNRA) is the primary nitrogen removal pathway that competes with anammox in the bioreactor. Increased replication rates of bacteria capable of DNRA leads to out-competition of annamox bacteria, which is the key source of fixed carbon, and the loss of reactor performance. Ultimately, our findings underline the importance of metabolic interdependencies related to carbon and nitrogen-cycling within anammox bioreactors and highlight the potentially detrimental effects of bacteria that are otherwise considered to be core community members.

Main Text

Anammox bacteria obtain energy from the conversion of ammonium and nitrite into molecular nitrogen gas[1]. The only currently known bacteria to catalyze this process are Planctomycetes[2,3], none of which have yet been isolated[3,4]. In practice, anammox bacteria are employed in combination with the partial nitritation (PN) process to remove ammonium from wastewaters or anaerobic digestor though side-streams. First, in PN, approximately half of the ammonium in solution is aerobically oxidized to nitrite. Second, in anammox, both ammonium and nitrite are anaerobically converted to N_2[5,6]. PN/anammox is beneficial because it consumes 60% less energy, produces 90% less biomass, and emits a significantly smaller volume of greenhouse gases than conventional nitrogen removal by nitrification and denitrification processes[7].

Even though over 100 full-scale PN/anammox processes have been installed across the globe at municipal and industrial wastewater treatment plants[8] anammox bacteria have very low growth rates within engineered environments and are easily inhibited by a variety of factors, including fluctuating substrate and metabolite concentrations[9,10]. Furthermore, recovery from an inhibition event can take up to six months, which is unacceptably long for municipalities who must meet strict nitrogen discharge limits[11]. These problems are compounded by what is currently only a cursory understanding of the

microbial communities responsible for stable, robust anammox performance. The broad application of PN/anammox in wastewater treatment processes requires a more comprehensive understanding of the complex interactions between the numerous bacterial species within the bioreactors.

Previous research suggests that a core microbial community exists within anammox bioreactors[12–16]. In the majority of anammox bioreactors, besides Planctomycetes, uncultured members of the phyla Bacteroidetes, Chloroflexi, Ignavibacteria, and Proteobacteria have been identified. Since these phyla have primarily been identified through 16S rRNA studies, their interplay on anammox performance have yet to be elucidated[12–16]. While anammox bacteria are responsible for the key nitrogen transformation processes, additional groups of anaerobic nitrogen-cycling bacteria cooperate to transform and remove nitrate, a product of anammox metabolism[17–19].

Here, we illuminate the metabolic relationships between the anammox bacteria and its supporting community members during the start-up and operation of a laboratory-scale anammox bioreactor. We used genome-centric metagenomics to recover 337 draft genomes from six time-points spanning 440 days of continuous bioreactor operation. Additionally, we combined our reconstruction of the microbial community's dynamic metabolic potential with bioreactor performance data and relative abundance profiles based on both metagenomic and 16S rRNA sequencing. As a result, we were able to identify core metabolic activities and potential interdependencies that inform the performance and stability of the anammox bioreactor. We found that certain metabolic interactions between anammox bacteria and associated community members may be responsible for the destabilization of anammox bioreactors. To our knowledge, this is the first time-series-based study to link anammox metagenomic insights and community composition to bioreactor functionality[20]. Our findings bolster the fundamental, community-level understanding of the anammox process. Ultimately, these results will enable better understanding of this important microbial process and a more comprehensive control of this promising technology that should help facilitate its widespread adoption at wastewater treatment plants.

Bioreactor performance. The performance of a laboratory-scale anaerobic membrane bioreactor (described in methods) was tracked for 440 days from initial inoculation, through several performance crashes, to stable and robust anammox activity (Figure 1). Performance was quantified in a variety of ways, including by its nitrogen removal rate (NRR, g-N L^{-1} d^{-1}). Bioreactor performance generally improved over the first 103 days of operation. At this point, the hydraulic residence time was reduced from 48 to 12 hours and influent concentrations were reduced to maintain a stable loading rate. Additional biomass from a nearby pilot-scale PN/anammox process was added on Day 145 and reactor performance improved enabling influent ammonium and nitrite concentrations to be steadily increased until the NRR approached 2 g-N L^{-1} d^{-1}. On Day 189 the bioreactor experienced a technical malfunction and subsequent performance crash, identified by a rapid decrease in the NRR and the effluent quality. On Day 203, the bioreactor was again amended with a concentrated stock of biomass and the NRR quickly recovered. Influent ammonium and nitrite concentrations were again increased until the NRR reached 2 g-N L^{-1} d^{-1}.

The bioreactor subsequently maintained steady performance for approximately 75 days, until Day 288, when effluent concentrations of ammonium and nitrite unexpectedly began to increase and nitrate concentrations disproportionately decreased. Seven days later, the NRR rapidly plummeted and had since no technical malfunctions had occurred this indicated that a destabilized microbial community may have been responsible for the performance crash. At that time, the cause of the performance decline was not understood, so the bioreactor was not re-seeded with biomass. After 50 days of limited performance, concentrations of copper, iron, molybdenum, and zinc in the bioreactor influent were increased[21–24] and the NRR rapidly recovered. Stable and robust bioreactor performance was subsequently maintained.

Metagenomic sequencing and binning. Whole community DNA was extracted and sequenced at six time-points throughout the study: Day 0 (D0), for inoculant composition; Day 82 (D82), during nascent, positive anammox activity; Day 166 (D166), three weeks after an additional biomass amendment; Day 284 (D284), after a long period of stable and robust anammox activity and just before destabilization; Day

328 (D328), in the midst of the destabilization period; and Day 437 (D437), during mature, stable, and robust anammox activity.

From all samples, 337 genomes were binned, 244 of which are estimated to be >70% complete. The genomes were further dereplicated across the six time-points into clusters at 95% average nucleotide identity (ANI). This resulted in 127 representative and unique genomes (Table 1), which were used for all downstream analyses. Mapping showed an average read recruitment of 76% to representative genomes (Table 2). The number of genomes present at each time-point (using threshold values of coverage > 1 and breadth > 0.5) ranged from 60 (D437) to 103 (D166). In addition, nine strains were detected that differed from the representative genome by 2% ANI (Supplemental Information, Supplemental Table 1). With the exception of the anammox bacterium, referred to at the genus level, all genomes are referred to at their phylum level.

Core anammox community. Resulting genomes from our study, in combination with genomes from two previous anammox metagenomic studies, Speth et al.[18] (22 genomes) and Lawson et al.[25] (15 genomes), provide strong evidence to support a core anammox community (Figure 2). The relative abundances of bacteria from the dominant phyla across these three reactors are fairly similar: in each reactor the anammox, along with Chloroflexi, Ignavibacteria, and Proteobacteria bacteria, compose >70% of the community (Figure 2B).

Due to the significantly larger genome yield and time-series analysis in this study, our metagenomes had more genomes in common with each of the other reactors than the other reactors shared between themselves. Nevertheless, three genomes were identified from bacteria that are closely related across all three reactors: *Brocadia* (responsible for anammox), an unclassified Chloroflexi, and an unclassified Ignavibacteria. All three of these genomes were present in our reactor during stable operation on D437, and two of them (*Brocadia* and an Ignavibacterium) are among the ten most abundant genomes at that time. In total, 21 genomes from our reactor are closely related to those from at least one of the two other reactors, 17 of which are present at D437 (Supplemental Table 2). The related bacteria account for

50% and 93% of the Speth et al. and Lawson et al. genomes, respectively. The reactor studied by Speth et al. was different from the other two reactors because it was amended with oxygen to perform partial nitritation and anammox within the same reactor, while the others performed anammox only.

A more focused phylogenetic tree of Planctomycetes shows that the *Brocadia* in our reactor and in the Lawson et al. reactor are the same species (*Brocadia sapporensis*[26]), while the *Brocadia* species from the Speth et al. reactor is different (*Brocadia sinica*) (Supplemental Figure 1).

Community dynamics. The relative abundances of organisms represented by genomes were calculated by multiplying genome coverage and breadth. Additional 16S rRNA gene sequencing, executed at 56 timepoints across the lifespan of the bioreactor, allowed us to expand our view of the relative abundances of bacteria over time. Thus, 38 of the 127 genomes contained 16S rRNA gene sequences that matched 16S rRNA gene sequencing efforts, and these 38 bacteria accounted for the majority of the bioreactor microbial community (Figure 3).

The *Brocadia* genus accounted for a small fraction of the bacteria in the inoculating biomass. Consistent with previous research of a combination PN/anammox bioreactor with oxygen amendment, members of the phyla Acidobacteria, Bacteroidetes, Ignavibacteriae, and Proteobacteria were also present. During the first 100 days of bioreactor operation, *Brocadia* increased in relative abundance. Its replication rate at D82 is high (Supplemental Table 3), which corroborates with its overall enrichment in the community. Following the reactor malperformance and biomass amendment on Day 147, the bioreactor became dominated by a bacterium represented by a single genome of the phylum Bacteroidetes (order Sphingobacteriales). The bacterium's calculated replication rate was low on D166, and over the next 100 days its relative abundance steadily declined. In contrast, the *Brocadia* replication rate was extremely high on D166 allowing it to once again dominate the microbial community. *Brocadia* remained dominant until Day 290, when the relative abundances of several Chloroflexi (most notably, one from the class Anaerolineae) and an Ignavibacteria dramatically increased. Shortly after this shift, the bioreactor experienced an unexplained period of performance decline and subsequent performance crash. During

this period the *Brocadia* replication rated dramatically declined, while the Chloroflexi replication rate increased (Supplemental Table 3). These shifts in replication rates six days before a response in relative abundance profiles and 12 days before a response in NRR are consistent with an instability in population dynamics having directly impacted the reactor performance.

The relative abundances of *Brocadia* and the Chloroflexi, as well as their replication rates, remained fairly constant over the next 44 days. After the influent media trace metal concentrations were increased, the relative abundance of Chloroflexi decreased and that of *Brocadia* increased. By D437, *Brocadia* again dominated the reactor, and the replication rates of *Brocadia* and Chloroflexi bacteria became similar.

Community grouping. Because both internal and external factors can work in combination to affect the structure of a bioreactor community we hypothesize that there are groups of bacteria (or sub-communities) associated with different phases of the reactor lifespan. To test for grouping, all of the genomes were cross-correlated (Figure 4A). The resulting heatmap revealed four distinct clusters (A-D) that are highly correlated. Cluster A was the largest, with 52 genomes, while Clusters B-D had 25, 24, and 26 genomes respectively.

To better examine the clustering of the genomes in relation to the different time-points, we ran nonmetric multidimensional scaling (nMDS) analyses on the abundance data (Figure 4B). The nMDS projection shows how genome groups are strongly associated with specific time-points. Group A are associated with the inoculant source biomass at D0 and D166 while Group C are associated with the nascent anammox community at D82. Group B are associated with the times of destabilized anammox performance (Days 284-328), and Group D are associated with the mature, stable anammox community at D437. *Brocadia* is part of Group D, although its location on the nMDS projection is skewed to the left, because of its high abundance throughout most of the experiment. Group A dominates at Days 0 and 166, but is highly reduced in other times (Figure 4C). Group B dominates D328, while maintaining a similar abundance in all other time points. Group C is mostly unique to D82 although a few of its members

remain in the reactor after the crash at low abundance. Group D bacteria show little change up to D284 (except a spike at D82), after which they increase in abundance.

It is interesting to note that the nascent anammox community is different from that of the destabilized and the mature anammox communities. Because the nascent anammox community was supplemented by a source inoculant biomass amendment, we cannot resolve a linear trajectory for the microbial community between the initial and final states. B and D groups, while distinct, share many similarities, and the majority of the genomes associated with group B were still present in the reactor on D437.

For all subsequent analyses, we split the genomes into two groups: those that are associated with the mature anammox community at D437 (Anammox Associated, AA, nMDS groups B and D), and those that are not (Source Associated, SA, nMDS groups A and C). The AA community includes all of the genomes that are present at D437 while the SA community includes the rest of the genomes that are not present at D437. Some of these genomes are associated with the sludge amendments, and some are associated with the nascent anammox community; at no point is there a community exclusively comprised of SA genomes. The relative abundance of each group with time aligns with the previous analysis (Figure 4C). Since the anammox bacterium dominates the community from D82 and onwards, it was removed from group D for the purpose of this comparison.

Metabolic profiles. For the purpose of analyzing the metabolic potential of the microbial community we evaluated only genomes with > 70% completeness (n = 88). Using Hidden Markov Model (HMM) searches of the KEGG database, we checked for the presence of genes (with KO number) and calculated KEGG module completeness[27,28]. The genomes were clustered by KO presence/absence (Supplemental Figure 2) and their module completeness (Figure 5). The clustering by the two methods resulted in similar groupings.

The module clustering resolved five groups (α, β, γ, δ, ε) (Figure 5A). Groups α and β contain more anammox-associated genomes (90% and 60% respectively) while groups γ, δ, and ε contain 65%,

70% and 60% of source-associated genomes. The taxonomy of the bacteria also strongly influenced the clustering (Figure 5B). Group α is composed solely of Gram (+) bacteria, while Group β is composed of Microgenomates (CPR bacteria). Group ɣ is composed entirely of Gram (-) bacteria (including *Brocadia*), Group δ is composed of Ignavibacteria and Bacteroidetes (other members of these phyla were clustered in Group ɣ). Only the Ignavibacteria of Group δ are associated with the AA group, so further analysis only included those. Group ε was composed of Proteobacteria.

Based on the KEGG module clustering, we reconstructed the representative metabolisms of the groups (Figure 6). We used a module completeness threshold of 67% per genome, and considered it representative if it was complete in >50% of its members. Group δ is not represented since it diverged from group ɣ by auxotrophies in several modules (Figure 5A, red rectangle). The *Brocadia* metabolism is shown in Supplemental Figure 3.

While module completeness was used for most of the analyses, in several cases it was not sufficient (e.g., overlap between modules, no module for path). In the cases of oxidative phosphorylation, fermentation, carbon fixation, several amino acid synthesis pathways, and nitrogen metabolism we analyzed gene presence manually.

Nitrogen cycling. We evaluated the quality genomes for the presence of all genes related to nitrogen metabolism appearing in KEGG (Figure 7). Four additional HMMs were added for anammox genes (hydrazine synthase subunit A (hzsA), hydrazine oxidoreductase subunit A (hzoA)), and nitrification (nitrite oxidoreductase subunits nrxA and B)[29]. For the latter, the similarity of the gene to the nitrate reductase narGH was taken into consideration.

With the exception of two CPR, all of the genomes in the reactor contained genes encoding assimilation of ammonia into glutamate (Figure 7A). More than half (49) of the bacteria could reduce nitrate, and the same number could further reduce nitrite to nitrogen monoxide (NO), however only 26 bacteria could do both steps. The most common gene encoding for nitrate reduction is narGH; niK is more common than nirS (36 and 19 occurrences, respectively). The remaining steps of denitrification are

encoded in a smaller number of genomes. The nrxAB gene was only identified in two genomes, one of which was *Brocadia*

One-step DNRA is identified in 22 genomes, predominantly with nrfAH. While ammonia assimilation and nitrate reduction are fairly similar in the AA and SA bacteria, DNRA is more common in AA and denitrification beyond nitrite in the SA genomes (Figure 7C).

Bacteria could improve reactor performance if they remove nitrate (nitrate reducers) and excess nitrite, but they could be detrimental if they compete with anammox for nitrite (DNRA and denitrification from nitrite). To check for changes in the abundance of these groups, we classified bacteria by the presence of genes encoding for DNRA, denitrification or nitrate reduction (Figure 7B). Some bacteria classified as denitrifiers or DNRA also encode nitrate reduction. A few genomes in the D0 sample encoded both denitrification and DNRA, but their abundances were negligible. The anammox bacterium has genes required for DNRA but, given the overall reactor performance, was expected to be primarily performing anammox for energy generation. DNRA could potentially be used by the anammox bacteria for detoxification by cycling potentially toxic excess nitrite back to ammonium where it could then participate in the anammox reactions[18,25].

In the inoculant source community, the nitrate reducers were the most dominant group (38%), with similar amounts of denitrifiers and DNRA (26% and 25% respectively). The abundance of anammox was consistent with the reactor performance (Figure 1). The denitrifying group of bacteria decreased in relative abundance to 8% at around D284. On the other hand, bacteria capable of DNRA were relatively abundant throughout the reactor start up. Most notably, these bacteria dominated the reactor during its destabilization, reaching 48% at D328, compared to 23% for the anammox bacteria. An increase of bacteria capable of DNRA is consistent with the reactor performance data which showed a decline in the amount of ammonium consumed relative to overall reactor performance. At this time period the following four DNRA bacteria were highly abundant: (anamox2_sub_Ignavibacterium_album_33_16_curated, anamox1_Bacteria_56_37_curated, LAC_NA06_sub_Chloroflexi_61_22_curated, and LAC_NA06_sub_Chloroflexi_59_14). Three of the four are group B bacteria, and one is group D. All

four bacteria show an increase in relative abundance between D284 and D328. Three of the four also had increased replication rates just before the onset of the crash, as mentioned above. The two other abundant bacteria (apart from the anammox bacterium) are LAC_NA07_Bacteria_70_305_curated (nitrate reducer) and LAC_NA07_Burkholderiales_70_312_curated (denitrifier). The former is among the most abundant when the community is not SA dominated, while the latter is always one of three most abundant bacteria. These 7 bacteria constitute >75% of the community at D328.

Carbon fixation. Several bacteria (n =12) in the community are potentially capable of carbon fixation, via the Wood-Ljungdahl pathway or the Calvin cycle. *Brocadia* was confirmed as a primary producer, fixing carbon via the Wood-Ljungdahl pathway using energy from the anammox pathway. All other bacteria had genes for reduction of nitrogen compounds. To confirm that these bacteria are likely autotrophs, we checked for genes conferring the ability to use inorganic electron donors. Three of these bacteria had no potential electron donor and therefore were classified as heterotrophs. The remainder had genes for oxidizing sulfide or hydrogen, and were classified as potential autotrophs. Of these nitrate reducing bacteria (n = 8), only one was relatively abundant after D166, and increased in abundance between D284 and D328.

LAC_NA07_Burkholderiales_70_312_curated, can fix carbon by the Calvin cycle, is a denitrifier, and can possibly oxidize sulfide to sulfite (dsrAB are present in the genome). This bacterium is among the most abundant at all time-points; it increased significantly in abundance between D284 and D328 and the increase continued to D437. However, the replication rate of the bacterium decreased from D166 onwards (Supplemental Table 3), so it is not likely competing with or destabilizing the anammox bacterium.

Electron transfer. Apart from nitrogen reduction, another common anaerobic respiration pathway was acetate fermentation (genes detected in 60% of the genomes). This process was much more common in AA (69%) bacteria than in SA (51%) bacteria. Ni-Fe Hydrogenase was present in 31% of the genomes,

but was most common among the Chloroflexi of group α (87% and 48% of all occurrences of hydrogenases found) (Figure 6B).

The majority of bacteria in the reactor are potentially facultative aerobes (58%). All have high affinity complex IV, which differed between AA and SA bacteria. In the AA bacteria, the bd type is found in all aerobic members of group α (one also has a cbb3 type) and the Ignavibacteria, and the cbb3 type occurs mostly in Proteobacteria. For the SA bacteria, the cbb3 type is found in 24/25 aerobes and the bd-type is only found in 6/25 (only in one bacterium it is the sole variant). Complex III, which is also essential to aerobic respiration, was only found in 14 Proteobacteria, one Actinobacterium, and one Chloroflexi. It is possible that other bacteria have an alternative Complex III[30] that cannot be found by current KEGG annotations. Complexes I/II are found in nearly all of the bacteria, except CPR. Only five bacteria lack the F-type ATPase; two have the V-type ATPase instead.

Central carbon metabolism. It is likely that nearly all bacteria (98%) can oxidize sugar by glycolysis (Figure 6A and E), while fewer bacteria (69%) have the pentose phosphate pathway (PPP). Acetyl-CoA could be synthesized from pyruvate (90% general, 98% AA, and 81% SA), or by beta-oxidation (49% general, 57% AA, and 43% SA). The majority of bacteria had the full TCA cycle (84%, or 88% after excluding CPR). A possible major carbon source for the bacteria in the reactor are amino acids (aa.), with 95% being able to incorporate aa. into their central carbon metabolism. The most common aa. (aspartate) can be converted into oxaloacetate and fed into the TCA cycle. Three aa. (serine, alanine, and cysteine) can be converted into pyruvate. Of these, only cysteine is unidirectional, so aa., as a carbon source, cannot be ascertained. Group α has additional genes that support a reliance on proteins for their metabolism (Figure 6B). They also have a set of peptidases, as well as multiple transporters covering all forms of aa., peptides, and polyamines.

Some metabolic groups can use aa. as precursors for synthesis of other metabolites. Glutamate and histidine can be converted to PRPP, and with glutamine to pyrimidines (Figure 6A). Groups ɣ and ε can use aspartate to synthesize NAD^+, and glutamine to synthesis IMP (Figure 6C). NAD^+ and IMP

cannot be synthesized by all of the bacteria, indicating that there are potential metabolic interdependencies in the community. Members of group ε (Figure 6D) can use leucine as a precursor to acetyl-CoA, lysine for acetoacetyl-CoA, glutamate for glutathione, and chorismate for ubiquinone. The last two are only synthesized by group ε, indicating additional potential metabolic interdependencies in the community.

Comparing AA and SA. To examine why certain bacteria were enriched in the reactor while others were removed, we compared the synthesis of metabolites to the utilization of nutrients in the reactor. For synthesis we checked 24 KEGG modules for aa., 16 modules for vitamins or cofactors, and 11 modules for lipids and fatty acids. For nutrient utilization we looked at 52 modules of transporters. A difference larger than 10% in the ratio of bacteria with or without a complete module was considered relevant.

In all synthesis categories, SA bacteria had higher completeness ratios in the majority of the modules investigated (14 of 24 aa. modules, 13 of 16 vitamins and cofactor modules, and 8 of 11 lipid and fatty acid modules). The transportation modules show an opposite trend, with 38 of 50 modules having higher completeness ratios in AA bacteria.

This comparison shows that the selective driver in the anammox community is the ability of the bacteria to acquire nutrients from the environment, rather than the ability to synthesize them. The larger ratio of bacteria with auxotrophies in the AA bacteria hints of a greater reliance on other members of the community.

Metabolic interdependencies. The bacteria in the AA community maintain a complex metabolic system. In the mature functioning reactor, the anammox bacterium is (almost) the only primary producer present. It is also the only bacterium capable of synthesizing vitamin B12. For most other metabolites (vitamins, and cofactors) the possible metabolic interdependencies[31] are less straightforward (Supplemental table 4). Seven of 20 aa. can be synthesized by the majority of all metabolic groups (Figures 5-6). Members of group δ have the largest set of auxotrophies, lacking the genes conferring the ability to synthesize eight

aa. The four other aa. can be synthesized by most group ε members and a few of the group ɣ members. Only a single cofactor (CoA) is commonly synthesized by all groups. Group α has auxotrophies for most other cofactors, with the exception of pimeloyl-ACP (in 45% of members). Most other vitamins and cofactors are commonly synthesized by only a single group, usually group ε. With lipids and fatty acids, many modules are irrelevant to compare since group α differs from all other groups as a solely Gram (+) bacteria, while the rest are all Gram (-). However even after the Gram (-) specific modules are excluded, group α still has multiple auxotrophies. The anammox bacterium and group ɣ also have few commonly complete modules.

When combining all of the above data, we found that groups ɣ and ε both had mutualistic associations with *Brocadia* (Figure 8). Group ε potentially provides more metabolites to *Brocadia* than it receives whereas groups α and δ are seem to gain more from *Brocadia* than they provide. Interestingly, four members of group α and one member of group δ were identified as the possible cause of the destabilization.

By the end of the experiment (D437) when reactor performance had stabilized, members of group α are the second most abundant group after *Brocadia* The ten most abundant bacteria at this point included four members of group δ and three members of group ε. Comparing these relative abundances to bacterial abundances during lowest reactor performance (D328) we find that *Brocadia* and group ε are reduced in abundance by about 50%, while groups α and δ are increased by 70% and 100% respectively.

Discussion

In this study we present an in-depth analysis of the development of an anammox community from seed to stable state (through several perturbations) in an anaerobic membrane bioreactor. By combining several methodologies, we were able to gain important insights into the dynamics and interactions of more than 100 species in the reactor community.

Previous studies have discussed a potential core anammox community[12–16]. With the exception of very few studies, all such work has been conducted with single gene markers. Our metagenomic analysis of an anammox community is the largest to-date and thus expands the ability to test this hypothesis. Our results support the existence of a core community, while identifying factors that differentiate communities. The high similarity between bacteria originating from three distinct anammox reactors[18,24] strongly suggests a global core anammox microbial community. In the construction of the phylogenetic tree we used >3000 reference genomes originating from diverse environments. Even with the sheer number and diversity of sources, the anammox community formed distinct clades at the species level. More than half of the bacteria did not have species level relatives, and an additional 26% only had a relative found in our anammox reactors or those from previous studies[18,24]. Together, nearly 80% of the bacterial are unique to anammox reactors so that it is clear that the anammox reactor selects for a unique set of bacteria. Parameters that increased the differences between communities were the species of the anammox bacterium and the reactor configuration. Since both parameters relate to the same reactor[18], we cannot conclude which would have a stronger effect.

We identified several potential bacterial destabilizers of the anammox process. Analysis of replication rate days prior to the destabilization event revealed that these bacteria increased their replication rate, while *Brocadia* nearly ceased replication. These results imply a causative nature to the change. Genes conferring DNRA capability were detected in these bacteria, which would allow them to compete with *Brocadia* for nitrite. This supposition is consistent with the reactor performance which exhibited decreased nitrogen removal and increased ammonium in the effluent during this period. The dominating bacteria during reactor malperformance were heterotrophs. In full-scale anammox reactors, where influent organic carbon is essentially ubiquitous, heterotrophic dominance could continue indefinitely without some sort of active countermeasure. Therefore, future research should target the inhibition of potential destabilizing heterotrophs.

A broader investigation of metabolic interdependencies within the community sheds light on the stability of the anammox community. *Brocadia* is the source of organic material in the community, but

obtains essential metabolites from some members, especially Proteobacteria. This forms a basis for a mutual symbiotic relationship. On the other hand, Chloroflexi, the largest group of bacteria besides *Brocadia*, receive numerous metabolites while apparently providing few in return. They are characterized by an array of extracellular proteases and amylases, likely used to breakdown the matrix formed by *Brocadia*. Chloroflexi as a group are most associated with anammox bacteria, and form a large fraction of the core community. They also account for the majority of the destabilizing bacteria. Together the results point to a parasitic symbiosis. Further investigation into these relations is warranted.

While anammox generates sufficient organic carbon to support the growth of its co-occurring heterotrophic microorganisms the tipping point between stable and unstable operation and the factors that control it have not been fully identified. Input changes may be able to restore anammox activity but this is just an empirical solution. Our findings improve the understanding of nitrogen-cycling within an anammox bioreactor and advance the comprehensive control of this promising technology. However further work is needed to elucidate the precise mechanisms that control community interactions.

Methods

Bioreactor operation. A laboratory-scale, anaerobic membrane bioreactor with a working volume of 1L was constructed and operated for over 440 days (Supplemental Figure 4). The bioreactor was originally inoculated with approximately 2 g VSS L^{-1} of biomass from a pilot-scale deammonification process treating sidestream effluent at San Francisco Public Utilities Commission (SFPUC) in San Francisco, CA. The bioreactor was re-inoculated with similar concentrations of biomass from the same source on Days 147 and 203. Synthetic media containing ammonium, nitrite, bicarbonate, and trace nutrients (meant to mimic sidestream effluent at a municipal wastewater treatment plant) was fed to the bioreactor (Supplemental Table 4). For the first 154 days of operation, the bioreactor was kept under nitrite-limiting conditions to prevent inhibitory conditions, and influent ammonium and nitrite concentrations ranged

from 200-300 mg N L^{-1} and 100-300 mg N L^{-1}, respectively. On Day 154, ammonium and nitrite concentrations were adjusted to the anammox stoichiometric ratio, 1:1.32. Afterwards, influent ammonium and nitrite concentrations were maintained at this ratio and ranged from 200-500 mg N L^{-1} and 265-660 mg N L^{-1}, respectively. On Day 353, influent concentrations of copper, iron, molybdenum, and zinc were increased based on literature suggestions[21–24].

The bioreactor was operated in a continuous flow mode. For the first 145 days, the hydraulic retention time (HRT) was maintained at 48 hours; afterwards it was reduced to 12 hours. No solids were removed from the bioreactor for the first 100 days of operation; afterwards, the solids retention time (SRT) was reduced to 50 days. A polyvinylidene fluoride hollow fiber membrane module with a 0.4 µm pore size and total surface area of 260 cm^2 (Litree Company, China) was mounted in the bioreactor. Temperature was maintained at 37° C with an electric heating blanket (Eppendorf, Hauppauge, NY). Mixing was provided by an impeller at a rate of 200 rpm. Mixed gas was supplied continuously to the bioreactor ($Ar:CO_2$ = 95:5; 50 mL min^{-1}) to eliminate dissolved oxygen and maintain pH at 7.2. Influent and effluent concentrations of ammonium, nitrite, and nitrate were measured approximately every other day using HACH test kits (HACH, Loveland, CO), as described in the manufacturer's methods 10031, 10019, and 10020, respectively.

Biomass collection and DNA extraction. Biomass samples were extracted via syringe from the bioreactor every 2-10 days, flash frozen in liquid nitrogen, and stored frozen at -80 ℃ until use. Genomic DNA was extracted from the samples using the DNeasy PowerSoil Kit (Qiagen, Carlsbad, CA), as described in the manufacturer's protocol. The concentration and purity of extracted DNA was measured with a NanoDrop Spectrophotometer (Thermo Scientific, Waltham, MA). The concentration of genomic DNA in all samples was normalized to 10 ng/µL with nuclease-free water (Thermo Scientific, Waltham, MA). All genomic DNA samples were stored at -20 ℃ until use.

Metagenomic sequencing, assembly, and binning. Genomic DNA samples from six time-points were sent to the Joint Genome Institute (JGI) in Walnut Creek, CA for sequencing on the Illumina HiSeq 2500 1T sequencer (Illumina, San Diego, CA). Resulting sequences from each time-point were processed separately, following the ggKbase SOP (https://ggkbase-help.berkeley.edu/overview/data-preparation-metagenome/). In summary, Illumina adapters and trace contaminants were removed (BBTools, GJI) and raw sequences were quality-trimmed with Sickle[32]. Paired-end reads were assembled using IDBA_UD with the pre-correction option and default settings[33]. For coverage calculations, reads were mapped with bowtie2[34]. Genes were predicted by Prodigal[35] and predicted protein sequences were annotated using usearch[36] against KEGG, UniRef100, and UniProt databases. The 16S rRNA gene and tRNA prediction was done with an in-house script and tRNAscanSE[37] respectively. At this point, the processed data was uploaded to ggKbase for binning.

Manual binning was performed using the ggKbase tool. The binning parameters for binning were GC% and coverage (CV) distribution, and phylogeny of the scaffolds. Quality of the manual bins was assessed by the number of Bacterial Single Copy Genes (BSCG) and Ribosomal Proteins (RP) found in each bin (aiming at finding the full set of genes, while minimizing the multiple copies). In addition to manual binning, automated binning was done, using four binners;ABAWACA1[38], ABAWACA2, CONCOCT[39], and Maxbin2[40]. For all, the default parameters were chosen.

All bins from both automatic and manual binning tools were input into DASTool[41] to iterate through bins from all binning tools and choose the optimal set of bins. checkM was run to analyze genome completeness[42]. The scaffold-to-bin file created by DASTool was uploaded back to ggKbase and all scaffolds were rebinned to match the DASTool output. Each of the new bins were manually inspected and scaffolds that were suspected to be falsely binned were removed.

After we inspected the first round of binning, we decided to improve the high coverage bins, by subsampling the read file, followed by the same SOP as above[43]. I addition, refinement of the Brocadia Genome bins was done with ESOMs[44] (Supplemental methods).

Post binning analysis. Unique representative genomes were determined by the dereplication tool, dRep[45], using a 95% threshold for species level clustering. Within each cluster, the representative genome was chosen based on their completeness, length, N50, contamination, and strain heterogeneity. In several clusters with higher heterogeneity, a second strain was chosen (Supplemental Table 1). The strain threshold was set at 2% difference (but lower than 5%).

All the representative and strain genomes were curated by correcting scaffolding errors introduced by idba_ud, using the ra2.py program[38]. Following curation, the genomes were processed again for gene calling and annotation (see above for details). Analysis of replication rates at different time-points was performed with the iRep program[46] using the default parameters.

Both raw reads and genomes were submitted to the National Center for Biotechnology Information (NCBI) Genbank, under project accession number PRJNA511011. In addition, the Representative and strains genomes were uploaded to ggkbase as two separate projects (https://ggkbase.berkeley.edu/LAC_reactor_startup/organisms and https://ggkbase.berkeley.edu/LAC_reactor_strains/organisms).

Phylogenetic analysis and core anammox analysis. The taxonomic affiliation of each genome was initially assigned in ggKbase. This was based on the taxonomic annotation of genes in the scaffolds. For each hierarchical taxonomic level, the taxonomy was decided if at least 50% of genes had a known taxonomic identification.

Phylogenetic analysis of the genomes (current study, Speth et al.[18], and Lawson et al.[25]) was based on a set of 15 ribosomal proteins[47]. Each gene was aligned separately to a set of 3225 reference genomes, followed by concatenation while keeping the aligned length of each gene intact. A preliminary tree was created by adding the queried genomes to the reference tree using pplacer v1.1.alpha19[48] and a set of in-house scripts. The tree was uploaded to iTOL[49] for visualization and editing. After initial inspection we decided to reduce the tree in preparation of creating a maximum likelihood tree. Large phyla with no representatives in an anammox sample were removed (approximately 1000 sequences). The

remaining sequences were aligned by MUSCLE[50] and a RAxML tree built in The CIPRES Science Gateway V. 3.3[50,51].

For the analysis of phylogenetic distance between different anammox community members, we used the APE package[52] in R[53,54] to extract the distance matrix. Species level distance was set at 5% of the longest measured distance on the tree.

16S rRNA gene sequencing, processing, and analysis. DNA samples, taken at 56 timepoints across the lifespan of the bioreactor, were sent to the Institute for Environmental Genomics at the University of Oklahoma (Norman, OK) for amplification of the variable 4 (V4) region of the 16S rRNA gene, library preparation, and amplicon sequencing. The full protocol was previously described in Wu et al. (Wu 2015). In summary, the V4 region of the bacterial 16S rRNA gene was amplified from DNA samples using primers 515F (5'-GTGCCAGCMGCCGCGG-3') and 806R (3'-TAATCTWTGGVHCATCAG-5'), with barcodes attached to the reverse primer. Amplicons were pooled at equal molality and purified with the QIAquick Gel Extraction Kit (QIAGEN Sciences, Germantown, MD). Paired-end sequencing was then performed on the barcoded, purified amplicons with the Illumina MiSeq sequencer (Illumina, San Diego, CA).

Subsequent sequence processing and data analysis were performed in-house using MOTHUR v.1.39.5, following the MiSeq SOP[55,56]. In summary, sequences were demultiplexed, merged, trimmed, and quality filtered. Unique sequences were aligned against the SILVA 16S rRNA gene reference alignment database[57]. Sequences that did not align to the position of the forward primer were discarded. Chimeras were detected and removed. Remaining sequences were clustered into operational taxonomic units (OTUs) within a 97% similarity threshold using the Phylip-formatted distance matrix. Representative sequences from each OTU were assigned taxonomic identities from the SILVA gene reference alignment database[57]. Sequences that were not classified as bacteria were removed. Remaining OTUs were counted, and the 137 most abundant OTUs (accounting for up to 99% of sequence reads within individual samples) were transferred to Microsoft Excel (Microsoft Office Professional Plus 2016)

for downstream interpretation and visualization. The 137 most abundant OTUs were up loaded to figshare (https://figshare.com/account/projects/59324/articles/7640396).

In order to correlate genome-based OTUs to 16S rRNA gene-based OTUs, 16S rRNA sequences were extracted from the representative genomes and combined with the representative sequences from the 137 most abundant 16S rRNA gene-based OTUs. If a representative genome did not contain the V4 region of the 16S rRNA gene, the region was pulled from another genome in the same cluster. The combined 16S rRNA sequences were aligned following the protocol described above, and those sharing at least 99% average nucleotide identity were assumed to represent the same microorganism (Evans 2006, Blast).

Community dynamics analysis. The paired sequence reads from all time-points were mapped to the set of reference genomes using bowtie2[34], followed by calculation of coverage (average number of reads mapped per nucleotide) and breadth (% of genome that was covered by at least one read in the mapping), for each genome per time-point[58]. The multiplication of the two values were then used to calculate the estimated abundance. This was done to negate biases that can be created repetitive sequences that more often occurs in very partial genome bins (i.e. only the repetitive sequences associated with the genome bin are found in a given time-point).

Association between genomes was tested by calculating pairwise correlation for all genomes by abundance. The Rho values (ranging from -1 to 1) were used to create a distance table (Euclidean distance), followed by clustering with the ward.D method. The resulting clusters were marked A-D. To test the association of genomes and clusters to time-points, we run a nMDS analysis (non-parametric MultiDimensional Scaling) with the genomes and the time-point. Each genome was colored by its abundance cluster on the 2D projection of the nMDS.

For relative abundance changes, the estimated abundance of genomes was divided by the sum of all estimated abundance values per time-point. For a more clear resolution of changes in the four abundance groups, the *Brocadia* (part of group D) was presented separately.

Metabolic analysis. The functional profiles of the genomes were evaluated using KEGG KAAS[59], with Hidden Markov Models for shared KEGG orthologies (KOs)[27,28,60]. From this, we received the KEGG annotation (KO number) for all open reading frames and a completeness value for each KEGG module. KO annotations that were questionable were removed from analysis.

From the KO list we created a presence absence matrix (Jaccard index), and clustered using the Complete method. From module completeness we created a Euclidean distance matrix, followed by clustering with the ward.D method. Based on module completeness clustering we assigned genomes to metabolic groups α-ε.

For each metabolic group a representative metabolic map was created. A module completeness greater than 67% in at least half of the group members, was considered as representative of the group. Once the modules were selected they were drawn and connected based on metabolic KEGG maps. Additional reaction, complexes and transporters were added according to KO presence (e.g a.a synthesis, oxidative phosphorylation complexes, flagellar motor, etc.).

For nitrogen metabolism all relevant KOs were examined. For the purpose of this study, nitrate reduction was consider as a separate path than denitrification/DNRA, since it could be the first step in both, using the same enzymes. Denitrifying bacteria were considered as bacteria capable of full conversion of nitrite to N_2. DNRA bacteria were considered as bacteria capable of conversion of nitrite to ammonium using the nrfAH enzymes. No partial nitrogen process in considered for this paper, although it is present, according to per step analysis.

Acknowledgements

This research was supported by the National Science Foundation through the Engineering Research Center for ReInventing the Nation's Water Infrastructure (ReNUWIt) ECC-1028962. This material is also based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1106400. Any opinion, findings, and conclusions or recommendations expressed in this

material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Author contribution

K.Y. and L.Z. supervised the study. L.Z., L.A-C., D.J., and K.Y designed the study. L.Z. built the reactor. R.K. analyzed metagenomics data and wrote the manuscript. J.L. analyzed 16S rRNA data, analyzed reactor performance, and wrote the manuscript. J.F.B supervised the metagenomics analysis. W.Z. contributed to reactor maintenance and analysis, sampling and 16S rRNA data analysis. All authors read the manuscript and contributed with inputs.

References

1. Mulder, A., Graaf, A. A., Robertson, L. A. & Kuenen, J. G. Anaerobic ammonium oxidation discovered in a denitrifying fluidized bed reactor. *FEMS Microbiol. Ecol.* **16**, 177–184 (1995).
2. Kuenen, J. G. Anammox bacteria: from discovery to application. *Nat. Rev. Microbiol.* **6**, 320–326 (2008).
3. Sonthiphand, P., Hall, M. W. & Neufeld, J. D. Biogeography of anaerobic ammonia-oxidizing (anammox) bacteria. *Front. Microbiol.* **5**, (2014).
4. Connan, R. *et al.* Batch enrichment of anammox bacteria and study of the underlying microbial community dynamics. *Chem. Eng. J.* **297**, 217–228 (2016).
5. Strous, M. *et al.* Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440**, 790–794 (2006).
6. Kartal, B. *et al.* Molecular mechanism of anaerobic ammonium oxidation. *Nature* **479**, 127–130 (2011).
7. Paques. Anammox sustainable nitrogen removal. Available at: https://en.paques.nl/mediadepot/1818a31cd232/WEBbrochureAnammox.pdf. (Accessed: 26th

September 2018)

8. Lackner, S. *et al.* Full-scale partial nitritation/anammox experiences – An application survey. *Water Res.* **55**, 292–303 (2014).

9. Ali, M. & Okabe, S. Anammox-based technologies for nitrogen removal: Advances in process start-up and remaining issues. *Chemosphere* **141**, 144–153 (2015).

10. Jin, R.-C., Yang, G.-F., Yu, J.-J. & Zheng, P. The inhibition of the Anammox process: A review. *Chem. Eng. J.* **197**, 67–79 (2012).

11. Li, X., Klaus, S., Bott, C. & He, Z. Status, Challenges, and Perspectives of Mainstream Nitritation-Anammox for Wastewater Treatment. *Water Environ. Res.* **90**, 634–649 (2018).

12. Li, X.-R. *et al.* The bacterial diversity in an anaerobic ammonium-oxidizing (anammox) reactor community. *Syst. Appl. Microbiol.* **32**, 278–289 (2009).

13. Gonzalez-Martinez, A. *et al.* Bacterial community structure of a lab-scale anammox membrane bioreactor. *Biotechnol. Prog.* **31**, 186–193 (2015).

14. Gonzalez-Martinez, A. *et al.* Comparison of bacterial diversity in full scale anammox bioreactors operated under different conditions. *Biotechnol. Prog.* **31**, 1464–1472 (2015).

15. Gonzalez-Martinez, A. *et al.* Microbial community analysis of a full-scale DEMON bioreactor. *Bioprocess Biosyst. Eng.* **38**, 499–508 (2015).

16. Pereira, A. D., Cabezas, A., Etchebehere, C., de Lemos Chernicharo, C. A. & de Araújo, J. C. Microbial communities in anammox reactors: a review. *Environmental Technology Reviews* **6**, 74–93 (2017).

17. Bagchi, S., Lamendella, R., Strutt, S., Van Loosdrecht, M. C. M. & Saikaly, P. E. Metatranscriptomics reveals the molecular mechanism of large granule formation in granular anammox reactor. *Sci. Rep.* **6**, 28327 (2016).

18. Speth, D. R., in ’t, M., Guerrero-Cruz, S., Dutilh, B. E. & Jetten, M. S. M. Genome-based microbial ecology of anammox granules in a full-scale wastewater treatment system. *Nat. Commun.* **7**, 11172 (2016).

19. Castro-Barros, C. M., Jia, M., van Loosdrecht, M. C. M., Volcke, E. I. P. & Winkler, M. K. H. Evaluating the potential for dissimilatory nitrate reduction by anammox bacteria for municipal wastewater treatment. *Bioresour. Technol.* **233**, 363–372 (2017).
20. Tang, X., Guo, Y., Jiang, B. & Liu, S. Metagenomic approaches to understanding bacterial communication during the anammox reactor start-up. *Water Res.* **136**, 95–103 (2018).
21. van de Graaf, A. A. *et al.* Anaerobic oxidation of ammonium is a biologically mediated process. *Appl. Environ. Microbiol.* **61**, 1246–1251 (1995).
22. Chen, H., Yu, J.-J., Jia, X.-Y. & Jin, R.-C. Enhancement of anammox performance by Cu(II), Ni(II) and Fe(III) supplementation. *Chemosphere* **117**, 610–616 (2014).
23. 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 (2014).
24. Zhang, X. *et al.* Impacts of the heavy metals Cu (II), Zn (II) and Fe (II) on an Anammox system treating synthetic wastewater in low ammonia nitrogen and low temperature: Fe (II) makes a difference. *Sci. Total Environ.* **648**, 798–804 (2018).
25. Lawson, C. E. *et al.* Metabolic network analysis reveals microbial community interactions in anammox granules. *Nat. Commun.* **8**, 15416 (2017).
26. Narita, Y. *et al.* Enrichment and physiological characterization of an anaerobic ammonium-oxidizing bacterium ‘ Candidatus Brocadia sapporoensis’. *Syst. Appl. Microbiol.* **40**, 448–457 (2017).
27. Burstein, D. *et al.* Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.* **7**, 10613 (2016).
28. Anantharaman, K. *et al.* Analysis of five complete genome sequences for members of the class Peribacteria in the recently recognized Peregrinibacteria bacterial phylum. *PeerJ* **4**, e1607 (2016).
29. Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).
30. Sun, C. *et al.* Structure of the alternative complex III in a supercomplex with cytochrome oxidase. *Nature* **557**, 123–126 (2018).

31. Zomorrodi, A. R. & Segrè, D. Genome-driven evolutionary game theory helps understand the rise of metabolic interdependencies in microbial communities. *Nat. Commun.* **8**, 1563 (2017).

32. Joshi NA, F. J. N. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle. (2011).

33. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).

34. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

35. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

36. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).

37. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).

38. Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211 (2015).

39. Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014).

40. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2015).

41. Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation, and scoring strategy. (2017). doi:10.1101/107789

42. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

43. Hug, L. A. *et al.* Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. *Environ. Microbiol.* **18**, 159–173 (2016).

44. Ultsch, A. & Mörchen, F. *ESOM-maps: Tools for Clustering, Visualization, and Classification with Emergent SOM*. (2005).

45. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).

46. Brown, C. T., Olm, M. R., Thomas, B. C. & Banfield, J. F. Measurement of bacterial replication rates in microbial communities. *Nat. Biotechnol.* **34**, 1256–1263 (2016).

47. Hug, L. A. *et al.* A new view of the tree of life. *Nat Microbiol* **1**, 16048 (2016).

48. Matsen, F. A., Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11**, 538 (2010).

49. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–5 (2016).

50. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

51. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. in *2010 Gateway Computing Environments Workshop (GCE)* (2010). doi:10.1109/gce.2010.5676129

52. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).

53. *RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA.*

54. R: The R Project for Statistical Computing. Available at: http://www.R-project.org. (Accessed: 26th September 2018)

55. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported

software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).

56. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).
57. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007).
58. Olm, M. R. *et al.* Identical bacterial populations colonize premature infant gut, skin, and oral microbiomes and exhibit different in situ growth rates. *Genome Res.* **27**, 601–612 (2017).
59. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–5 (2007).
60. Finn, R. D. *et al.* HMMER web server: 2015 update. *Nucleic Acids Res.* **43**, W30–8 (2015).

Figures

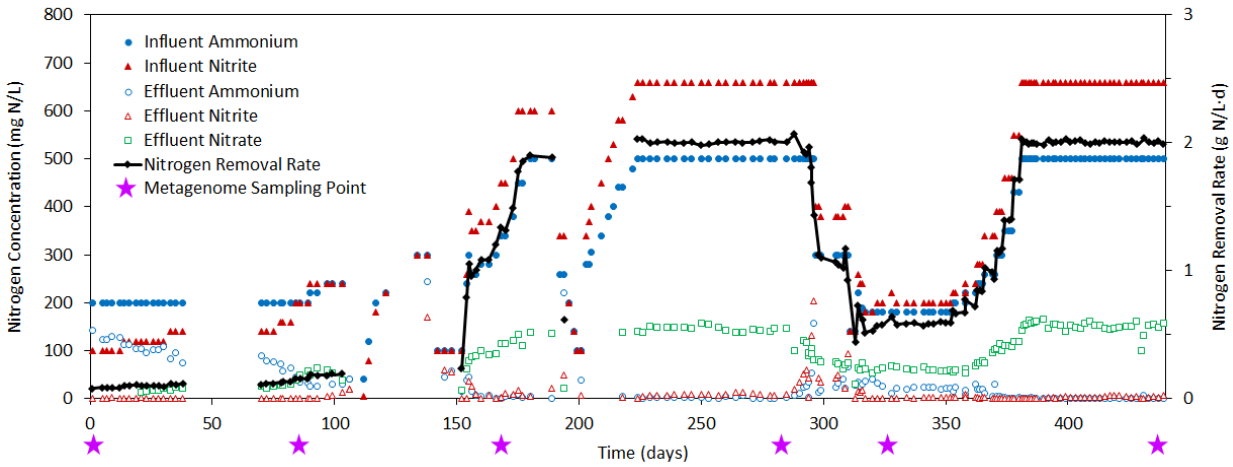
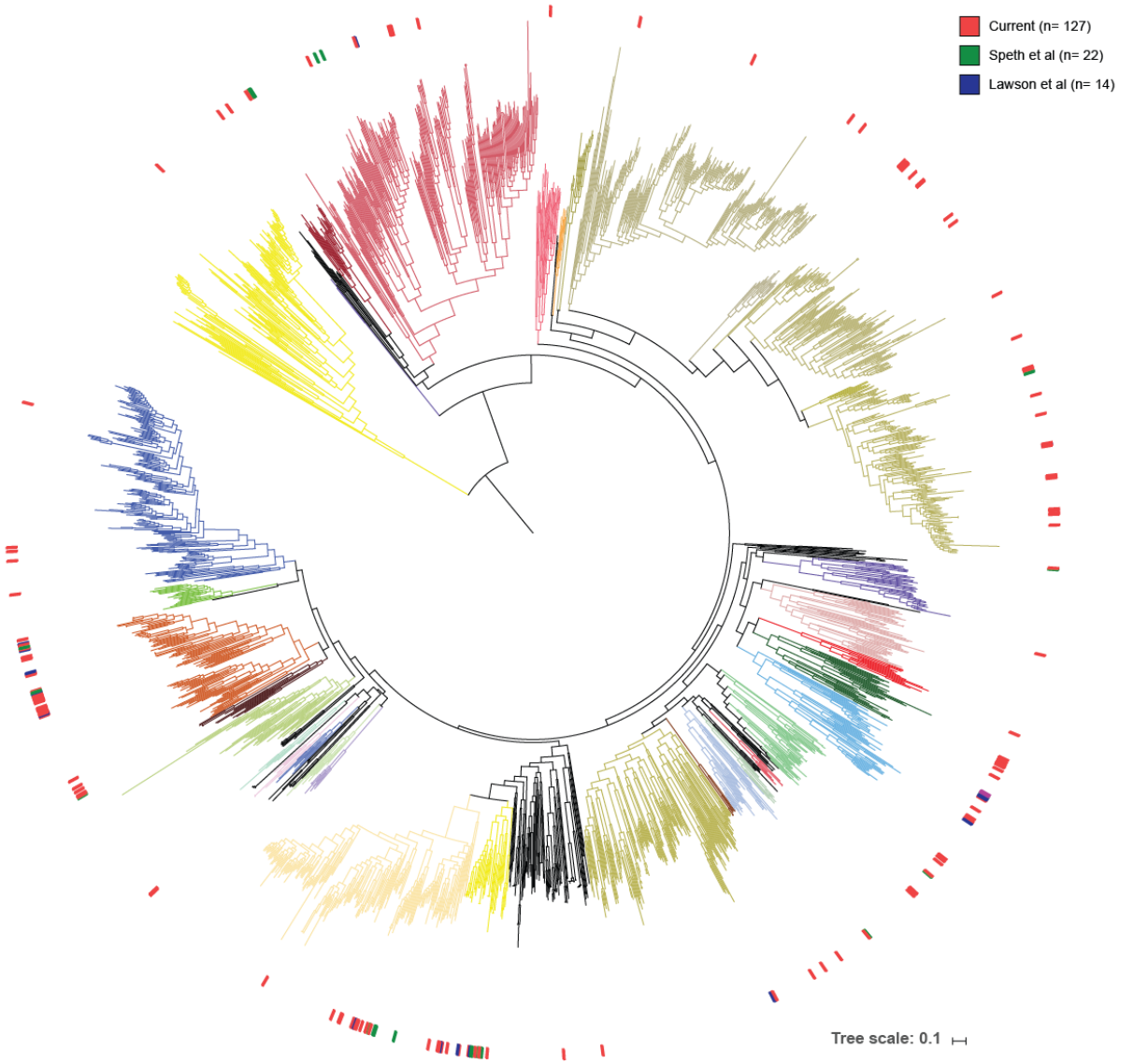
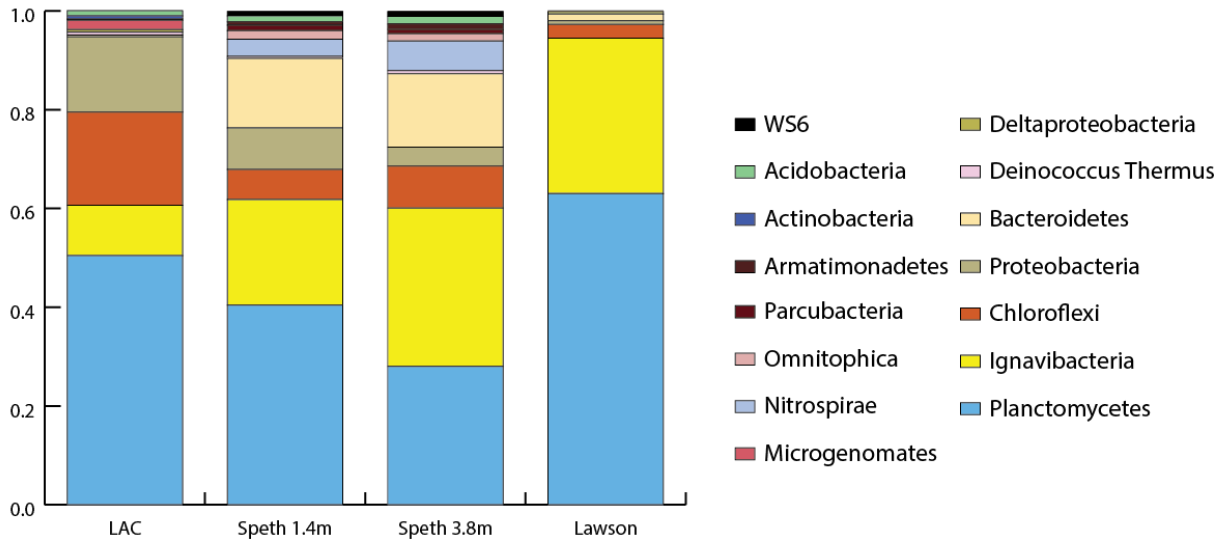


Figure 1 | Performance of the anaerobic membrane bioreactor. Influent and effluent concentrations of ammonium, nitrite, and nitrate (all as N) (primary y-axis) within the anaerobic membrane bioreactor performing anammox monitored over a period of 440 days. The influent did not contain nitrate, so influent nitrate is not plotted. The nitrogen removal rate (NRR), is plotted against the secondary y-axis. Sampling time points for metagenomes are indicated with purple stars below the x-axis.

A

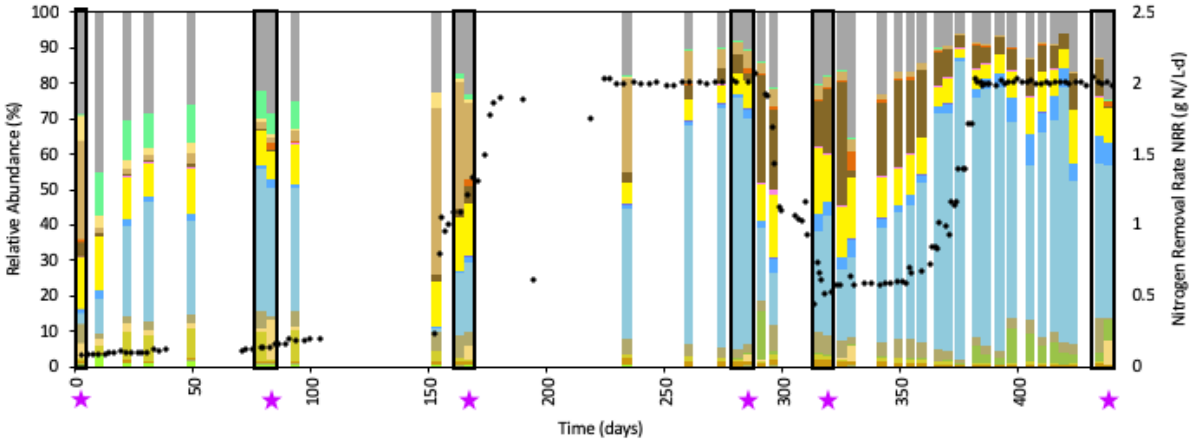

Current (n= 127)
Speth et al (n= 22)
Lawson et al (n= 14)
Tree scale: 0.1

B


1.0
0.8
0.6
0.4
0.2
0.0
LAC
Speth 1.4m
Speth 3.8m
Lawson
WS6
Acidobacteria
Actinobacteria
Armatimonadetes
Parcubacteria
Omnitophica
Nitrospirae
Microgenomates
Deltaproteobacteria
Deinococcus Thermus
Bacteroidetes
Proteobacteria
Chloroflexi
Ignavibacteria
Planctomycetes

730

Figure 2 | Phylogenetic analysis of three anammox microbial communities. (A) A maximum likelihood tree based on the alignment of 15 concatenated ribosomal proteins. In the construction of the tree 3225 reference sequences were used, with genomes from current and previous genome-centric studies on anammox communities. Genomes from the current anammox community are marked with a red dash, genomes from two previously studied communities; Speth et al. and Lawson et al., are marked with green and blue dashes respectively. (B) relative abundance of major phyla in the three microbial communities. Current community reference data was calculated from day 437 only. The relative abundance *Brocadia* sp. comprises nearly all of the abundance attributed to phylum Planctomycetes (with small contribution from other members of the phylum). The most abundant phyla (Chloroflexi, Ignavibacteria, and Proteobacteria) consistently account for >70% of the communities. The phyla colors follow the ggkbase color scheme and the major phyla are shown in the legend.



Phylum	Class	Order	Genus
Unmatched			
Acidobacteria	-	-	-
Bacteroidetes	-	-	-
Bacteroidetes	Sphingobacteria	Sphingobacteriales	-
Chloroflexi	-	-	-
Chloroflexi	Anaerolineae	-	-
Deinococcus-Thermus	Deinococci	Deinococcales	Truepera
Ignavibacteriae	Ignavibacteria	-	-
Nitrospirae	Nitrospira	Nitrospirales	Nitrospira
Planctomycetes	-	-	-
Planctomycetes	Planctomycetia	Brocadiales	Brocadia
Proteobacteria	-	-	-
Proteobacteria	Alphaproteobacteria	-	-
Proteobacteria	Betaproteobacteria	Burkholderiales	-
Proteobacteria	Betaproteobacteria	Rhodocyclales	-
Proteobacteria	Gammaproteobacteria	-	-
Proteobacteria	Gammaproteobacteria	Xanthomonadales	-
Verrucomicrobia	-	-	-

Figure 3 | Relative abundances of bacterial taxa over the lifespan of the bioreactor. Relative abundances of bacterial taxa, as identified by metagenomic and 16S rRNA gene sequencing, are plotted against the primary y-axis. Results derived from metagenomic sequencing are indicated with a purple star below the x-axis; all remaining results are derived from 16S rRNA gene sequencing. For visual clarity, sequencing results falling within three days of each other have been merged. “Unmatched” includes the OTUs and genomes that were not able to be matched across the two sequencing platforms. The similar relative abundance profiles at shared time points across metagenomic and 16S rRNA gene sequencing platforms (highlighted in the black boxes) provided us with the confidence to extrapolate high-resolution relative abundance profiles of our representative genomes from our 16S rRNA gene sequencing efforts.

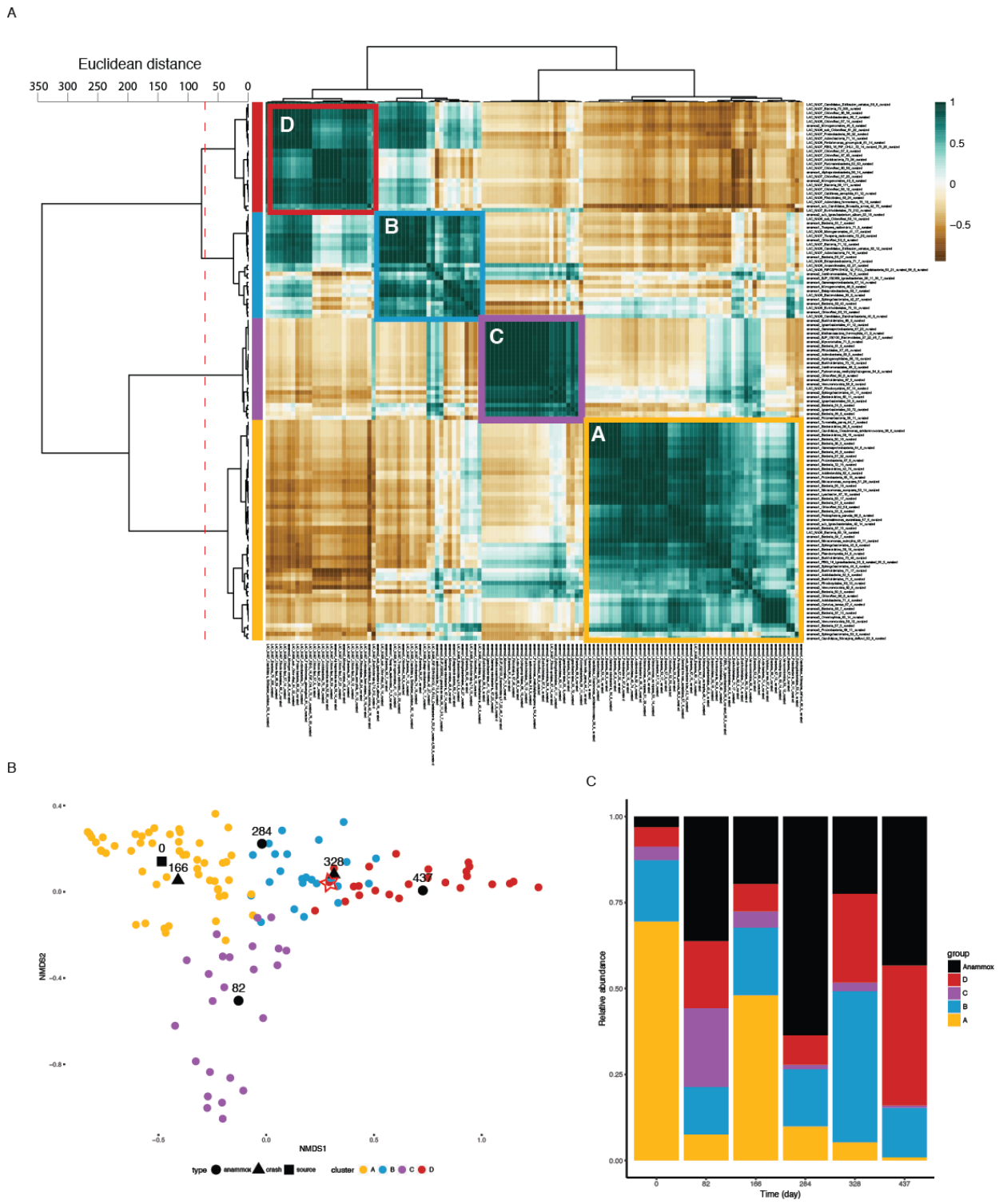
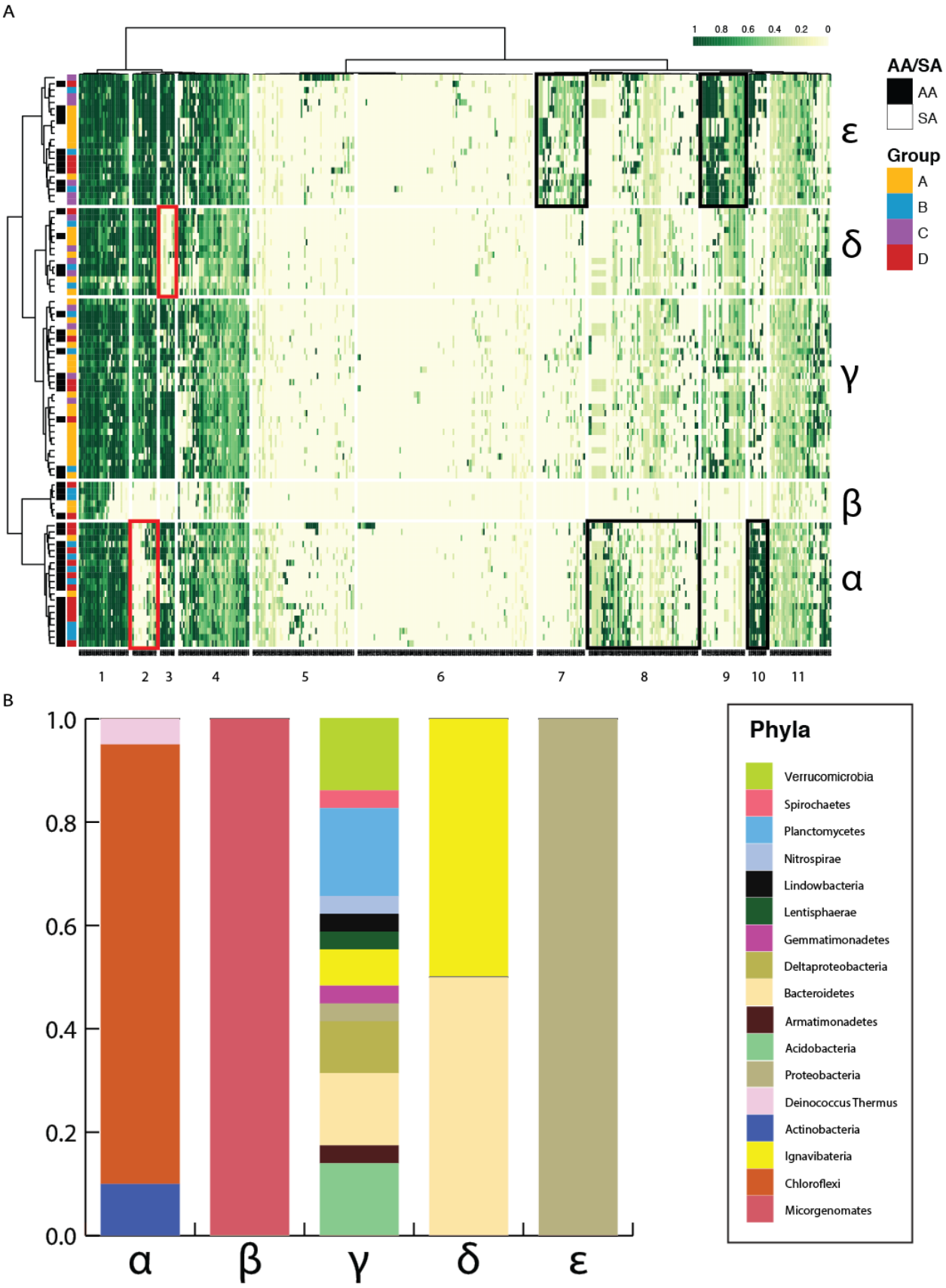


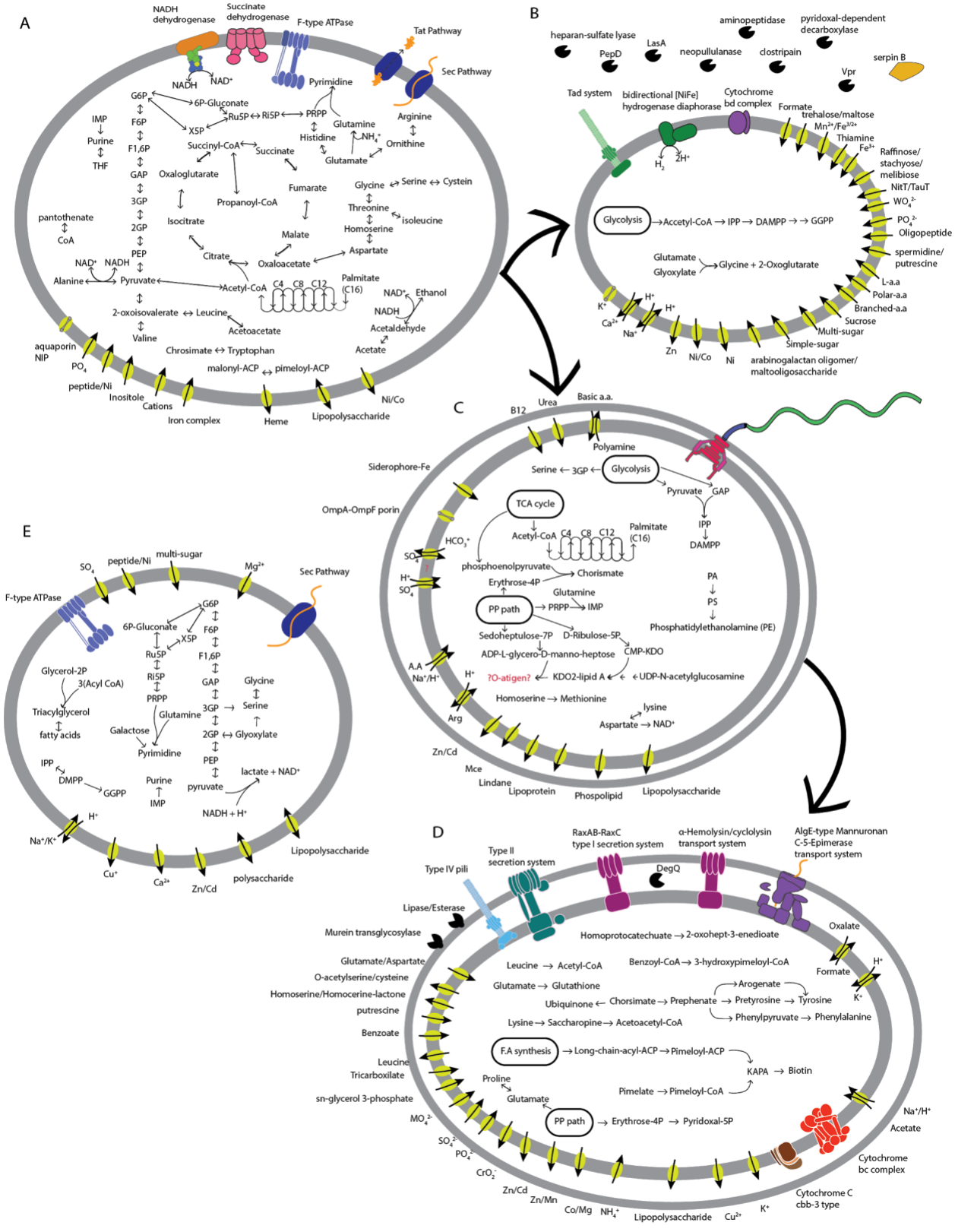
Figure 4| Analysis of the reactor community dynamics by the estimated abundance of the bacteria. (A) clustering heatmap of bacteria based on pairwise cross correlations in the six time points (matrix values are Rho values). Color scale mark high positive correlation in green and high negative correlation in brown. The row and column dendrograms are identical. Row dendrogram shows the calculated distance between the clusters with a dashed red line marking the threshold distance for clustering. Three columns

of annotations between the row dendrogram and heatmap; Core- labels bacteria related to previously studied bacteria (see relevant section); AA/SA- association of bacteria with either mature anammox or source inoculum; Group- assigned group based on correlation and clustering. (B) Two dimensional nMDS projection of bacteria and time points, showing the association of the bacteria (and abundance groups to certain time points). Each colored dot represents the centroid of a bacterium, with colors matching the abundance group. The location of Brocadia is marked with a red star. (C) Relative abundance of Groups A-D by time points. The anammox bacterium (of group D) is presented apart from the group to more easily show changes in the other group members.


A
1 0.8 0.6 0.4 0.2 0
AA/SA
AA
SA
Group
A
B
C
D
ε
δ
γ
β
α
1 2 3 4 5 6 7 8 9 10 11
B
1.0
0.8
0.6
0.4
0.2
0.0
α β γ δ ε
Phyla
Verrucomicrobia
Spirochaetes
Planctomycetes
Nitrospirae
Lindowbacteria
Lentisphaerae
Gemmatimonadetes
Deltaproteobacteria
Bacteroidetes
Armatimonadetes
Acidobacteria
Proteobacteria
Deinococcus Thermus
Actinobacteria
Ignavibateria
Chloroflexi
Micorgenomates

793

Figure 5| Metabolic profiling of bacterial community based on KEGG module completeness. (A) heatmap showing the reciprocal clustering of genomes (rows) and KEGG modules (columns). Heatmap is based on a Euclidean distance matrix and clustering with the ward.D method. Genome clustering resulted in 5 clusters (groups α-ε). Module clustering resulted in 11 clusters (blocks 1-11). Black rectangles on heatmap show module blocks that have increased completeness in a group of bacteria (compared to the others), and red rectangles show decreased completeness. The three columns on the left of the heatmap denote core association, AA/SA division, and abundance grouping respectively. B) Relative abundance by phyla of members in the metabolic clusters.


A
NADH dehydrogenase
Succinate dehydrogenase
F-type ATPase
Tat Pathway
Sec Pathway
NADH
NAD+
Pyrimidine
G6P
6P-Gluconate
Ru5P
RiSP
PRPP
Glutamine
Arginine
X5P
Histidine
NH4+
Ornithine
IMP
Purine
THF
F6P
F1,6P
Succinyl-CoA
Succinate
Glutamate
GAP
Oxaloglutarate
Fumarate
Glycine
Serine
Cystein
3GP
Propanoyl-CoA
Threonine
isoleucine
pantothenate
CoA
Isocitrate
Malate
Homoserine
2GP
PEP
Citrate
Oxaloacetate
Aspartate
NAD+
NADH
Alanine
Pyruvate
Acetyl-CoA
C4
C8
C12
Palmitate (C16)
NAD+
Ethanol
NADH
2-oxoisovalerate
Leucine
Acetoacetate
Acetaldehyde
Valine
Chrosimate
Tryptophan
Acetate
aquaporin NIP
PO4
malonyl-ACP
pimeloyl-ACP
peptide/Ni
Inositole
Cations
Iron complex
Heme
Lipopolysaccharide
Ni/Co
B
heparan-sulfate lyase
PepD
LasA
neopullulanase
aminopeptidase
clostripain
pyridoxal-dependent decarboxylase
serpin B
Vpr
Tad system
bidirectional [NiFe] hydrogenase diaphorase
Cytochrome bd complex
Formate
trehalose/maltose
Mn2+/Fe3/2+
Thiamine
Fe3+
Raffinose/ stachyose/ melibiose
NitT/TauT
WO4 2-
PO4 2-
Oligopeptide
spermidine/ putrescine
L-a.a
Polar-a.a
Branched-a.a
Sucrose
Multi-sugar
Simple-sugar
arabinogalactan oligomer/ maltooligosaccharide
Ni
Ni/Co
Zn
H2
2H+
Glycolysis
Accetyl-CoA
IPP
DAMPP
GGPP
Glutamate
Glyoxylate
Glycine + 2-Oxoglutarate
K+
Ca2+
Na+
H+
C
B12
Urea
Basic a.a.
Polyamine
Siderophore-Fe
Serine
3GP
Glycolysis
Pyruvate
GAP
OmpA-OmpF porin
TCA cycle
Acetyl-CoA
C4
C8
C12
Palmitate (C16)
IPP
DAMPP
HCO3+
SO4
H+
SO4
phosphoenolpyruvate
Chorismate
Erythrose-4P
PA
PS
Glutamine
PP path
PRPP
IMP
Phosphatidylethanolamine (PE)
Sedoheptulose-7P
D-Ribulose-5P
ADP-L-glycero-D-manno-heptose
CMP-KDO
A.A
Na+/H+
H+
?O-atigen?
KDO2-lipid A
UDP-N-acetylglucosamine
Homoserine
Methionine
lysine
Aspartate
NAD+
Arg
Zn/Cd
Mce
Lindane
Lipoprotein
Phospolipid
Lipopolysaccharide
D
RaxAB-RaxC type I secretion system
α-Hemolysin/cyclolysin transport system
AlgE-type Mannuronan C-5-Epimerase transport system
Type II secretion system
Type IV pili
DegQ
Lipase/Esterase
Murein transglycosylase
Homoprotocatechuate
2-oxohept-3-enedioate
Oxalate
Glutamate/Aspartate
Leucine
Acetyl-CoA
Benzoyl-CoA
3-hydroxypimeloyl-CoA
Formate
H+
O-acetylserine/cysteine
Glutamate
Glutathione
Arogenate
Homoserine/Homocerine-lactone
Ubiquinone
Chorsimate
Prephenate
Pretyrosine
Tyrosine
K+
putrescine
Lysine
Saccharopine
Acetoacetyl-CoA
Phenylpyruvate
Phenylalanine
Benzoate
F.A synthesis
Long-chain-acyl-ACP
Pimeloyl-ACP
Leucine
Tricarboxilate
KAPA
Biotin
Proline
Pimelate
Pimeloyl-CoA
sn-glycerol 3-phosphate
Glutamate
PP path
Erythrose-4P
Pyridoxal-5P
MO4 2-
SO4 2-
PO4 3-
CrO2-
Na+/H+
Acetate
Cytochrome bc complex
Zn/Cd
Zn/Mn
Co/Mg
NH4+
Lipopolysaccharide
Cu2+
K+
Cytochrome C cbb-3 type
E
SO4
peptide/Ni
multi-sugar
Mg2+
Sec Pathway
F-type ATPase
G6P
6P-Gluconate
F6P
X5P
Ru5P
F1,6P
Glycerol-2P
3(Acyl CoA)
Ri5P
GAP
Glycine
PRPP
Triacylglycerol
Glutamine
3GP
Serine
fatty acids
Galactose
2GP
Glyoxylate
Pyrimidine
IPP
PEP
DMPP
Purine
lactate + NAD+
GGPP
IMP
pyruvate
NADH + H+
H+
Na+/K+
Lipopolysaccharide
Cu+
Ca2+
Zn/Cd
polysaccharide

804

Figure 6| Representative metabolic maps of bacterial groups in the reactor. To prevent redundancy, the metabolism is presented in a nested approach with each panel showing only paths unique to the relevant metabolic group. Two exceptions are group β (all detected paths are shown), and group δ. The latter is not presented here since it shares all paths with group ɣ and only differs by auxotrophies. (A) Metabolic map of paths that are common to all bacteria in the reactor (except Microgenomates and *Brocadia* sp.). The vast majority of bacteria in the reactor are heterotrophs, capable of carbohydrate-based metabolism (glycolysis, pentose phosphate pathway) and amino acid-based metabolism. Some bacteria can respire oxygen, but can also ferment (acetate/alanine). (B) Paths unique to group α. The bacteria have genes for hydrogen oxidation, supporting anaerobic growth, as well as genes for oxidative phosphorylation with cytochrome BD complex. These bacteria have a cassette of extracellular proteases and decarboxylases, paired with a wide array of transporters. The bacteria are also potentially capable of synthesizing long chain isoprenoids. (C) Paths found in Gram (-) bacteria (groups ɣ, δ, and ε). Most paths are related to fatty acid and lipid synthesis. Several important precursors (chorismate and IMP) can potentially be synthesized by these bacteria. Motility is also a common feature in these bacteria (via a flagellar motor) (D) Unique paths of group ε (Proteobacteria). This group has the potential to synthesize multiple vitamins and cofactors (biotin, pyridoxal, glutathione, etc.), as well as several aa. (tyrosine, phenylalanine, proline). Another unique feature is the multiple secretion systems present in the bacteria. (E) Metabolic profile of CPR bacteria (Microgenomates). The bacteria are obligate anaerobes that ferment pyruvate. They can only utilize carbohydrates as carbon source. Some of the bacteria in this group might also be able to synthesize long chain isoprenoids, in the same path as group α.

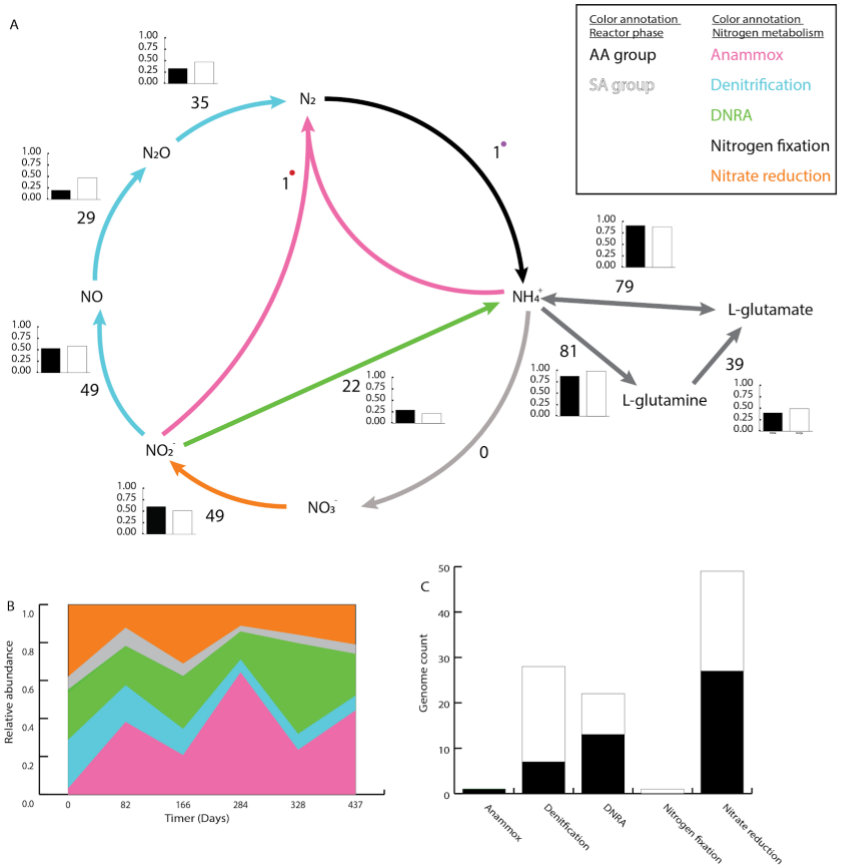


Figure 7| Nitrogen cycle in the anammox reactor. (A) The steps in the nitrogen cycle are color coded by their association to different types of metabolism. The number of bacteria with genes encoding a given

step is listed and the bar chart depicts the ratio of bacteria within the AA/SA groups associated with the step. (B) Changes in relative abundance of bacterial groups by their nitrogen metabolism. Only a single path was assigned to each genome for the purpose of this analysis. Since nitrate reduction is also considered a first step in denitrification and DNRA is was assigned only when other paths were not present. Bacteria with no complete metabolic path are depicted in light grey. Anammox is the dominant nitrogen metabolic path at Days 82, 284, and 437. This matches the reactor performance monitoring (Figure 1). At times when the source community is predominant (Days 0 and 166), nitrogen reduction is the most common metabolic path, followed by DNRA. During the period of reactor destabilization (Day 328) the DNRA bacteria dominate the community. (C) The number of bacteria in which a given metabolic path was detected. The bars are divided by AA/SA association. Unlike panel B, overlap of functions was allowed for the genome counts. The largest group are the nitrate reducers, followed by denitrifiers and DNRA. Denitrifiers are more common among the SA group, while DNRA are more common among AA group.

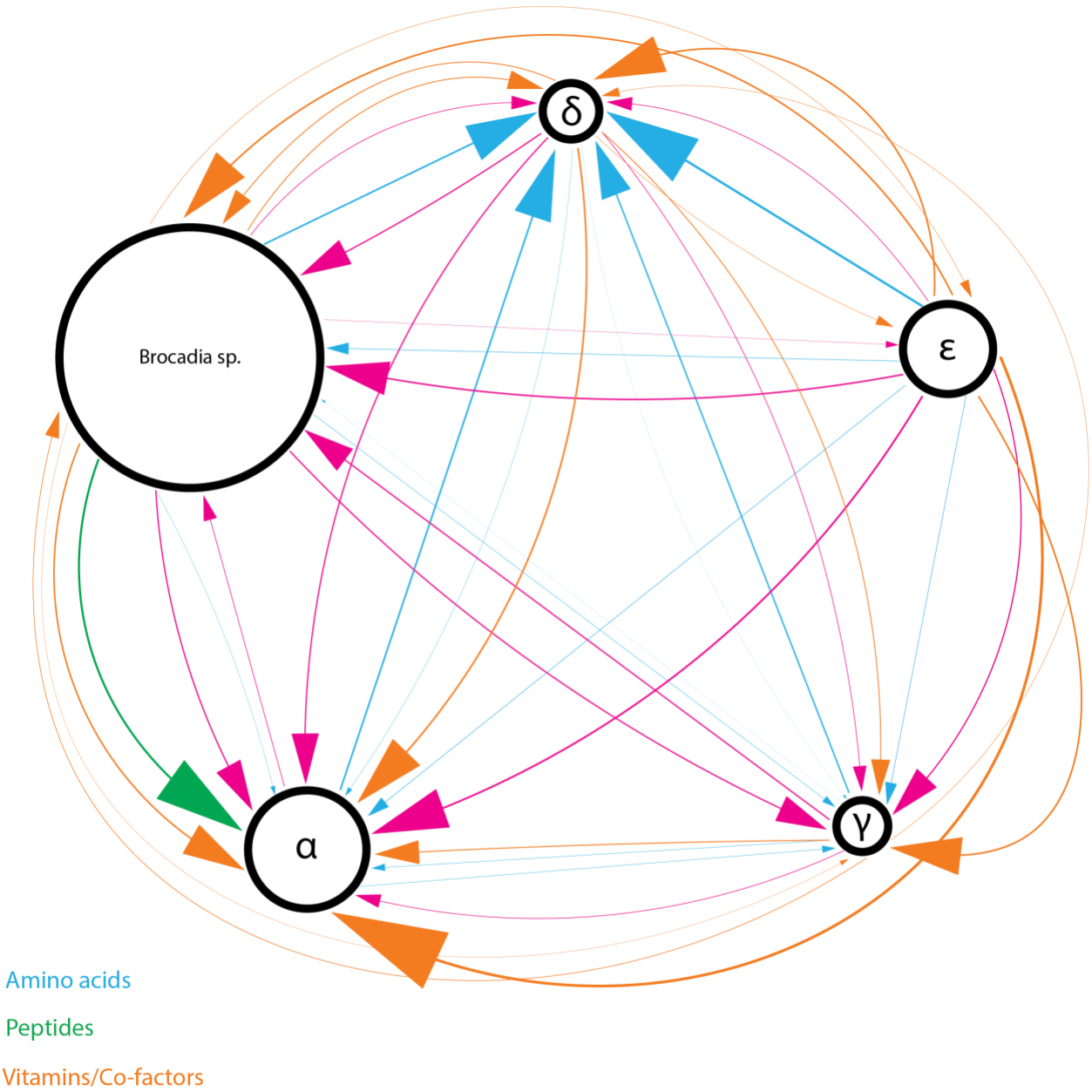


Figure 8| Potential metabolic hand-offs between the MO groups in the anammox reactor. Arrows were assigned according to absence of ability to synthesis a metabolite and connect to all groups that do have the ability (meaning there is redundancy in arrows). The arrowhead points at the group that receives the metabolite. The width of the arrow is proportional to the ratio of metabolites of a given type that are provided; amino acids - 20 metabolites; Peptides - deduced from proteases and transporters (Figure 6B); Vitamins/Co-factors - 10 metabolites; Lipids/Fatty acids - 7 metabolites. The size of each group is proportional to their relative abundance at Day 437. Group β is not shown since the assumption is that its members obtain all of their nutrients and metabolites from a their host. Overall, groups α and δ receive the most metabolites and group ε the least. Group δ has the highest number of aa. synthesis auxotrophies and can potentially acquire these from many other community members. Group ε has only a single auxotrophy in vitamin/Co-factor synthesis while most other groups have multiple auxotrophies (group α capable of only a single metabolite). *Brocadia* sp. is the only bacterium capable of vitamin B12 synthesis.

Tables

Table 1 | Genome parameters for the representative bacteria from the anammox reactor

Genome	Taxonomy (ggKbase)	Genome length (bp)	GC%	# contigs	# features	longest contig	Completeness (checkM)
anamox1_Acidimicrobiia_62_4_curated	Acidimicrobiia, Actinobacteria, Bacteria	298722	58.08	173	496	5546	65.54
anamox1_Acidobacteria_62_5_curated	Acidobacteria, Bacteria	3111074	61.59	565	3251	20877	65.54
anamox1_Bacteria_33_9_curated	Bacteria	775494	32.38	45	793	69755	70.85
anamox1_Bacteria_45_8_curated	Bacteria	2579669	45	166	2221	61316	87.77
anamox1_Bacteria_50_18_curated	Bacteria	820566	49.55	22	874	204925	77.9
anamox1_Bacteria_53_17_curated	Bacteria	3465172	53.24	12	2895	903458	100
anamox1_Bacteria_55_18_curated	Bacteria	4100779	54.56	216	3365	90642	98.28
anamox1_Bacteria_56_37_curated	Chloroflexi, Bacteria	4970250	55.88	43	4404	437022	94.83
anamox1_Bacteria_57_32_curated	Bacteria	2939326	56.91	76	2571	186230	95.69
anamox1_Bacteria_57_5_curated	Bacteria	3332363	56.3	813	3846	19658	68.62
anamox1_Bacteria_57_9_curated	Bacteria	3469651	57.28	74	3012	281471	93.65
anamox1_Bacteria_64_7_curated	Bacteria	3988790	63.42	306	3679	75294	93.03
anamox1_Bacteria_65_5_curated	Bacteria	2089721	64.54	434	2237	15423	51.28
anamox1_Bacteria_72_15_curated	Bacteria	2937143	72.45	24	2507	371794	87.93
anamox1_Bacteroidetes_38_5_curated	Bacteroidetes, Bacteria	1414251	37.34	293	1501	13351	62.96
anamox1_Bacteroidetes_39_16_curated	Bacteroidetes, Bacteria	2896628	38.84	40	2472	542771	100
anamox1_Bacteroidetes_63_11_curated	Bacteroidetes, Bacteria	3356307	63.21	73	2915	220821	100
anamox1_Betaproteobacteria_69_7_curated	Betaproteobacteria, Proteobacteria, Bacteria	2106569	66.39	531	2636	12349	66.55
anamox1_Burkholderiales_70_40_curated	Burkholderiales, Betaproteobacteria, Proteobacteria, Bacteria	3601855	70.09	60	3356	242838	100
anamox1_Burkholderiales_71_17_curated	Burkholderiales, Betaproteobacteria, Proteobacteria, Bacteria	3639459	70.81	114	3448	149014	99.22
anamox1_Candidatus_Cloacimonas_acidaminovorans_38_6_curated	Candidatus Cloacimonas acidaminovorans, Candidatus Cloacimonas, Cloacimonetes, Bacteria	1192038	35.28	246	1132	15612	56.47
anamox1_Chloroflexi_52_59_curated	Chloroflexi, Bacteria	2864537	52.47	96	2759	102167	100
anamox1_Gammaproteobacteria_64_6_curated	Gammaproteobacteria, Proteobacteria, Bacteria	2447777	64.02	391	2594	28213	84.25
anamox1_Gemmatimonas_aurantiaca_57_6_curated	Gemmatimonas aurantiaca, Gemmatimonas, Gemmatimonadales, Gemmatimonadetes, Gemmatimonadetes, Bacteria	2631895	56.94	255	2549	56382	91.22
anamox1_Lysobacter_67_10_curated	Lysobacter, Xanthomonadales, Gammaproteobacteria, Proteobacteria, Bacteria	783854	66.76	125	840	26187	34.64
anamox1_Nitrosomonas_europaea_50_14_curated	Nitrosomonas europaea, Nitrosomonas, Nitrosomonadales, Betaproteobacteria, Proteobacteria, Bacteria	2128781	50.43	37	2006	177381	98.28
anamox1_Planctomycetia_64_8_curated	Planctomycetia, Planctomycetes, Bacteria	4138393	63.58	783	3420	35628	87.62
anamox1_Proteobacteria_65_15_curated	Proteobacteria, Bacteria	2240508	65.45	348	2476	63177	58.59
anamox1_Proteobacteria_67_8_curated	Proteobacteria, Bacteria	2912885	66.93	203	2910	122132	81.27
anamox1_Pyrinomonas_methylaliphatogenes_54_8_curated	Pyrinomonas methylaliphatogenes, Pyrinomonas, Acidobacteria, Bacteria	2071852	54.23	87	1989	105081	95.69
anamox1_RBG_16_Ignavibacteria_36_9_curated_35_5_curated	RBG_16_Ignavibacteria_36_9_curated, Ignavibacteriales, Ignavibacteria, Ignavibacteriae, Bacteria	1163158	34.03	347	1374	12559	70.14
anamox1_Rhodocyclales_69_13_curated	Rhodocyclales, Betaproteobacteria, Proteobacteria, Bacteria	2263347	68.53	185	2364	100335	96.55
anamox1_Sphingobacteriales_42_27_curated	Sphingobacteriales, Sphingobacteriia, Bacteroidetes, Bacteria	3490519	42	95	3068	154972	100
anamox1_Sphingobacteriales_43_8_curated	Sphingobacteriales, Sphingobacteriia, Bacteroidetes, Bacteria	3156182	42.95	130	2826	130033	97.41
anamox1_Truepera_radiovictrix_71_5_curated	Truepera radiovictrix, Truepera, Deinococcales, Deinococci, Deinococcus-Thermus, Bacteria	435656	69.12	317	694	3402	24.49
anamox1_Turneriella_parva_44_7_curated	Turneriella parva, Turneriella, Spirochaetia, Spirochaetes, Bacteria	2667359	44.12	240	2748	87610	92.24
anamox2_Actinobacteria_65_5_curated	Bacteria	2799571	63.61	662	3147	20411	67.46
anamox2_Bacteria_34_5_curated	Bacteria	434538	33.34	88	508	19334	50.78
anamox2_Bacteria_61_6_curated	Bacteria	2415824	60.87	385	2646	23948	65.91
anamox2_Bacteria_68_6_curated	Bacteria	2291336	67.96	460	2621	29805	64.03
anamox2_BJP_IG2103_Bacteroidetes_37_22_46_7_curated	BJP_IG2103_Bacteroidetes_37_22, Bacteroidetes, Bacteria	2490972	45.52	190	2226	53299	93.1
anamox2_Burkholderiales_67_5_curated	Burkholderiales, Betaproteobacteria, Proteobacteria, Bacteria	1246117	66.28	298	1434	12340	38.05
anamox2_Burkholderiales_68_9_curated	Burkholderiales, Betaproteobacteria, Proteobacteria, Bacteria	3527974	68.25	181	3452	116403	91.22
anamox2_Burkholderiales_70_13_curated	Burkholderiales, Betaproteobacteria, Proteobacteria, Bacteria	3074974	70.33	39	2841	297272	84.48
anamox2_Chloroflexi_60_8_curated	Chloroflexi, Bacteria	2023353	60.06	111	1885	73961	47.49
anamox2_Gammaproteobacteria_67_23_curated	Gammaproteobacteria, Proteobacteria, Bacteria	2663638	67.06	116	2584	140445	79.31
anamox2_Hydrogenophilales_66_19_curated	Hydrogenophilales, Betaproteobacteria, Proteobacteria, Bacteria	2257110	65.94	122	2331	92964	91.95
anamox2_Ignavibacteriales_33_72_curated	Ignavibacterium album, Ignavibacterium, Ignavibacteriales, Ignavibacteria, Ignavibacteriae, Bacteria	3331053	33.19	161	2995	201964	98.28
anamox2_Ignavibacteriales_33_9_curated	Ignavibacteriales, Ignavibacteria, Ignavibacteriae, Bacteria	2936140	33.09	166	2753	82489	86.05
anamox2_Ignavibacteriales_41_12_curated	Ignavibacteriales, Ignavibacteria, Ignavibacteriae, Bacteria	3156563	41.32	42	2805	297034	96.55
anamox2_Methanosarcina_thermophila_41_9_curated	Methanosarcina thermophila, Methanosarcina, Methanosarcinales, Methanomicrobia, Euryarchaeota, Archaea	2959285	41.19	82	2703	143092	35.55
anamox2_Microgenomates_45_6_curated	Microgenomates, Bacteria	740215	44.28	99	904	30865	73.35
anamox2_Microgenomates_49_6_curated	Microgenomates, Bacteria	950204	49.23	109	1082	36723	68.42
anamox2_Myxococcales_71_5_curated	Myxococcales, Deltaproteobacteria, Proteobacteria, Bacteria	1434218	70.21	495	1687	8213	30.21
anamox2_Rhizobiales_67_45_curated	Mesorhizobium, Rhizobiales, Alphaproteobacteria, Proteobacteria, Bacteria	4837226	66.56	24	4668	717639	100
anamox2_Roizmanbacteria_38_11_curated	Roizmanbacteria, Microgenomates, Bacteria	663231	37.62	14	693	114830	58.31
anamox2_Sphingobacteriales_41_11_curated	Sphingobacteriales, Sphingobacteriia, Bacteroidetes, Bacteria	2768577	41.16	129	2490	79962	87.07
anamox2_sub_Ignavibacterium_album_33_16_curated	Ignavibacterium album, Ignavibacterium, Ignavibacteriales, Ignavibacteria, Ignavibacteriae, Bacteria	3391206	33.43	195	3029	116818	97.81
anamox2_Verrucomicrobia_58_8_curated	Verrucomicrobia, Bacteria	3592680	57.68	250	3293	90598	87.93
anamox2_Verrucomicrobia_62_8_curated	Verrucomicrobia, Bacteria	3685167	61.78	161	3134	100565	98.28
anamox2_Xanthomonadales_68_9_curated	Xanthomonadales, Gammaproteobacteria, Proteobacteria, Bacteria	1242229	67.63	21	1168	141855	50
anamox2_Xanthomonadales_70_8_curated	Xanthomonadales, Gammaproteobacteria, Proteobacteria, Bacteria	2605543	68.54	346	2555	32162	82.76
anamox3_Acidobacteria_71_4_curated	Acidobacteria, Bacteria	1334157	69.62	772	1759	7429	39.71
anamox3_Bacteria_50_5_curated	Bacteria	1042006	49.54	203	1076	20276	42.87
anamox3_Bacteria_66_7_curated	Bacteria	2266342	65.35	286	2150	45201	81.9
anamox3_Bacteria_67_13_curated	Bacteria	1642722	66.5	47	1442	89691	67.24
anamox3_Bacteria_67_15_curated	Bacteria	2765727	66.48	144	2980	91127	95.69
anamox3_Bacteroidetes_39_15_curated	Bacteroidetes, Bacteria	2533542	39.26	23	2188	636516	100
anamox3_BJP_JG2069_Ignavibacteriae_38_11_30_7_curated	BJP_JG2069_Ignavibacteriae_38_11, Ignavibacteriae, Bacteria	1803433	30.38	132	1600	73197	73.51
anamox3_Burkholderiales_71_6_curated	Burkholderiales, Betaproteobacteria, Proteobacteria, Bacteria	1367674	69.86	648	1904	18351	32.85
anamox3_Chloroflexi_59_6_curated	Chloroflexi, Bacteria	1407767	53.68	351	1507	12267	40.22
anamox3_Chloroflexi_68_6_curated	Chloroflexi, Bacteria	2294191	67.35	479	2330	19055	53.92
anamox3_Nitrosomonas_europaea_51_28_curated	Nitrosomonas europaea, Nitrosomonas, Nitrosomonadales, Betaproteobacteria, Proteobacteria, Bacteria	2372894	50.54	93	2302	91833	98.12
anamox3_Omnitrophica_63_14_curated	Bacteria	365510	62.81	4	402	274544	25.86
anamox3_Opitutus_terrae_67_4_curated	Opitutales, Opitutae, Verrucomicrobia, Bacteria	780321	65.72	503	1046	5713	24.42
anamox3_Pedosphaera_parvula_66_5_curated	Pedosphaera parvula, Pedosphaera, Verrucomicrobiales, Verrucomicrobiae, Verrucomicrobia, Bacteria	1233438	65.98	543	1384	10642	56
anamox3_Proteobacteria_68_11_curated	Proteobacteria, Bacteria	3218375	68.07	66	3078	349840	100
anamox3_Sphingobacteriales_44_6_curated	Sphingobacteriales, Sphingobacteriia, Bacteroidetes, Bacteria	3033899	42.77	512	2768	29727	73.9
anamox3_Sphingobacteriales_50_9_curated	Sphingobacteriales, Sphingobacteriia, Bacteroidetes, Bacteria	4305233	49.73	216	2999	129109	99.14
anamox3_sub_Ignavibacteriales_42_14_curated	Ignavibacteriales, Ignavibacteria, Ignavibacteriae, Bacteria	3163923	42.21	22	2498	452672	96.55
anamox3_Verrucomicrobia_59_12_curated	Bacteria	2712092	59.27	50	2592	173980	96.55
anamox4_Alphaproteobacteria_58_14_curated	Alphaproteobacteria, Proteobacteria, Bacteria	2593178	57.86	5	2558	1525488	98.28
anamox4_Bacteria_63_7_curated	Bacteria	3451341	62.64	351	3276	43084	93.1
anamox4_Bacteria_69_43_curated	Bacteria	3827900	69.37	15	3388	1008601	86.21
anamox4_Bacteroidetes_40_74_curated	Bacteroidetes, Bacteria	2627547	39.91	15	2219	539103	100
anamox4_Candidatus_Nitrospira_defluvii_60_9_curated	Candidatus Nitrospira defluvii, Nitrospira, Nitrospirales, Nitrospira, Nitrospirae, Bacteria	3085099	60.3	94	3084	176639	95.69
anamox4_Chloroflexi_66_15_curated	Chloroflexi, Bacteria	4383590	65.96	133	3602	139118	91.22
anamox4_Gammaproteobacteria_67_14_curated	Gammaproteobacteria, Proteobacteria, Bacteria	3265689	66.91	73	3049	145604	98.28
anamox4_Microgenomates_48_8_curated	Microgenomates, Bacteria	999881	47.64	49	1055	99690	80.02
anamox4_Nitrosomonas_eutropha_48_11_curated	Nitrosomonas eutropha, Nitrosomonas, Nitrosomonadales, Betaproteobacteria, Proteobacteria, Bacteria	1880399	48.46	103	1862	88908	94.83
anamox4_sub_Candidatus_Brocadia_sinica_42_75_curated	Candidatus Brocadia sinica, Candidatus Brocadia, Candidatus Brocadiales, Planctomycetia, Planctomycetes, Bacteria	3107335	42.29	64	2859	160056	98.28
LAC_NA06_Anaerolineales_42_27_curated	Anaerolineales, Anaerolineae, Chloroflexi, Bacteria	2409642	41.78	210	2380	75774	93.1
LAC_NA06_Bacteria_60_18_curated	Bacteria	3689918	59.98	50	3221	327084	94.83
LAC_NA06_Bacteroidetes_30_9_curated	Bacteroidetes, Bacteria	2233050	29.92	225	2055	41036	85.06
LAC_NA06_Betaproteobacteria_71_7_curated	Betaproteobacteria, Proteobacteria, Bacteria	4060980	69.45	1841	5695	17332	69.75
LAC_NA06_Burkholderiales_73_13_curated	Burkholderiales, Betaproteobacteria, Proteobacteria, Bacteria	3708868	72.67	335	3469	53457	94.51
LAC_NA06_Candidatus_Saccharibacteria_40_6_curated	Candidatus Saccharibacteria, Bacteria	463567	39.93	148	580	16266	58.46
LAC_NA06_Candidatus_Solibacter_usitatus_62_12_curated	Acidobacteria, Bacteria	4513182	61.85	167	4021	141967	91.38
LAC_NA06_Chloroflexi_57_14_curated	Chloroflexi, Bacteria	2719486	57.25	159	2567	157158	86.91
LAC_NA06_Fimbriimonas_ginsengisoli_61_14_curated	Fimbriimonas ginsengisoli, Fimbriimonas, Fimbriimonadales, Fimbriimonadia, Armatimonadetes, Bacteria	2673573	61.04	54	2542	245447	88.45
LAC_NA06_Microgenomates_41_17_curated	Microgenomates, Bacteria	1110978	41.17	7	1216	569682	80.88
LAC_NA06_Rhizobiales_66_24_curated	Rhizobiales, Alphaproteobacteria, Proteobacteria, Bacteria	3228463	66.32	148	3270	125494	98.28
LAC_NA06_RIFCSPHIGHO2_12_FULL_Dadabacteria_53_21_curated_58_6_curated	RIFCSPHIGHO2_12_FULL_Dadabacteria_53_21_curated, Dadabacteria, Bacteria	869333	55.01	539	1358	4902	33.68
LAC_NA06_sub_Chloroflexi_59_14_curated	Chloroflexi, Bacteria	2594752	59.45	164	2359	101841	91.22
LAC_NA06_sub_Chloroflexi_61_22_curated	Chloroflexi, Bacteria	2755473	61.22	170	2707	68305	96.55
LAC_NA07_Acidobacteria_70_38_curated	Acidobacteria, Bacteria	3262818	69.91	40	2917	270970	94.83
LAC_NA07_Actinobacteria_71_14_curated	Actinobacteria, Actinobacteria, Bacteria	3047960	70.65	64	3117	193413	93.97
LAC_NA07_Actinobacteria_74_18_curated	Actinobacteria, Actinobacteria, Bacteria	1235923	73.53	141	1261	48493	64.86
LAC_NA07_Actinotalea_fermentans_75_19_curated	Actinotalea fermentans, Actinotalea, Micrococcales, Actinobacteria, Actinobacteria, Bacteria	2905150	75.39	31	2745	349240	100
LAC_NA07_Bacteria_38_171_curated	Bacteria	2383463	37.56	22	2157	344212	98.28
LAC_NA07_Bacteria_70_305_curated	Bacteria	2792777	70.13	54	2465	247763	98.28
LAC_NA07_Bacteria_71_12_curated	Bacteria	2770642	70.73	206	2877	57439	81.41
LAC_NA07_Burkholderiales_70_312_curated	Burkholderiales, Betaproteobacteria, Proteobacteria, Bacteria	2688511	69.75	290	2827	57610	93.32
LAC_NA07_Caldilinea_aerophila_61_12_curated	Caldilinea aerophila, Caldilinea, Caldilineales, Caldilineae, Chloroflexi, Bacteria	3885866	60.59	83	3168	234454	80.88
LAC_NA07_Candidatus_Solibacter_usitatus_59_6_curated	Candidatus Solibacter usitatus, Candidatus Solibacter, Solibacterales, Solibacteres, Acidobacteria, Bacteria	1779371	57.94	742	2242	9528	56.02
LAC_NA07_Chloroflexi_57_23_curated	Chloroflexi, Bacteria	6358497	57.22	93	5201	299767	89.66
LAC_NA07_Chloroflexi_57_9_curated	Chloroflexi, Bacteria	2081669	55.57	325	2239	31406	71.76
LAC_NA07_Chloroflexi_58_12_curated	Chloroflexi, Bacteria	2711149	57.43	154	2459	125878	60.42
LAC_NA07_Chloroflexi_60_59_curated	Chloroflexi, Bacteria	3858643	60.1	13	3478	869846	94.83
LAC_NA07_Chloroflexi_65_58_curated	Chloroflexi, Bacteria	3316034	65.08	327	3017	48064	87.93
LAC_NA07_Chloroflexi_67_63_curated	Chloroflexi, Bacteria	3446202	66.61	65	2852	450062	96.55
LAC_NA07_Proteobacteria_68_32_curated	Proteobacteria, Bacteria	2747459	68.4	14	2541	557607	98.28
LAC_NA07_RBG_16_RIF_CHLX_72_14_curated_75_20_curated	RBG_16_RIF_CHLX_72_14_curated, RIF-CHLX, Bacteria	2506614	74.69	63	2330	167021	91.38
LAC_NA07_Rhodobacterales_68_7_curated	Rhodobacterales, Alphaproteobacteria, Proteobacteria, Bacteria	1308110	67.47	473	1651	13528	52.93
LAC_NA07_Rhodocyclales_67_14_curated	Rhodocyclales, Betaproteobacteria, Proteobacteria, Bacteria	2723825	66.65	212	2909	51017	92.95
LAC_NA07_Roizmannbacteria_52_60_curated	Roizmannbacteria, Microgenomates, Bacteria	957307	52.33	6	1013	885168	73.98
LAC_NA07_Truepera_radiovictrix_72_29_curated	Truepera radiovictrix, Truepera, Deinococcales, Deinococci, Deinococcus-Thermus, Bacteria	1353035	72.21	165	1373	45517	82.6

Table 2 | Read counts to representative genomes across time points

Time piont (days)	# Total reads	# Total maped reads	% Mapped to rep	# Rep number
0	55398280	40291503	72.73	92
82	62544544	43877427	70.15	82
166	60931806	46030350	75.54	103
284	56282006	48644523	86.43	68
328	127048582	95132145	74.88	87
437	119945232	93737087	78.15	60

*** number of representatives based on threshold of >1 coverage and > 0.5 breadth**