Metagenomic analysis of anammox communities in three different microbial aggregates

Jianhua Guo[1,2*], Yongzhen Peng[1], Lu Fan[2], Liang Zhang[1], Bing-Jie Ni[2], Boran Kartal[3,4], Xin Feng[5], Mike S.M. Jetten[3], Zhiguo Yuan[1,2]

[1] Key Laboratory of Beijing for Water Quality Science and Water Environmental Recovery Engineering, Engineering Research Center of Beijing, Beijing University of Technology, Beijing 100124, PR China;

[2] Advanced Water Management Centre (AWMC), The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia;

[3] Microbiology, IWWR, Faculty of Science, Radboud University Nijmegen, Heyendaalseweg 135, 6525 AJ Nijmegen, the Netherlands

[4] Department of Biochemistry and Microbiology, Laboratory of Microbiology, Gent University, Gent, Belgium, 9000

[5] Beijing Genomics Institute (BGI)-Shenzhen, China

*Correspondence: J Guo, Key Laboratory of Beijing for Water Quality Science and Water Environmental Recovery Engineering, Engineering Research Center of Beijing, Beijing University of Technology, Beijing 100124, PR China.

E-mail: gjh@bjut.edu.cn; j.guo@awmc.uq.edu.au

Running title: Metagenomic analysis of anammox communities

Keywords: ammonia-oxidizing bacteria (AOB); Anammox; biofilm; biological nitrogen removal; granule; metagenomic;

Summary: There is great potential to understand the functional diversity of microorganisms that are involved in wastewater treatment through metagenomic analyses. This study presents the first metagenomic comparison of taxonomic and functional profiles of the microbial communities occurring in different aggregates from anaerobic ammonium-oxidizing (anammox) bioreactors. The anammox bacterial communities in both biofilm, developed in a carrier, and granule sludge samples showed relatively high abundance and diversity compared to floccular sludge. Four of the five known genera of anammox bacteria were detected in the three cultures except *Candidatus* Jettenia, which was absent in the granules. *Candidatus* Kuenenia comprised the major population of anammox bacteria in these three sludges, independent of their growth morphologies. The genome assembled for the *Candidatus* Kuenenia in the granule was very similar to the published reference genome of *Candidatus* K. stuttgartiensis. Genes involved in the metabolism of the anammox process were highly detected in the biofilm and granule sludges. In particular, the abundance of hydrazine synthase gene (*hzs*) in the biofilm was around 486 times more pronounced than that in the granules. The knowledge gained in this study highlights an important role of sludge aggregate in affecting community structure and metabolic potential of anammox systems.

Introduction

Understanding of the global nitrogen cycle has been fundamentally altered by discovery of anaerobic ammonium-oxidizing (anammox) bacteria (Mulder et al., 1995). Anammox bacteria have the unique ability to convert ammonium and nitrite to form dinitrogen gas (Strous et al., 1999). The identification of the responsible chemolithoautotrophic bacteria as member of the order *Brocadiales* (Strous et al., 1999; Jetten et al., 2010) has exposed their ecological and applied significance. Anammox bacteria play a critical role in global climate change due to

2

their interaction between nitrogen and carbon cycling. It is estimated that marine anammox bacteria contribute up to 50% in the loss of fixed nitrogen from the oceans globally (Kuypers et al., 2003; Lam and Kuypers, 2011).

Biological wastewater treatment is the largest application of environmental biotechnology in the world (Guo et al., 2013). Conventionally, nitrogen in wastewater is removed through nitrification and denitrification. Ammonium is sequentially oxidized to nitrate by ammonia-oxidizing bacteria (AOB) and nitrite-oxidizing bacteria (NOB) via nitrite. Nitrate is then reduced to dinitrogen gas by heterotrophic denitrifying bacteria with organic carbon as the electron donor. However, biological nitrogen removal in this way substantially reduces the amount of organic carbon available for bioenergy recovery, consumes large amounts of energy and electricity for aeration and produces substantial waste sludge and potential greenhouse gases (e.g. N_2O) (Guo et al., 2013). The anammox process provides a more sustainable solution to these problems, which has been widely applied for nitrogen removal from wastewater as a promising alternative process (Ni et al., 2009; Kartal et al., 2010; Kartal et al., 2013; Chen et al., 2014), since it offers the advantages of less sludge production, decreased oxygen consumption and reduced N_2O emissions.

To date, several cultivation systems for anammox bacteria have been developed, including upflow anaerobic sludge blanket (UASB), fixed biofilm bed, fluidized bed, sequencing batch reactor (SBR), gas-lift reactors, rotating biological contactors and membrane bioreactors (Strous et al., 1997; Schmidt et al., 2003; van der Star et al., 2008; Ni et al., 2010; Kartal et al., 2011a). In these different reactors the morphologies of the microbial aggregates enriched with anammox bacteria can be categorized into three different types, i.e. flocculent sludge, attached biofilm and granular sludge. The macroscopic architecture of the aggregates affects the ecology of the microbial consortia (Rittmann and McCarty, 2001). Although many aspects of the engineering and design of anammox systems have been improved (Hu et al., 2013a), the microbial community and functional profiles of these various anammox microbial

3

aggregates have not been fully explored. The application of reactors with different designs and operation conditions most likely results in the enrichment of distinct microbial communities and specific anammox groups with different functional potentials (Vlaeminck et al., 2010; Buergmann et al., 2011). Hence, study of the taxonomic and functional traits of the microbiomes is essential for optimizing reactor design and performance (Vanwonterghem et al., 2014b). Specifically, different anammox cultures may be characterized by different abundances of anammox bacteria, by the presence of specific anammox species or strains, alternative enzymes and pathways, and by variations of interactions between the microbes (Jetten et al., 2009; Hu et al., 2010).

Currently, the anammox bacteria are detected as five candidate genera in the phylum *Planctomycetes*, and the order *Brocadiales* (Jetten et al., 2010). These genera are *Brocadia*, *Kuenenia*, *Scalindua*, *Anammoxoglobus* and *Jettenia* (Kuenen, 2008; Kartal et al., 2012). In contrast to their distribution in mostly marine and some freshwater environments, where only members of *Scalindua* have been detected (van de Vossenberg et al., 2008; van de Vossenberg et al., 2013), all these five genera of anammox bacteria have been reported in wastewater treatment systems (Schmid et al., 2003; Kartal et al., 2012). These studies used molecular methods targeting the 16S ribosomal rRNA genes or other marker genes to reveal microbial community phylogeny. The methods included denaturing gradient gel electrophoresis (DGGE) (Tal et al., 2003; Schmid et al., 2005; Hu et al., 2010; Li and Gu, 2011), fluorescence *in situ* hybridization (FISH) (Pavlekovic et al., 2009; Almstrand et al., 2014), and sequencing after cloning (Kartal et al., 2011a; Zhu et al., 2013). However, most of these methods are gene amplification-based that may cause technical biases (Aird et al., 2011; Ye et al., 2012) and thus provide a skewed community structure or even fail to detect certain anammox groups (Han et al., 2013). Moreover, the marker-gene based studies provide limited information about the functional characteristics of anammox cultures. Metagenomic studies have been conducted on full-scale wastewater treatment plants (WWTPs) (Albertsen et al.,

4

2012; Yu and Zhang, 2012; Zhang et al., 2012), however, community-level functional profiling has been studied only on a few anammox reactors (Strous et al., 2006; Gori et al., 2011; van de Vossenberg et al., 2013). In particular, the genomes of anammox bacteria *Candidatus* K. stuttgartiensis, B. fulgida*, and* S. profunda have been successfully assembled through the sequencing of anammox culture metagenomes (Strous et al., 2006; Gori et al., 2011; van de Vossenberg et al., 2013).

This study aimed to use metagenomic deep-sequencing analysis to reveal phylogenetic and functional traits in three different anammox cultures (i.e., flocs, attached biofilm and granules). We attempt to link reactor operation conditions and reactor performances to anammox populations, functional traits and their cooperation or competition with AOB and/or NOB in nitrogen metabolism.

Results and Discussion

Performance and morphologies of the anammox microbial aggregates

Anammox sludge samples with the three different growth forms of floccular sludge, attached biofilm and granular sludge, were collected from a pilot-scale integrated fixed-film activated sludge (IFAS) reactor (flocs and biofilm) and from a lab-scale UASB-type granule anammox reactor (granules). Before the sampling, the pilot-scale IFAS anammox reactor and the lab-scale UASB anammox reactor had been operated continuously for more than 470 and 1400 days, respectively for nitrogen removal (Tables S1 and S2). Stable anammox performance in terms of total nitrogen removal efficiency was averaged at 83% and 72% in the IFAS and UASB respectively, for the periods of operation (Figure S1). When the IFAS anammox reactor was fed total nitrogen concentrations of between 600~938 mg/L, the average effluent concentrations of ammonium, nitrite and nitrate were 77.8 mg/L, 14.5 mg/L and 36.3 mg/L, respectively. The obtained maximum nitrogen conversion rate was 1.2 kg N/m^3/d.

The UASB laboratory reactor was also operated for nitrogen removal. In order to get granules to form in the reactor, the upflow fluid velocity within the reactor was increased by

5

changing recirculation rates from 300% to 500%. This was done on day 760 of operation and in the following 40 days a part (~30%) of the suspended biomass was washed out, and anammox granules formed and grew bigger. At the sampling time, on around day 1400 of operation, this reactor achieved a nitrogen removal rate of 0.6 kg N/m^3/d, and a red granular sludge had developed (Figure 1). The average effluent concentrations of ammonium, nitrite and nitrate were 72.0, 23.4 and 35.0 mg/L, respectively, and the average nitrogen removal efficiency was 72% (Figure S2).

The morphologies of the three anammox communities were different (Figure 1). The flocs had relatively small diameters. The average floc size was 176.5 μm and the volume percent of flocs with diameter lower than 300 μm was 75.3%, based on the size distribution measurement using a laser scattering particle analyzer (Malven, UK). The flocs were reddish-brown in color, likely due to the presence of anammox bacteria. In the IFAS reactor polyethylene sponges were used as biomass carriers for the development of biofilm from suspended sludge. The attachment of anammox cells to these bio-carriers was observed in the beginning of the cultivation period. After more than one year of continuous operation, a dense and thick biofilm was formed on the inner porous matrix of the sponge carriers (Figure 1B). The anammox granules taken from the UASB had diameters of between 1.0-2.5 mm, with an average of 1.9± 0.6 mm, and the percentage of granules with diameter larger than 1.5 mm was about 62%. These granules had a porous inner structure, which could facilitate the diffusion of ammonium and nitrite as well as the release of N$_2$.

The overall taxonomic structure and functional profile

In this study, high-throughput sequencing was used to explore microbial community structures, the relationship among the key microorganisms and pathways involved in nitrogen metabolism in the three different anammox microbial aggregates. The aggregates were the flocculent sludge, the attached biofilm and the granular sludge. Over 5 gigabases of metagenomic sequence data were generated with the Illumina HiSeq 2000 platform. About

6

4.3 Gb of high quality sequence data were obtained after read quality control and about 67-96 thousand contigs longer than 500 bp were obtained from the three samples (Table 1).

Taxonomic assignments indicated that bacteria dominated all three samples (Table S3), in which *Proteobacteria* and *Planctomycetes* were the most abundant bacterial populations at the phylum level while their percentages varied between samples (Figure S3). Approximately 92.0% of the sequence reads from the floc sample were assigned to *Proteobacteria* (Figure S3). In comparison, the most abundant phyla were *Planctomycetes* (63.0% in the biofilm and 89.6% in the granule), followed by *Proteobacteria* (32.4% in the biofilm and 10.1% in the granule) and *Firmicutes* (0.75% in the biofilm and 0.07% in the granule). Other bacterial phyla were generally very low in abundance except *Actinobacteria*, which comprised about 2% in the biofilm sample. *Beta-* and *Delta-proteobacteria* were the major classes in the phylum of *Proteobacteria*. High abundance of *Planctomycetes* was associated with high abundance of anammox bacteria in the biofilm and granule samples. This was consistent with the observed higher anammox activity in these two sludge samples in comparison to the activity of the flocs (Table S2, Figures S1 and S2). At the genus level, there were 885 genera detected in the flocs, 815 genera in the biofilm and 730 genera in the granules, demonstrating the microbial diversity within the three sludges. The top 24 abundant taxa in the three samples were selected and at the genus level, *Candidatus* Kuenenia was the most dominant taxon in both the biofilm and granules (Figure 2). In contrast, *Nitrosomonas* was the most dominant genus in the flocs.

In total, 183,398, 170,725 and 146,983 unique genes were predicted from the floc, biofilm and granule sludges respectively (Table 1). The genes were functionally annotated using the KEGG and eggNOG databases and it was apparent that most detected functions were shared across all the sludge samples, indicating potential similarities in metabolic capabilities of the different anammox microbial aggregates (Figure S4, Tables S4 and S5).

In particular, COG and KEGG functional clustering showed more similarity between the floc and the biofilm sludge (Figure 3). These two sludges were obtained from the same reactor, and predictably, displayed a greater similarity in overall functional profile. These floc and biofilm sludges were different to granules in that they had more coenzymes and lipid transport and metabolism genes than that of granules. While genes involved in amino acid-, energy-, and carbohydrate- metabolisms were higher in flocs than in the biofilm and granular sludge. Conversely, genes assigned into cell growth and death, cell motility, membrane transport and signal transduction had higher abundances in the biofilm and granules relative to the floc sludge (Figure S4, Tables S4 and S5).

The Anammox, AOB, and NOB populations

We specifically analyzed the anammox populations in the three samples based on metagenomic sequences (Figure 4), as well as quantitative real-time PCR (qPCR) (Table S6). The abundance of AOB and NOB detected by metagenomic sequencing were relatively higher than that by qPCR. In contrast, the qPCR approach had higher abundances of anammox bacteria in both the biofilm and granule, compared with the metagenomic analysis (87.8% vs 65.5% in the biofilm, while 94.3% vs 91.4% in the granule).

To date, five Candidatus genera in the phylum *Planctomycetes* are reported to specifically conduct the anammox process, these genera are *Candidatus* Brocadia, *Candidatus* Kuenenia, *Candidatus* Scalindua, *Candidatus* Anammoxoglobus and *Candidatus* Jettenia (Kuenen, 2008; Kartal et al., 2012). In this study, 13 species belonging to the four genera of *Candidatus*- Brocadia, Kuenenia, Scalindua and Anammoxoglobus were detected in all samples. The genus *Jettenia*, was also detected, but this was absent from the granules (Table S7 and S8). *Candidatus* Kuenenia appeared to be the most dominant anammox species in all three samples, suggesting this selection occurred irrespective of the type of growth mode. In previous studies on anammox bioreactors based on 16S rRNA gene analysis (such as FISH or PCR), it is also revealed that the anammox process in lab- or pilot-scale reactors is performed

8

mainly by bacteria of the genera *Candidatus* Kuenenia or *Candidatus* Brocadia (Kuenen, 2008; Hu et al., 2010; Hu et al., 2013b). In contrast to our findings, other granular sludge anammox reactors are dominated by *Candidatus* J. asiatica (around 50%) as detected by both16S rRNA gene quantitative PCR assays (Quan et al., 2008) and by metagenome analysis (Hu et al., 2012).

In natural and manmade habitats, anammox bacteria often coexist with AOB and NOB. AOB and NOB can be involved in anammox-based nitrogen removal processes, such that occurred in the IFAS reactor of this study. In the aerobic regions of the reactor, AOB would provide nitrite for the NOB and anammox bacteria. NOB would then compete with AOB for oxygen and with anammox bacteria for nitrite. In the IFAS reactor, anammox bacteria dominated over AOB in the biofilm, while AOB were more abundant than anammox bacteria in the flocs, where the proportion of AOB was 54.6% and anammox bacteria only accounted for 2.4% in terms of qPCR results (Table S6). In the biofilm, the proportion of AOB was 2.6%, while anammox bacteria were predominant at 87.8% (Table S6). *Nitrosomonas*, was the predominant AOB genus in the flocs and biofilm detected at 73.2% and 18.1% of those communities based on metagenomic data respectively. However, it was less abundant in the granule community at only 0.26% (Table S7, Figure 4B). Ammonia-oxidizing archaea were not detected by high-throughout sequencing in the IFAS reactor. While some reads were classified to the NOB genus *Nitrospira* (Table S7, Figure 4C), NOB were generally in very low abundance in the three sludges (Figure 4D). This reflected the different design and operational conditions in the reactors. In the IFAS reactor, it appears that nitritation by AOB would mostly take place in the flocs in the suspended phase, whereas the anammox process would mainly occur in the biofilm in the biomass carriers (Figure S6). This spatial distribution observed in the IFAS reactor was likely due to easier access of AOB for the substrates ammonium and oxygen occurring in the bulk liquid. It is reported that in oxygen-limited anammox reactors, nitritation driven by AOB is often restricted to an outer shell of aggregates

or in the smaller aggregates. In those reactors the anammox process is found in the central anoxic parts of the aggregates or in the larger aggregates (Nielsen et al., 2005; Kindaichi et al., 2007; Vlaeminck et al., 2008). Consequently, the different bacteria here are locating according to oxygen requirements. Additionally, anammox microorganisms are known to grow slowly (Kuenen, 2008; Kartal et al., 2012). For anammox to occur, the reactor operation would need to accommodate organisms with slow growth rates. Granules and fixed biofilm reactors will support organisms with slow growth rates. However, in floc based systems the growth rates can be manipulated and faster. Granules are also a type of large, self-supporting biofilm with structure that may have functional and phylogenetic layering (Ni et al., 2010). Such biofilm environments are favorable for cell-cell interactions, intercellular communication signaling, exchange of genetic material, and metabolite exchange (Hall-Stoodley et al., 2004).

Analysis of key nitrogen cycle inventory

To better understand the functional diversity, especially the nitrogen metabolism pathways in the samples, we calculated the abundance of the key enzymes for nitrification, denitrification and anammox (Figure 5). Most subunits of the genes that encode catabolic enzymes specific for denitrification were highly abundant. This included: *nar*G, with 12394 hits in flocs and 20025 hits in biofilm and 21463 hits in granules; *nor*B with 3397 hits in flocs, 908 hits in biofilm and 2423 hits in granules; and *nos*Z with 4212 hits in floc, 1628 hits in biofilm, and 7593 hits in granules. However, the genes encoding key enzymes of aerobic ammonium oxidation were relatively low in abundance in flocs and biofilm, with *amo*ABC having 5356 hits in flocs and 1070 hits in the biofilm.

The abundance of genes involved in the anammox reactions were quantified and the gene encoding hydrazine dehydrogenase (HDH), the key enzyme that catalyses the oxidation of the intermediate hydrazine (N_2H_4) to dinitrogen gas (N_2) (Kartal et al., 2011b; Kartal et al., 2013), was more abundant in the granules (1115 hits) and biofilm (332 hits) than in the flocs (70 hits).

10

The hydrazine synthase gene (*hzs*), which is unique for anammox bacteria (Strous et al., 2006; Kartal et al., 2011b), was detected in the anammox communities of both the biofilm and granules but not in the flocs. It should be noted that the abundance of *hzs* in the biofilm (11679 hits) was distinctly higher than in the granules (24 hits). Nitrite reductase genes, both the NO-producing *nir*K and *nir*S and the ammonium-producing *nrf*A, were detected in all three samples. In particular, the gene *nrf*A, encoding for the key enzyme of dissimilatory nitrate reduction to ammonium (DNRA) (Simon, 2002; Kraft et al., 2014; Welsh et al., 2014; Yoon et al., 2014), was found in higher abundance in the three samples in comparison to the other nitrite reductases. It should be noted that the metabolic pathways analyzed in this study were based on genomic DNA analysis, such that our interpretation implies functional capabilities of these microbiomes rather than actual activities. Further studies directly based on RNA, protein or metabolite levels are necessary to further explore the active functions in the different anammox microbial aggregates (van de Vossenberg et al., 2013).

Taxonomic origins of the functions in nitrogen metabolism

In addition to the anammox reactions, which were performed by *Candidatus* Brocadiaceae in these sludges, it was evident that other nitrogen based transformations were conducted by a diverse range of bacteria (Figure 6). For example, bacterial AmoA in both flocs and biofilm were mostly associated with *Nitrosomonadaceae*. Dissimilatory nitrite reduction, mediated by NrfA, could be carried out by other *Planctomycetaceae*-like organisms in the flocs and biofilm, but in the granules *Chloroflexi*-like populations had this ability. In addition, NirS in both biofilm and granules was mostly distributed to *Candidatus* Brocadiaceae and *Hydrogenophilaceae*, while NirK in flocs was assigned to *Nitrosomonadaceae*. NorB and NorZ were assigned to several phyla, including *Bacteroidetes*, *Betaproteobacteria* and *Gammaproteobacteria*.

These results showed an important phenomenon that certain biochemical processes in bioreactors could be conducted by various microbial groups including currently unclassified

taxa and taxa that previously thought not being able to carry a certain bioprocess. These taxa are almost always missed in 16S rRNA based approaches such as FISH or qPCR using clade-specific primers that are designed based on existing information gathered from public databases (Han et al., 2013). The 16S rRNA gene amplicon sequencing enables the community compositions to be determined, but it likely underestimates the bio-diversity of the system being studied as previously shown in (Orphan, 2009; Vanwonterghem et al., 2014a). Compared to traditional PCR based approaches, metagenomic analyses result in greater sequence depth and additionally provide an overview of the genetic capabilities. Interestingly, the metagenomic approach used in this study along with the reactor performance data provide a more comprehensive picture of the taxonomic and functional diversity for the key bioprocesses of nitrogen metabolism in the sludge samples.

Comparative genomic analyses

The publication of the first draft genome of *Candidatus* Kuenenia stuttgartiensis, which was assembled from a lab-scale enrichment culture, resulted in a new hypothesis on how anaerobic ammonium oxidation would function (Strous et al., 2006). Based on this genome information, using physiological and biochemical experiments and metaproteomic and metatranscriptomic analyses, the molecular mechanism of anaerobic ammonium oxidation was further delineated (Kartal et al., 2011b). Recently, the draft genome of *Candidatus* Scalindua profunda, prominent in a marine anammox enrichment culture, was also constructed (van de Vossenberg et al., 2013). Here we compared the genome assemblies of *Candidatus* Kuenenia from the three sludge samples to the published draft genome of *Candidatus* Kuenenia stuttgartiensis. In general, *Candidatus* Kuenenia sp. from the three samples was highly similar to the reference and to each other. More than 90% of the genes in the three genomic assemblies, respectively, were found in the reference genome (Table 2). However, due to the fragmented nature of the assemblies, lower coverage of the reference genome was observed in the samples. There was also high degree of collinearity between the metagenome assemblies and the reference

12

genome, and only limited genome-wise rearrangements were observed (Figure 7a).

We distinguished the common genes (i.e. genes shared between samples) and the sample specific genes (i.e. genes exclusively hosted by respective samples) among the three sludges (Figure 7b). COG category analysis showed that the three *Candidatus* Kuenenia species obtained from the sludges had genes unique to each other in most of the functional categories. In particular, the granular *Candidatus* Kuenenia had more specific genes for the functions of inorganic ion and amino acid transport and metabolism, signal transduction, and defense mechanisms. This observation suggests evidence of intensive interactions of this anammox bacterium with coexisting microorganisms. In contrast, the *Candidatus* Kuenenia in the biofilm and floc sludges had more genes unique for them involved in lipid, coenzyme, carbohydrate and nucleotide transport and metabolism, and for energy conversion. This implyies a more diverse cellular metabolism in the IFAS sludges. Of the genome assemblies of *Candidatus* Brocadia detected in the sludges, these had had very low coverage rates (4.7~7.2%) to the reference genome *Candidatus* 'Brocadia sinica' (Oshiki et al., 2015) (Figure S5).

A metagenomics-assisted optimisation of the Anammox process

Anammox is a promising energy-efficient nitrogen removal process (Kartal et al., 2010), however, there is interest to improve the performance of these systems (Wett et al., 2013). Currently, we know relatively little about the regulation and adaptation of the microbial communities present in reactors with different anammox aggregate morphologies. In this study, high-throughput sequencing was used to explore microbial community structure, the relationships between the key microorganisms, and the nitrogen metabolic pathways in the three different anammox microbial aggregates. In the IFAS reactor it was demonstrated that nitrogen removal using the synergy of aerobic ammonium oxidation and anaerobic ammonium oxidation can be successfully achieved in a hybrid system, where both flocculent sludge aggregates enriched with AOB and attached biofilm enriched with anammox bacteria

13

could form a cooperative relationship. Granular sludge in the UASB reactor also harbored abundant anammox bacteria to achieve autotrophic nitrogen removal. Like floccular sludge, granules are a suspended biofilm. As biofilms are characterized by substrate diffusion limitations and redox variations, these can harbor microorganisms with very different growth kinetics and different electron acceptor preferences (Hall-Stoodley et al., 2004). Since anammox microorganisms are known to grow slowly and are inhibited by oxygen, the development of biofilm or granule structure would be critical to retain anammox microorganisms in high abundance. To date, there are some attempts to optimize the operation of the anammox process based on microbial community information obtained from bioreactors. With the decreasing costs of metagenomic sequencing and with the development of diagnostic tools based on metagenomic approaches, it will be promising to monitor the biomass status in reactors systems in the future. Together with detailed reactor performance data, a metagenomics-assisted optimisation approach should be applied to further improve existing anammox reactor robustness.

Experimental Procedures

Reactors operation and anammox microbial aggregates sampling

A pilot-scale single-stage autotrophic nitrogen removal reactor with a working volume of 12 m^3 was setup at the site of a full-scale WWTP in Beijing, China. This IFAS reactor applied the combined activated sludge and biofilm systems (Figure S6). It was operated in continuous mode with polyethylene (PE) sponges as the biomass carriers. The DO was maintained at 0.5-1.5 mg/L through an aeration system that introduced air at the bottom of the reactor. The influent domestic wastewater for the reactor was collected from a primary sedimentation tank of the WWTP. Ammonium carbonate was added to the domestic wastewater to obtain an ammonium concentration of 600−1000 mg/L. The mean concentrations of chemical oxygen demand (COD) and phosphorus were 200 and 5 mg/L, respectively. After long-term

14

cultivation, this reactor achieved stable autotrophic nitrogen removal mediated by AOB and anammox bacteria.

A lab-scale UASB reactor with a working volume of 8.0 L was installed in the same full-scale WWTP as above. This reactor was fed with the same domestic wastewater influent with additions so final concentrations of ammonium and nitrate were 150-300 mg/L and 150-330 mg/L respectively. Before sampling for DNA extractions, the UASB had been operated for longer than two years and achieved stable autotrophic nitrogen removal performance driven by anammox bacteria.

Anammox sludge samples were collected from the pilot-scale IFAS nitrogen removal reactor for flocs and biofilm, and from the lab-scale UASB anammox reactor for granules respectively. The samples were pooled from 5 specimens taken from different parts of the reactors during the steady-state operational phase of the reactors. Microbial cells were pelleted by centrifugation at 10,000 x g for 3 minutes. The supernatant was discarded and the cell pellets were stored at -80 $^{\circ}$C until DNA extraction.

DNA extraction, library preparation, high throughout sequencing and read processing

Genomic DNA extraction was conducted quickly after sampling using the FastDNA SPIN Kit for Soil (QBIOgene Inc., Carlsbad, CA, USA), according to the manufacturer's instructions. Following DNA extraction, the integrity of the DNA was verified using gel electrophoresis and concentration and purity were determined using a Qubit Fluorometer (Thermo, USA). Illumina shotgun DNA library construction and sequencing was conducted by the Beijing Genomic Institute at Shenzhen, China. Specifically, after fragmentation, paired-end fragment library in length of ~170 bp was constructed. Adaptor-appended fragments were sequenced on the Illumina HiSeq 2000 platform and reads of average length of 90 bp for each end were generated. Reads were excluded from further analysis if they: were shorter than 35 bp, had more than 3 ambiguous nucleotides, had 15 bp or more overlapping regions with adapter sequences, had more than 36 nucleotides with a quality value lower than 20, or were potential

duplicated reads from amplification artifacts. The sequences were deposited to the MG-RAST sever with accession numbers 4536155.3, 4536156.3 and 4536157.3.

Read assembly and gene prediction from contigs

After quality filtering describe above, the cleaned sequence reads were assembled into contigs using SOAPdenovo assembler (v 1.05, with settings of -p 8 -F -M 3 -D 1 -L 90 -u) (Li et al., 2010). To obtain the maximum contig lengths and the maximum values of N50 (Miller et al., 2010), a range of k-mer size values (23−57) were investigated (Figure S7). This analysis showed that the optimal K value was '27' for all three samples. Only contigs longer than 500 bp were used for further analysis. SOAP2 (Li et al., 2008) was then used to align all reads to the contigs for each sample (with settings of -s 40 −v 5 –l 32 −r 2 –m 130 –x 210 –M 4 –p 1) (Li et al., 2009). The successfully aligned reads were assigned as 'assembled reads'.

Open reading frames (ORFs) were predicted from contigs using MetaGeneMark (version 2.10) using default settings (Zhu et al., 2010). The predicted ORFs longer than 100 nt were translated into protein sequences based on the NCBI translation table 11 (Junjie et al., 2010). CD-HIT (version 4.6.1), was then used to remove 'redundant' (or highly similar) sequences, and determine gene abundance and statistics among the samples (Li and Godzik, 2006). All of the predicted ORFs were imported into CD-HIT. Reads were mapped back to the non-redundant ORF set for each sample and the coverage for each ORF was calculated as the number of mapped reads.

Taxonomic classification of reads

Taxonomic classification of clean reads was conducted firstly by searching the reads against the NCBI NT database (10 May 2013) using SOAP2 (v2.21, with settings of -s 40 -v 5 -l 32 -r 2 -m 130 -x 210 -M 4 -p 1) (Li et al., 2009), and then further using MEGAN (version 5.3) with the default settings (Huson and Weber, 2013). The sequences were assigned to NCBI taxa using the lowest common ancestor algorithm and the default parameters. To predict the phylogenetic origins of the functional genes, their protein sequences were searched against the

NCBI-NR database (10 May 2013) using BLASTP (version 2.2.23) (Altschul et al., 1990). Sequences were assigned to NCBI taxonomies with MEGAN by using the lowest common ancestor algorithm and the default parameters.

Functional annotation

Protein sequences of the predicted genes were used to search against the NCBI NR (10 May 2013), the Non-supervised Orthologous Groups (eggNOG, version 3.0) (Powell et al., 2012) and against the Kyoto Encyclopedia of Genes and Genomes (KEGG, version 59) (Tatusov et al., 2000; Kanehisa et al., 2006) databases using BLASTP with the E-value cutoff of 10^{-5}. The abundance of a certain COG or KEGG entry in each sample was calculated by the total number of found genes weighted by their coverage. Ammonia monooxygenase (Amo), hydrazine synthase (*hzs*), nitrate reductase (*nar*), periplasmic nitrate reductase (*nap*), nitrite reductase (*nir*), nitric oxide reductase (*nor*), nitrous oxide reductase (*nos*) and dissimilatory nitrite reductase (*nrf*) lacked good representative sequences in the eggNOG and KEGG databases at the time of this study. To accurately detect them in our samples, we manually obtained their abundance from the NCBI NR database search results. Consensus of function was determined by considering the results from all the three databases. Hydrazine dehhydrogenase (Hdh) is currently the only hydroxylamine oxidoreductase (HAO)-like protein that is determined to be a functional hydrazine dehydrogenase. Consequently we manually expanded a Hdh sequence database by using kustc0694 sequence (Kartal et al., 2011b; Kartal et al., 2013) as the seed to recruit sequences from the NCBI NT database based on BLASTP search with similarity cutoff of > 80%. This database was then used to identify potential Hdh sequences in our samples by using BLASTP with similarity cutoff of > 80%.

Sample clustering by functional profiles

COG and KEGG annotations were first filtered to remove potential background noise as described here. A COG entry was selected only if it had no less than ten hits in each of the three samples. For KO entries, no less than five hits in each sample were required. After noise

filtering, only COG (129) or KO (214) entries with standard deviation bigger than 30% were selected and standardized to *z* scores before heatmap plotting and clustering using Cluster 3 (Eisen et al., 1998).

Taxonomic analysis of key enzymes in nitrogen metabolisms

Nucleotide sequences of genes encoding selected key enzymes based on the above KEGG and eggNOG annotations were searched against the NCBI NR database (9 Jun 2014) using BLASTX (version 2.2.23). The top 50 hits with E-values < 0.001 were imported to MEGAN. The LCA parameters were set at a Min Score of 30, Top Percent as 1, and Min Support as 1. Potential miss-assignments were corrected manually. Sequences were only classified to the family level to ensure accuracy. The abundance of each enzyme in a certain sample was calculated by multiplying the genes annotated as this enzyme with the abundance of the gene. The reads per kilo-basepairs per 10 million reads (RPK10M) was then calculated for each gene.

Comparative genomic analyses

Contigs were searched against the NCBI NT database using MegaBLAST (version 2.2.23). The taxonomic assignment of the contigs was then obtained using MEGAN. The contigs assigned to *Candidatus* Kuenenia were then compared to the reference genomic sequence of *Candidatus* Kuenenia stuttgartiensis using BLASTN (version 2.2.23) to calculate the collinearity between genomes/assemblies. Coverage between genomes/assemblies were calculated based on the result of gene mapping using BLASTP and it was considered as successful mapping when amino acid sequences of two genes shared similarity of > 80%. Core genes were those that shared 80% or more similarity from the Kuenenia assemblies between samples, while the rest were considered as sample specific Kuenenia genes.

AOB, NOB and Anammox abundance analysis

Two approaches were applied to quantify the abundances of AOB, NOB and Anammox in

these three samples. The first method was based on metagenomic sequencing data. According to taxonomic classification results, the annotated read numbers of AOB, NOB and anammox bacteria was counted and the percentage of each group was obtained by normalizing to the number of total bacterial reads in each sample. The second approach was based on qPCR. Genes targeting AOB, NOB (*Nitrobacter and Nitrospira*), anammox bacteria, and the universal bacterial 16S rRNA genes for the total bacteria were quantified using an ABI7300 Real Time PCR system (ABI, USA) using specific primer sets and thermal profiles for the reactions (Table S9) (Zhu et al., 2013). Standard plasmids carrying the target genes were obtained by TA cloning and these were extracted using a TIANpure Mini Plasmid kit (Tiangen, China). For the qPCR the 25 μL each reaction contained 1×Sybr Green I, 1×Dye (Takara), 200 nM of each primer, 0.5 mg mL^{-1} BSA, and 2 μL of the DNA template. Triplicate assays were conducted for the decimally diluted standard plasmids, the appropriately diluted samples, and the negative controls. Standard curves covered 5-8 orders of magnitude with R^2 higher than 0.99. The amplification efficiencies were based on slopes between 93.4% and 96.5% (Table S10).

Chemical analysis and floc size measurement

The reactor solutions were analysed for COD, NH_4^+, NO_3^-, NO_2^-, MLSS and volatile suspended solids (VSS). These were measured according to standard methods (APHA, 1995). The total nitrogen (TN) was measured using a Multi N/C 3100 meter (Jena, Germany). The volumetric floc and granule size distributions were measured using a Malvern Mastersizer 2000 (Malvern Instruments Ltd., UK). The Malvern Mastersizer 2000 uses light scattering and returns a volume fraction for each of the 100 size bands of between 0.01 and 10,000 μm.

Conflict of Interest

The authors declare no conflict of interest.

19

Supplementary Information accompanies this paper on Environmental Microbiology website.

References

Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., and Gnirke, A. (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* **12**: R18.

Albertsen, M., Hansen, L.B.S., Saunders, A.M., Nielsen, P.H., and Nielsen, K.L. (2012) A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. *ISME J* **6**: 1094-1106.

Almstrand, R., Persson, F., Daims, H., Ekenberg, M., Christensson, M., Wilen, B.M., Sorensson, F., and Hermansson, M. (2014) Three-dimensional stratification of bacterial biofilm populations in a moving bed biofilm reactor for nitritation-Anammox. *Int J Mol Sci* **15**: 2191-2206.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

APHA (1995) *Standard Methods for the Examination for Water and Wastewater American Public Health Association* Washington, DC.: American Public Health Association.

Buergmann, H., Jenni, S., Vazquez, F., and Udert, K.M. (2011) Regime shift and microbial dynamics in a sequencing batch reactor for nitrification and anammox treatment of urine. *Appl Environ Microb* **77**: 5897-5907.

Chen, X., Guo, J., Shi, Y., Hu, S., Yuan, Z., and Ni, B.-J. (2014) Modeling of simultaneous anaerobic methane and ammonium oxidation in a membrane biofilm reactor. *Environ Sci Technol* **48**: 9540-9547.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *P Natl Acad Sci USA* **95**: 14863-14868.

Gori, F., Tringe, S.G., Kartal, B., Marchiori, E., and Jetten, M.S.M. (2011) The metagenomic basis of anammox metabolism in Candidatus 'Brocadia fulgida'. *Biochem Soc T* **39**: 1799-1804.

Guo, J., Peng, Y., Wang, S., Ma, B., Ge, S., Wang, Z., Huang, H., Zhang, J., and Zhang, L. (2013) Pathways and organisms involved in ammonia oxidation and nitrous oxide emission. *Crit Rev Env Sci Tec* **43**: 2213-2296.

Hall-Stoodley, L., Costerton, J.W., and Stoodley, P. (2004) Bacterial biofilms: From the natural environment to infectious diseases. *Nature Reviews Microbiology* **2**: 95-108.

Han, P., Huang, Y.T., Lin, J.G., and Gu, J.D. (2013) A comparison of two 16S rRNA gene-based PCR primer sets in unraveling anammox bacteria from different environmental samples. *Appl Microbiol Biot* **97**: 10521-10529.

Hu, B.-L., Zheng, P., Tang, C.-J., Chen, J.-w., van der Biezen, E., Zhang, L., Ni, B.-J., Jetten, M.S.M., Yan, J., Yu, H.-Q., and Kartal, B. (2010) Identification and quantification of anammox bacteria in eight nitrogen removal reactors. *Water Res* **44**: 5014-5020.

Hu, Z., Lotti, T., van Loosdrecht, M., and Kartal, B. (2013a) Nitrogen removal with the anaerobic ammonium oxidation process. *Biotechnol Lett* **35**: 1145-1154.

Hu, Z., Speth, D.R., Francoijs, K.-J., Quan, Z.-X., and Jetten, M. (2012) Metagenome analysis of a complex community reveals the metabolic blueprint of anammox bacterium '*Candidatus* Jettenia asiatica'. *Front Microbiol* **3**: 366.

Hu, Z., Lotti, T., de Kreuk, M., Kleerebezem, R., van Loosdrecht, M., Kruit, J., Jetten, M.S.M., and Kartal, B. (2013b) Nitrogen Removal by a Nitritation-Anammox Bioreactor at Low Temperature. *Appl Environ Microb* **79**: 2807-2812.

Huson, D.H., and Weber, N. (2013) Microbial community analysis using MEGAN. In *Microbial metagenomics, metatranscriptomics, and metaproteomics*. DeLong, E.F. (ed), pp. 465-485.

Jetten, M., Camp, H., Kuenen, J., and Strous, M. (2010) Description of the order brocadiales. *Krieg, NR; Ludwig, W; Whitman, WB [et al](ed), The bacteroidetes, spirochaetes, tenericutes (mollicutes), acidobacteria, fibrobacteres, fusobacteria, dictyoglomi, gemmatimonadetes, lentisphaerae, verrucomicrobia, chlamydiae, and planctomycetes*: 596-603.

Jetten, M.S., Niftrik, L.v., Strous, M., Kartal, B., Keltjens, J.T., and Op den Camp, H.J. (2009) Biochemistry and molecular biology of anammox bacteria. *Crit Rev Biochem Mol* **44**: 65-84.

Junjie, Q., Ruiqiang, L., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Junhua, L., Junming, X., Shaochuan, L., Dongfang, L., Jianjun, C., Bo, W., Huiqing, L., Huisong, Z., Yinlong, X., Tap, J., Lepage, P., Bertalan, M., Batto, J.M., Hansen, T., Paslier, D.l., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Hongmei, Z., Chang, Y., Shengting, L., Min, J., Yan, Z., Yingrui, L., Xiuqing, Z., Songgang, L., Nan, Q., Huanming, Y., Jian, W., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S.D., Jun, W., and Meta, H.I.T.C. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature, UK* **464**: 59-65.

Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* **34**: D354-D357.

Kartal, B., Kuenen, J.G., and van Loosdrecht, M.C.M. (2010) Sewage Treatment with Anammox. *Science* **328**: 702-703.

Kartal, B., Geerts, W., and Jetten, M.S.M. (2011a) Cultivation, detection, and ecophysiology of anaerobic ammonium-oxidizing bacteria. In *Methods in Enzymology: Research on Nitrification and Related Processes, Vol 486, Part A*. Klotz, M.G. (ed). San Diego: Elsevier Academic Press Inc, pp. 89-108.

Kartal, B., van Niftrik, L., Keltjens, J.T., den Camp, H., and Jetten, M.S.M. (2012) Anammox-Growth Physiology, Cell Biology, and Metabolism. In *Advances in Microbial Physiology, Vol 60*. Poole, R.K. (ed). London: Academic Press Ltd-Elsevier Science Ltd, pp. 211-262.

Kartal, B., de Almeida, N.M., Maalcke, W.J., Op den Camp, H.J.M., Jetten, M.S.M., and Keltjens, J.T. (2013) How to make a living from anaerobic ammonium oxidation. *Fems Microbiology Reviews* **37**: 428-461.

Kartal, B., Maalcke, W.J., de Almeida, N.M., Cirpus, I., Gloerich, J., Geerts, W., den Camp, H.J.M.O., Harhangi, H.R., Janssen-Megens, E.M., Francoijs, K.-J., Stunnenberg, H.G., Keltjens, J.T., Jetten, M.S.M., and Strous, M. (2011b) Molecular mechanism of anaerobic ammonium oxidation. *Nature* **479**: 127-U159.

Kindaichi, T., Tsushima, I., Ogasawara, Y., Shimokawa, M., Ozaki, N., Satoh, H., and Okabe, S. (2007) In situ activity and spatial organization of anaerobic ammonium-oxidizing (anammox) bacteria in biofilms. *Appl Environ Microb* **73**: 4931-4939.

Kraft, B., Tegetmeyer, H.E., Sharma, R., Klotz, M.G., Ferdelman, T.G., Hettich, R.L., Geelhoed, J.S., and Strous, M. (2014) The environmental controls that govern the end product of bacterial nitrate respiration. *Science* **345**: 676-679.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009) Circos: An information aesthetic for comparative genomics. *Genome Res* **19**: 1639-1645.

Kuenen, J.G. (2008) Anammox bacteria: from discovery to application. *Nature Reviews Microbiology* **6**: 320-326.

Kuypers, M.M.M., Sliekers, A.O., Lavik, G., Schmid, M., Jorgensen, B.B., Kuenen, J.G., Damste, J.S.S., Strous, M., and Jetten, M.S.M. (2003) Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. *Nature* **422**: 608-611.

Lam, P., and Kuypers, M.M.M. (2011) Microbial Nitrogen Cycling Processes in Oxygen Minimum Zones. In *Annual Review of Marine Science, Vol 3*. Carlson, C.A., and Giovannoni, S.J. (eds), pp. 317-345.

Li, M., and Gu, J.D. (2011) Advances in methods for detection of anaerobic ammonium oxidizing (anammox) bacteria. *Appl Microbiol Biot* **90**: 1241-1252.

Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008) SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**: 713-714.

Li, R., Yu, C., Li, Y., Lam, T.-W., Yiu, S.-M., Kristiansen, K., and Wang, J. (2009) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**: 1966-1967.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J., and Wang, J. (2010) De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res* **20**: 265-272.

Li, W., and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658-1659.

Miller, J.R., Koren, S., and Sutton, G. (2010) Assembly algorithms for next-generation sequencing data. *Genomics* **95**: 315-327.

Mulder, A., van de Graaf, A.A., Robertson, L.A., and Kuenen, J.G. (1995) Anaerobic ammonium oxidation discovered in a denitrifying fluidized bed reactor. *FEMS Microbiol Ecol* **16**: 177-183.

Ni, B.-J., Chen, Y.-P., Liu, S.-Y., Fang, F., Xie, W.-M., and Yu, H.-Q. (2009) Modeling a granule-based anaerobic ammonium oxidizing (ANAMMOX) process. *Biotechnol Bioeng* **103**: 490-499.

Ni, B.-J., Hu, B.-L., Fang, F., Xie, W.-M., Kartal, B., Liu, X.-W., Sheng, G.-P., Jetten, M., Zheng, P., and Yu, H.-Q. (2010) Microbial and physicochemical characteristics of compact anaerobic ammonium-oxidizing granules in an upflow anaerobic sludge blanket reactor. *Appl Environ Microb* **76**: 2652-2656.

Nielsen, M., Bollmann, A., Sliekers, O., Jetten, M., Schmid, M., Strous, M., Schmidt, I., Larsen, L.H., Nielsen, L.P., and Revsbech, N.P. (2005) Kinetics, diffusional limitation and microscale distribution of chemistry and organisms in a CANON reactor. *FEMS Microbiol Ecol* **51**: 247-256.

Orphan, V.J. (2009) Methods for unveiling cryptic microbial partnerships in nature. *Curr Opin Microbiol* **12**: 231-237.

Oshiki, M., Shinyako-Hata, K., Satoh, H., and Okabe, S. (2015) Draft Genome Sequence of an Anaerobic Ammonium-Oxidizing Bacterium, "Candidatus Brocadia sinica". *Genome announcements* **3**.

Pavlekovic, M., Schmid, M.C., Schmider-Poignee, N., Spring, S., Pilhofer, M., Gaul, T., Fiandaca, M., Loffler, F.E., Jetten, M., Schleifer, K.H., and Lee, N.M. (2009) Optimization of three FISH procedures for in situ detection of anaerobic ammonium oxidizing bacteria in biological wastewater treatment. *J Microbiol Meth* **78**: 119-126.

Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., Arnold, R., Rattei, T., Letunic, I., Doerks, T., Jensen, L.J., von Mering, C., and Bork, P. (2012) eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res* **40**: D284-D289.

Quan, Z.X., Rhee, S.K., Zuo, J.E., Yang, Y., Bae, J.W., Park, J.R., Lee, S.T., and Park, Y.H. (2008) Diversity of ammonium-oxidizing bacteria in a granular sludge anaerobic ammonium-oxidizing (anammox) reactor. *Environ Microbiol* **10**: 3130-3139.

Rittmann, B., and McCarty, P. (2001) *Environmental biotechnology: Principles and applications*. New York: McGraw-Hill Book Company.

Schmid, M., Walsh, K., Webb, R., Rijpstra, W.I.C., van de Pas-Schoonen, K., Verbruggen, M.J., Hill, T., Moffett, B., Fuerst, J., Schouten, S., Damste, J.S.S., Harris, J., Shaw, P., Jetten, M., and Strous, M. (2003) Candidatus "Scalindua brodae", sp nov., Candidatus "Scalindua wagneri", sp nov., two new species of anaerobic ammonium oxidizing bacteria. *Syst Appl Microbiol* **26**: 529-538.

Schmid, M.C., Maas, B., Dapena, A., de Pas-Schoonen, K.V., de Vossenberg, J.V., Kartal, B., van Niftrik, L., Schmidt, I., Cirpus, I., Kuenen, J.G., Wagner, M., Damste, J.S.S., Kuypers, M., Revsbech, N.P., Mendez, R., Jetten, M.S.M., and Strous, M. (2005) Biomarkers for in situ detection of anaerobic ammonium-oxidizing (anammox) bacteria. *Appl Environ Microb* **71**: 1677-1684.

Schmidt, I., Sliekers, O., Schmid, M., Bock, E., Fuerst, J., Kuenen, J.G., Jetten, M.S.M., and Strous, M. (2003) New concepts of microbial treatment processes for the nitrogen removal in wastewater. *FEMS Microbiol Rev* **27**: 481-492.

Simon, J. (2002) Enzymology and bioenergetics of respiratory nitrite ammonification. *FEMS Microbiol Rev* **26**: 285-309.

Strous, M., VanGerven, E., Zheng, P., Kuenen, J.G., and Jetten, M.S.M. (1997) Ammonium removal from concentrated waste streams with the anaerobic ammonium oxidation (anammox) process in different reactor configurations. *Water Res* **31**: 1955-1962.

Strous, M., Fuerst, J.A., Kramer, E.H.M., Logemann, S., Muyzer, G., van de Pas-Schoonen, K.T., Webb, R., Kuenen, J.G., and Jetten, M.S.M. (1999) Missing lithotroph identified as new planctomycete. *Nature* **400**: 446-449.

Strous, M., Pelletier, E., Mangenot, S., Rattei, T., Lehner, A., Taylor, M.W., Horn, M., Daims, H., Bartol-Mavel, D., Wincker, P., Barbe, V., Fonknechten, N., Vallenet, D., Segurens, B., Schenowitz-Truong, C., Medigue, C., Collingro, A., Snel, B., Dutilh, B.E., Op den Camp, H.J.M., van der Drift, C., Cirpus, I., van de Pas-Schoonen, K.T., Harhangi, H.R., van Niftrik, L., Schmid, M., Keltjens, J., van de Vossenberg, J., Kartal, B., Meier, H., Frishman, D., Huynen, M.A., Mewes, H.W., Weissenbach, J., Jetten, M.S.M., Wagner, M., and Le Paslier, D. (2006) Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440**: 790-794.

Tal, Y., Watts, J.E.M., Schreier, S.B., Sowers, K.R., and Schreier, H.J. (2003) Characterization of the microbial community and nitrogen transformation processes associated with moving bed bioreactors in a closed recirculated mariculture system. *Aquaculture* **215**: 187-202.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**: 33-36.

van de Vossenberg, J., Rattray, J.E., Geerts, W., Kartal, B., van Niftrik, L., van Donselaar, E.G., Damste, J.S.S., Strous, M., and Jetten, M.S.M. (2008) Enrichment and characterization of marine anammox bacteria associated with global nitrogen gas production. *Environ Microbiol* **10**: 3120-3129.

van de Vossenberg, J., Woebken, D., Maalcke, W.J., Wessels, H., Dutilh, B.E., Kartal, B., Janssen-Megens, E.M., Roeselers, G., Yan, J., Speth, D., Gloerich, J., Geerts, W., van der Biezen, E., Pluk, W., Francoijs, K.J., Russ, L., Lam, P., Malfatti, S.A., Tringe, S.G., Haaijer, S.C.M., Op den Camp, H.J.M., Stunnenberg, H.G., Amann, R., Kuypers, M.M.M., and Jetten, M.S.M. (2013) The metagenome of the marine anammox bacterium 'Candidatus Scalindua profunda' illustrates the versatility of this globally important nitrogen cycle bacterium. *Environ Microbiol* **15**: 1275-1289.

van der Star, W.R.L., Miclea, A.I., van Dongen, U.G.J.M., Muyzer, G., Picioreanu, C., and van Loosdrecht, M.C.M. (2008) The membrane bioreactor: A novel tool to grow anammox bacteria as free cells. *Biotechnol Bioeng* **101**: 286-294.

Vanwonterghem, I., Jensen, P.D., Ho, D.P., Batstone, D.J., and Tyson, G.W. (2014a) Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. *Curr Opin Biotech* **27**: 55-64.

Vanwonterghem, I., Jensen, P.D., Dennis, P.G., Hugenholtz, P., Rabaey, K., and Tyson, G.W. (2014b) Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters. *ISME J* **8**: 2015-2028.

Vlaeminck, S.E., Cloetens, L.F.F., Carballa, M., Boon, N., and Verstraete, W. (2008) Granular biomass capable of partial nitritation and anammox. *Water Sci Technol* **58**: 1113-1120.

Vlaeminck, S.E., Terada, A., Smets, B.F., De Clippeleir, H., Schaubroeck, T., Bolca, S., Demeestere, L., Mast, J., Boon, N., Carballa, M., and Verstraete, W. (2010) Aggregate Size and Architecture Determine Microbial Activity Balance for One-Stage Partial Nitritation and Anammox. *Appl Environ Microb* **76**: 900-909.

Welsh, A., Chee-Sanford, J.C., Connor, L.M., Loffler, F.E., and Sanford, R.A. (2014) Refined NrfA Phylogeny Improves PCR-Based nrfA Gene Detection. *Appl Environ Microb* **80**: 2110-2119.

Wett, B., Omari, A., Podmirseg, S.M., Han, M., Akintayo, O., Brandon, M.G., Murthy, S., Bott, C., Hell, M., Takacs, I., Nyhuis, G., and O'Shaughnessy, M. (2013) Going for mainstream deammonification from bench to full scale for maximized resource efficiency. *Water Sci Technol* **68**: 283-289.

Ye, L., Zhang, T., Wang, T.T., and Fang, Z.W. (2012) Microbial structures, functions, and metabolic pathways in wastewater treatment bioreactors revealed using high-throughput sequencing. *Environ Sci Technol* **46**: 13244-13252.

Yoon, S., Cruz-Garcia, C., Sanford, R., Ritalahti, K.M., and Loffler, F.E. (2014) Denitrification versus respiratory ammonification: environmental controls of two competing dissimilatory NO3-/NO2- reduction pathways in Shewanella loihica strain PV-4. *ISME J*.

Yu, K., and Zhang, T. (2012) Metagenomic and Metatranscriptomic Analysis of Microbial Community Structure and Gene Expression of Activated Sludge. *PLoS One* **7**: e38183.

Zhang, T., Shao, M.-F., and Ye, L. (2012) 454 Pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. *ISME J* **6**: 1137-1147.

Zhu, G., Wang, S., Wang, W., Wang, Y., Zhou, L., Jiang, B., Op den Camp, H.J.M., Risgaard-Petersen, N., Schwark, L., Peng, Y., Hefting, M.M., Jetten, M.S.M., and Yin, C. (2013) Hotspots of anaerobic ammonium oxidation at land-freshwater interfaces. *Nature Geoscience* **6**: 103-107.

Zhu, W., Lomsadze, A., and Borodovsky, M. (2010) Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* **38**: e132.

Titles and legends to figures

Table 1 Sludge sample metagenome sequencing, assembly and annotation statistics

Table 2 The genomic coverage rates for the three samples relatively

Figure 1 Macroscopic morphologies of the three distinct anammox sludge cultures: A, flocs; B, biofilm; and C, granules.

Figure 2 Relative abundances of the Top 24 genera in the three different anammox aggregates. The data were visualized using Circos (Krzywinski et al., 2009). The width of bars from each genus indicates the relative abundances of that phylum in the three samples.

Figure 3 Clustering of the sludge samples based on functional annotation of the genes by COG (A) and KEGG pathways (B) and on the gene abundance. The color code indicates the relative abundance of the genes after z-score normalization.

Figure 4 Abundances of AOB, NOB and Anammox microorganisms in the three samples as determined from metagenome sequences: A, Anammox; B, AOB; C, NOB; D, Key microorganisms involved in nitrogen cycle.

Figure 5 Abundance of key enzymes for anammox, nitrification, denitrification and DNRA. Pathways and key enzymes are colored differentially. The abundance of the enzymes in each sample is shown in the color bars, with the location in bar determining the sludge type with left for flocs, middle for biofilm, and right for granules. AmoA, bacterial ammonia monooxygenase subunit A; Hao, hydroxylamine oxidoreductase; NarG, cytoplasmic nitrate reductase alpha subunit; NapA, periplasmic nitrate reductase precursor; NirK, copper-containing nitrite reductase; NirS, cytochrome cd_1 nitrite reductase; cNor/qNor, nitric oxide reductase cytochrome or quinol-dependent; NosZ, nitrous oxide reductase; NrfA, ammonia-forming dissimilatory nitrite reductase; Hzs, hydrazine synthase; and Hdh, hydrazine dehhydrogenase. RPK10M, hit reads per kilo-basepairs per 10 million reads.

Figure 6 Taxonomic origins of the key enzymes in nitrogen metabolism detected in the anammox sludges. AmoA, bacterial ammonia monooxygenase subunit A; NarG, cytoplasmic nitrate reductase alpha chain; NapA, periplasmic nitrate reductase precursor; NirK, copper nitrite reductase; NirS, cytochrome cd1 nitrite reductase; NorB/NorZ, nitric oxide reductase cytochrome or quinol-dependent; NosZ, nitrous oxide reductase; NrfA, ammonia-forming dissimilatory nitrite reductase; Hzs, hydrazine synthase; and Hdh, hydrazine dehhydrogenase. RPK10M, hit reads per kilo-basepairs per 10 million reads.

Figure 7 (a) Collinearity comparison of the *Candidatus* Kuenenia genome assemblies, obtained from the sludge samples, to the *Candidatus* 'Kuenenia stuttgartiensis' reference genome. (b) COG category analysis of the common genes and the sample specific genes among the three sludge samples.
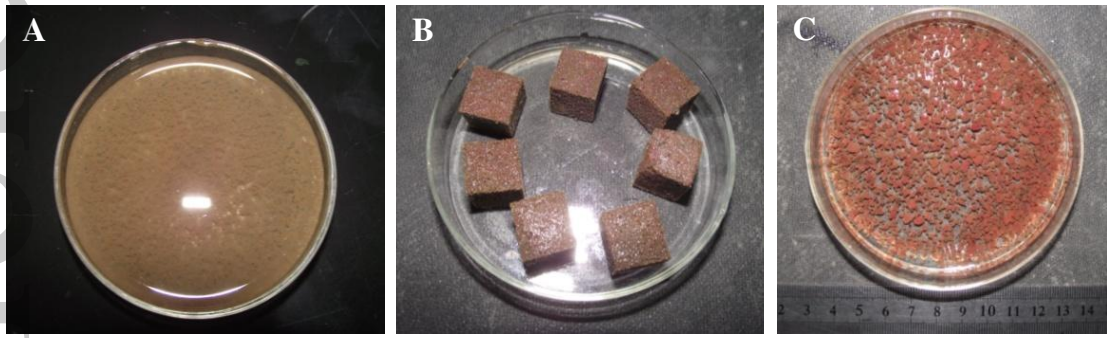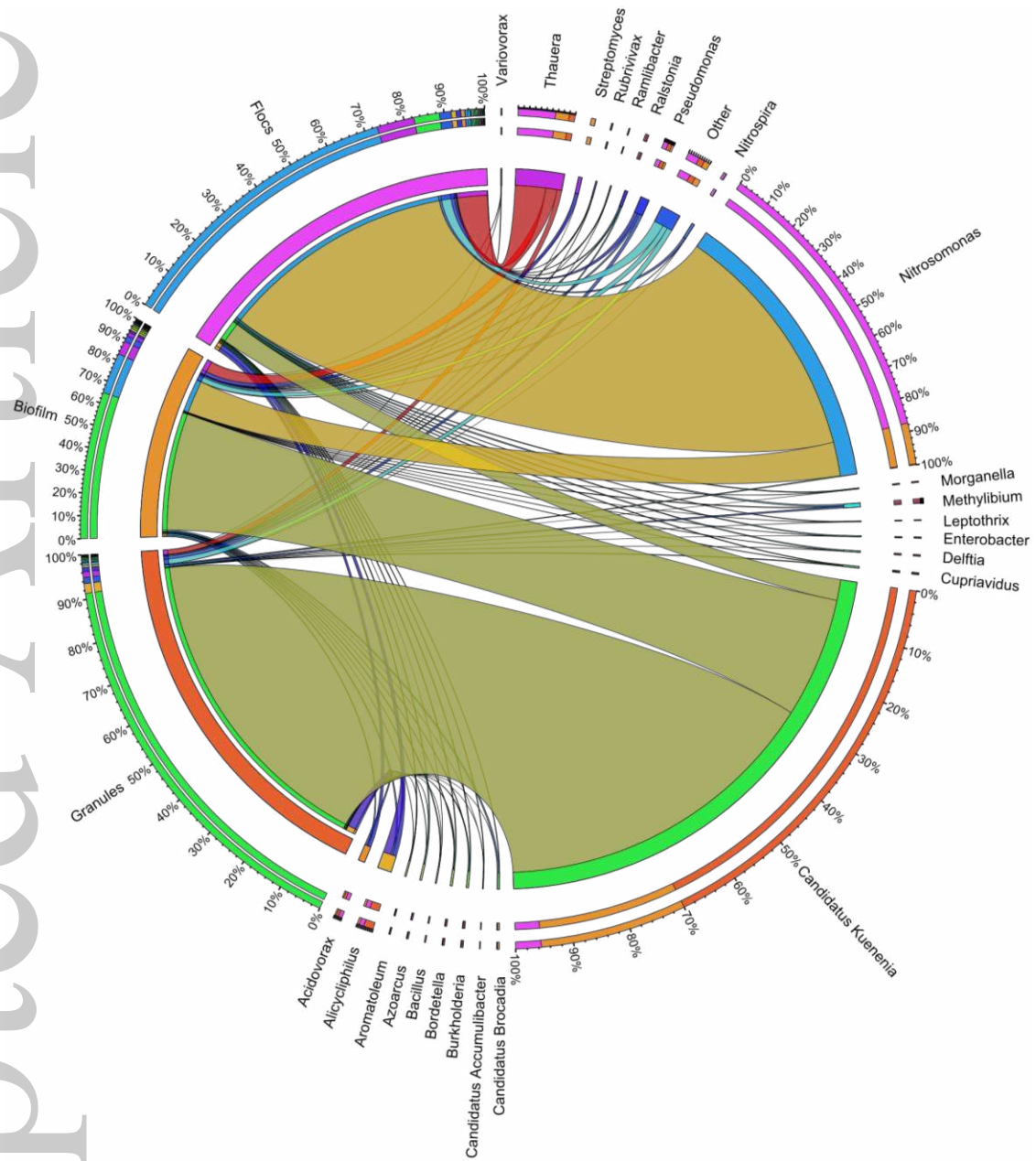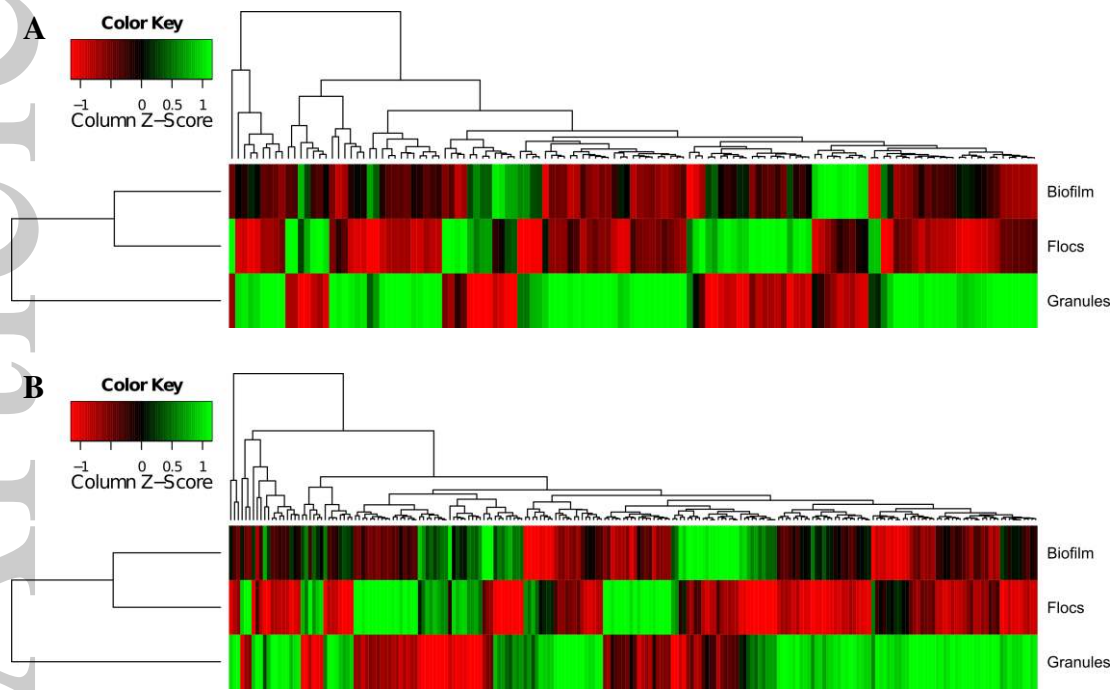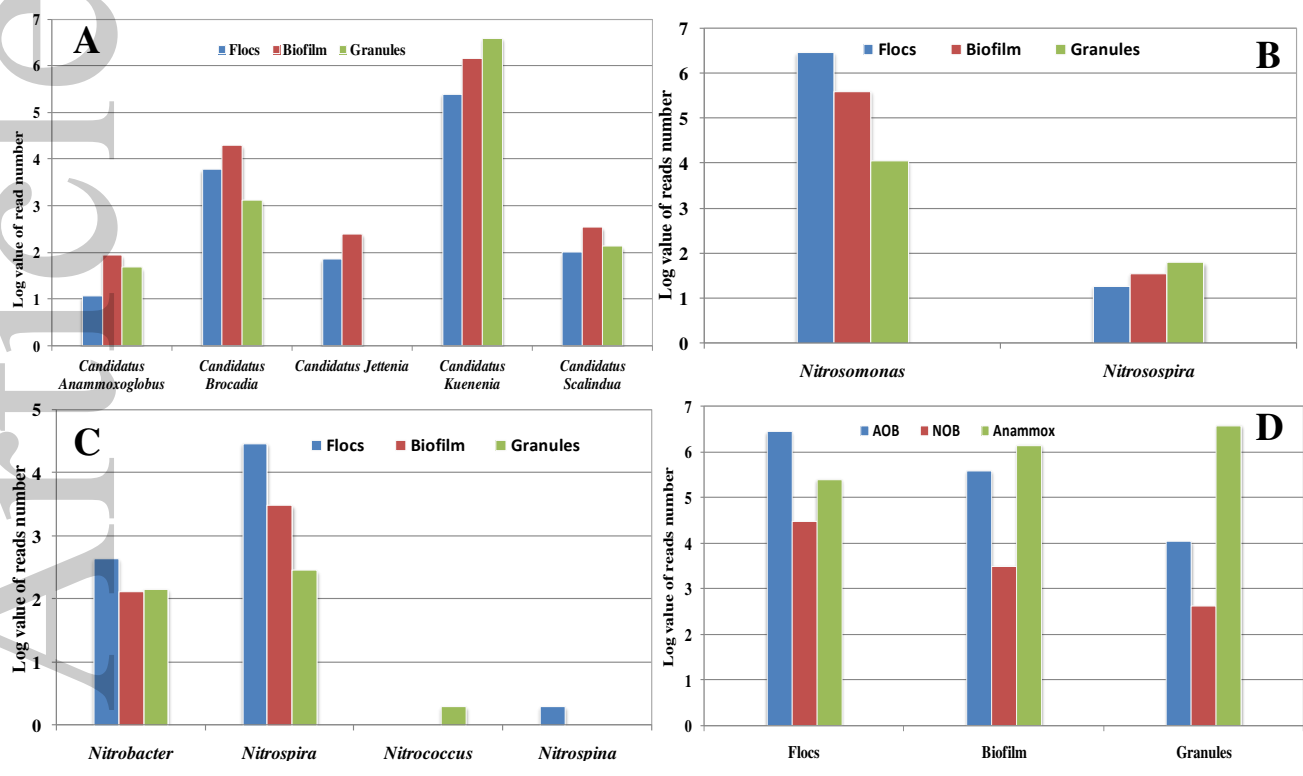
Figure 1

Figure 2

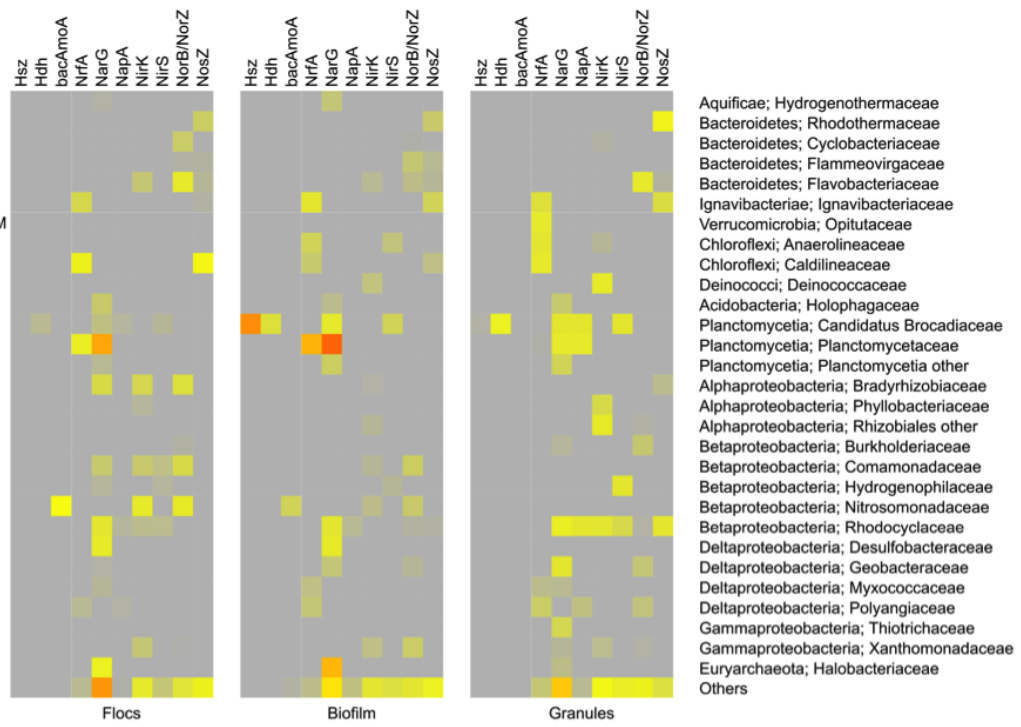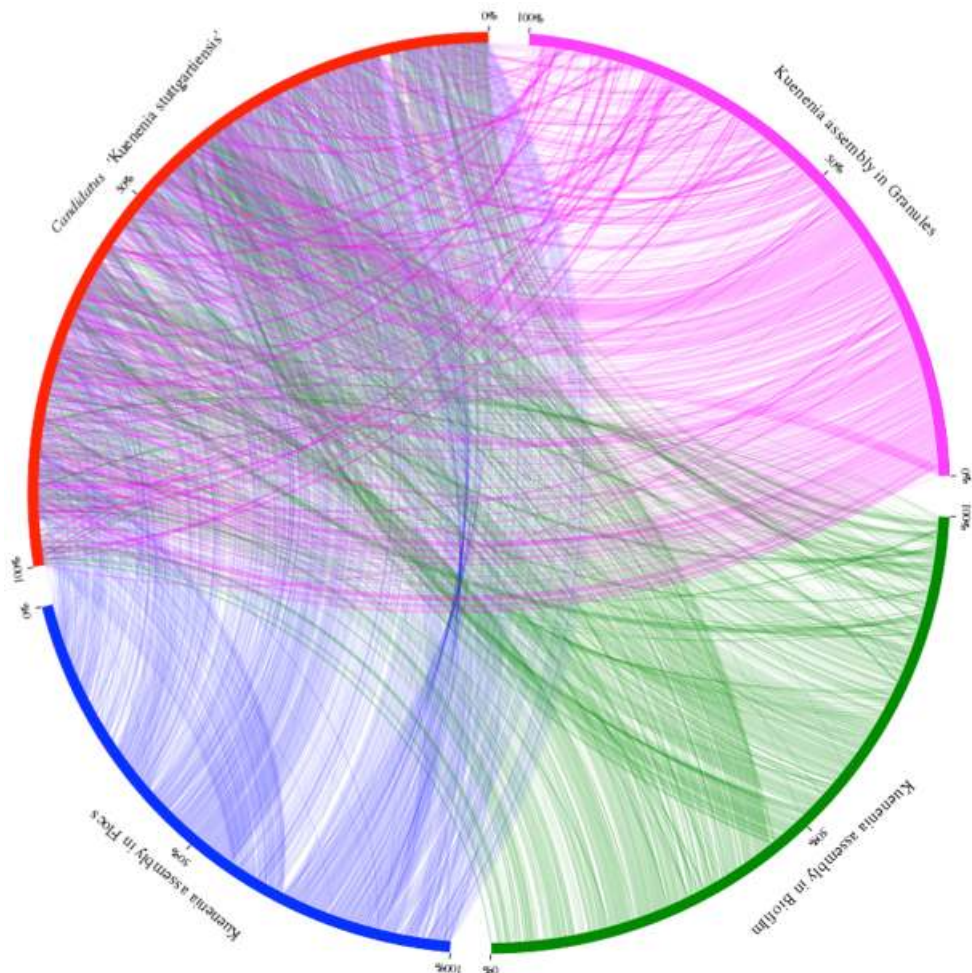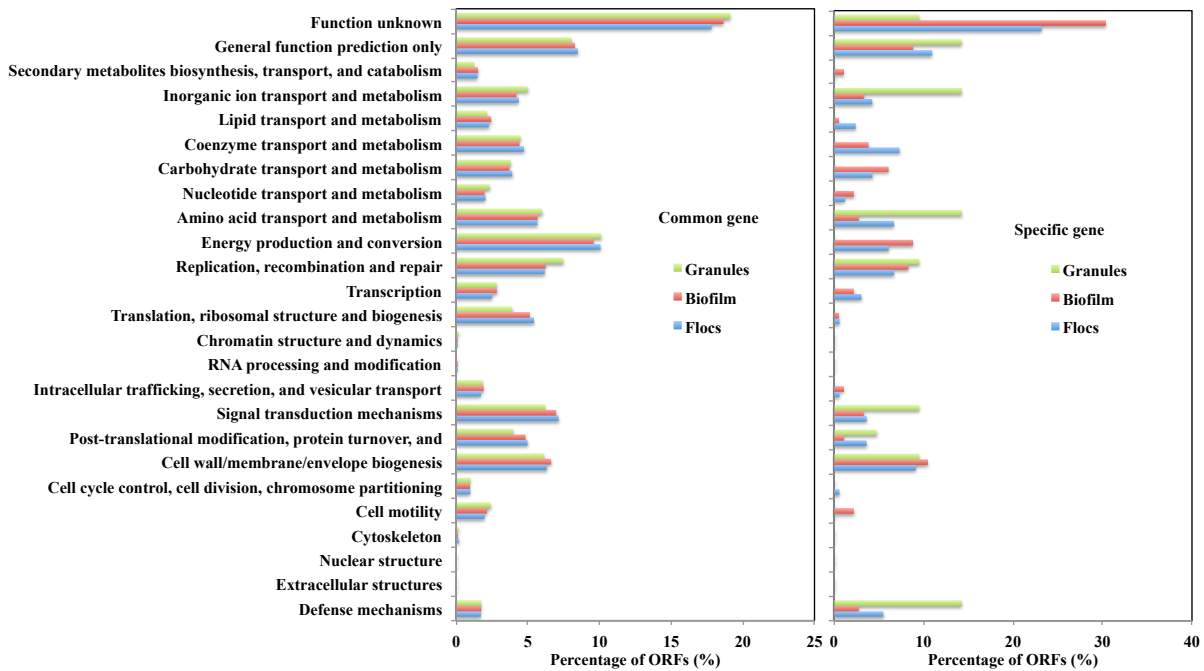**Figure 3**

Figure 4
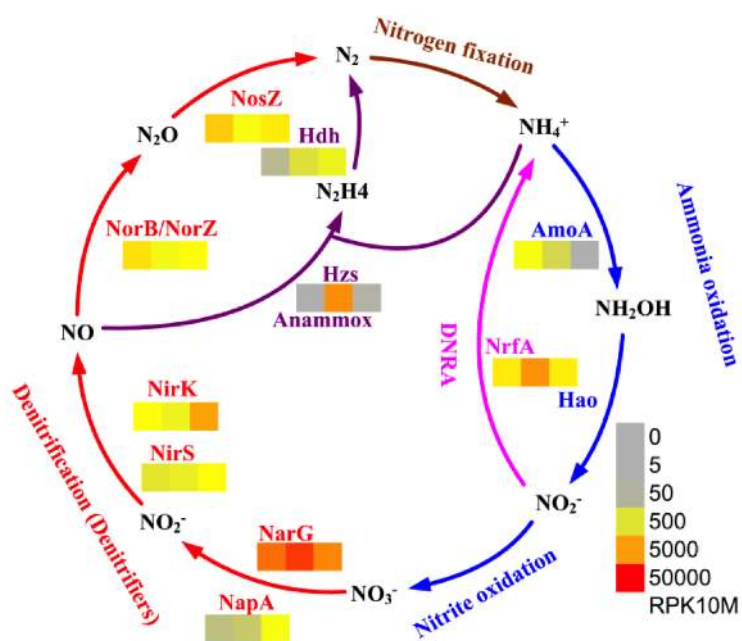
Figure 5

Figure 6

(a)

(b)

Figure 7

Table 1 Sequencing, assembly and annotation statistics

Sample	High quality reads	Taxonomically classified reads (%)	Assembled reads (%)	Contigs > 500bp	Contig N50 (bp)	Reads mapped to contigs (%)	ORFs total	Annotated ORFs by eggNOG (%)	Annotated ORFs by KEGG (%)
Flocs	11,513,079	35.0	50.46	96,488	1494	50.46	183398	82.4	32.7
Biofilm	11,439,096	19.7	54.49	88,007	1650	54.49	170725	94.6	33.6
Granules	14,229,888	30.4	62.8	67,319	2318	62.8	146983	87	34.5

Table 2 The genomic coverage rates for the three samples relatively

Parameters	Samples	*Candidatus* Brocadia	*Candidatus* Kuenenia
N50 (bp)	Flocs	1,399	1,398
	Biofilm	5,740	5,740
	Granules	8,308	8,308
Genome (bp)	Flocs	3,018,134	3,012,785
	Biofilm	3,693,500	3,690,351
	Granules	3,523,381	3,522,726
Assembly Genome Coverage by References (%)	Flocs	7.20	90
	Biofilm	6.75	92
	Granules	4.74	98
References Coverage by Assembly Genome (%)	Flocs	5.48	71
	Biofilm	5.80	85
	Granules	4.71	88