

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/157011>

Please be advised that this information was generated on 2017-12-12 and may be subject to change.

ARTICLE

Received 10 Jul 2015 | Accepted 26 Feb 2016 | Published 31 Mar 2016

DOI: 10.1038/ncomms11172

OPEN

Genome-based microbial ecology of anammox granules in a full-scale wastewater treatment system

Daan R. Speth¹, Michiel H. in 't Zandt¹, Simon Guerrero-Cruz¹, Bas E. Dutilh^{2,3,4,*} & Mike S.M. Jetten^{1,5,*}

Partial-nitritation anammox (PNA) is a novel wastewater treatment procedure for energy-efficient ammonium removal. Here we use genome-resolved metagenomics to build a genome-based ecological model of the microbial community in a full-scale PNA reactor. Sludge from the bioreactor examined here is used to seed reactors in wastewater treatment plants around the world; however, the role of most of its microbial community in ammonium removal remains unknown. Our analysis yielded 23 near-complete draft genomes that together represent the majority of the microbial community. We assign these genomes to distinct anaerobic and aerobic microbial communities. In the aerobic community, nitrifying organisms and heterotrophs predominate. In the anaerobic community, widespread potential for partial denitrification suggests a nitrite loop increases treatment efficiency. Of our genomes, 19 have no previously cultivated or sequenced close relatives and six belong to bacterial phyla without any cultivated members, including the most complete *Omnitrophica* (formerly OP3) genome to date.

¹Department of Microbiology, Institute for Water and Wetland Research, Radboud University, Heyendaalseweg 135, 6525AJ Nijmegen, The Netherlands.

²Theoretical Biology and Bioinformatics, Utrecht University, Padualaan 8, 3584CH Utrecht, The Netherlands. ³Centre for Molecular and Biomolecular Informatics, Radboud University Medical Centre, Geert Grooteplein Zuid 26-28, 6525GA Nijmegen, The Netherlands. ⁴Instituto de Biologia, Universidade Federal do Rio de Janeiro, Avenida Carlos Chagas Filho 373, Rio de Janeiro 21941-902, Brazil. ⁵Department of Biotechnology, Delft University of Technology, Julianalaan 67, 2628BC Delft, The Netherlands. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to D.R.S. (email: d.speth@science.ru.nl).

The vast microbial metabolic diversity is a rich source for industrial application and innovation. One large-scale example is wastewater treatment^{1,2}, where microorganisms remove pollutants, including ammonium, from municipal or industrial wastewater. The paradigm for ammonium removal from wastewater is currently shifting, from removal based on conventional two-step nitrification/denitrification³ to a one-step system based on anaerobic ammonium oxidation (anammox)⁴.

In one-step anammox, also known as partial-nitritation/anammox (PNA), aerobic ammonium-oxidizing bacteria (AOB) oxidize part of the influent ammonium to nitrite. Subsequently, anammox bacteria convert the nitrite and remaining ammonium to dinitrogen gas in the absence of oxygen^{5,6}. The niche differentiation required for these aerobic and anaerobic processes to occur in a single PNA reactor can be achieved in various ways, such as biofilms on a carrier or in granular biomass^{7,8}. In all single-reactor PNA variants, an outer layer of aerobic organisms consumes the available oxygen, leaving the interior of the biofilm or granule anaerobic for anammox⁹.

PNA has several advantages over conventional ammonium removal via nitrification/denitrification. As no additional electron donors such as methanol are required and nitrous oxide (N₂O) is not an intermediate of the anammox process, PNA emits less greenhouse gases than conventional systems¹⁰. In addition, PNA has a lower cost and energy requirement, because the process can take place in a single reactor with limited aeration^{2,11}. In the decade since the first pilot plant, these benefits have resulted in the implementation of the PNA process for nitrogen removal in over 100 full-scale plants around the world¹². However, despite the increasing global importance of PNA systems, a comprehensive study of the microbial community facilitating nitrogen removal is lacking.

Previous studies of the PNA microbial community reported on the organisms responsible for the key processes in PNA systems: AOB and anammox bacteria (examples include refs 13,14). In addition, fluorescence *in situ* hybridization and clone libraries revealed the presence of nitrite-oxidizing bacteria (NOB) in PNA systems and various studies showed that uncultured members of the phyla *Bacteroides*, *Chlorobi* and *Chloroflexi* are omnipresent in anammox bioreactors^{13,15–17}. Moreover, four recent studies reported 16S rRNA gene amplicon sequencing on two lab-scale and two full-scale PNA reactors, allowing a more detailed insight in the community composition^{18–21}. However, these insights into the microbial community were based on 16S rRNA gene

inventories only, either through PCR or fluorescence *in situ* hybridization, and most of the detected organisms were only distantly related to cultured organisms. Thus, both the functional content of their genomes and their role in PNA systems remain unknown. To gain comprehensive insight in the function of the total community in a full-scale PNA reactor, we used a shotgun metagenomics approach followed by a genome-centred metagenome analysis pipeline to retrieve near-complete genome sequences from members of the microbial community. Based on these genome sequences, we present an ecological model of the PNA wastewater treatment system.

Our model system was the Olburgen reactor, a full-scale (600 m³) PNA reactor treating wastewater from a potato-processing plant in The Netherlands²². This reactor is of particular interest, as its sludge is used to inoculate PNA reactors in other wastewater treatment plants around the world (Paques b.v.). In the Olburgen PNA reactor, the required niche differentiation is achieved using granular sludge without carrier. In addition to the granular sludge, the Olburgen reactor and other granule-based PNA systems also contain less dense, flocculent biomass (Fig. 1). In our experimental design we enriched the anaerobic community in part of the samples by washing the granules before DNA extraction, thereby removing the flocculent biomass. This enabled the identification of two distinct communities, the micro-aerobic community in the flocs and on the granular surface, and the anaerobic community in the granule core. Moreover, it enabled us to use differential coverage binning to identify contigs that were derived from the same microbial genome^{23–25}. This approach, combined with sequence composition-based methods, enabled us to extract 23 draft genomes representing the majority (59% of the reads) of the microbial community in the Olburgen PNA reactor. Together, the obtained draft genomes provide a comprehensive, system-wide overview of the community in the PNA reactor.

Results

Community overview. Sequencing, assembly and binning of the metagenomic sequencing reads (Supplementary Table 1) resulted in 23 high-quality (estimated >80% completeness) draft genomes (Table 1 and Supplementary Fig. 1). These 23 draft genomes accounted for 59% of the original sequencing data from the untreated samples (Supplementary Table 2). This indicates that our sequencing effort provided sufficient resolution to obtain a comprehensive insight in the prokaryotic microbial community of the system.

As expected, three of the most abundant bacteria in the system are relatively well-studied nitrogen cycle organisms: anammox bacterium ‘*Candidatus Brocadia sinica*’, aerobic AOB *Nitrosomonas europaea* and NOB *Nitrospira* sp. (Table 1). The remaining 20 genomes represent organisms of the phyla *Bacteroidetes* (five genomes), non-phototrophic members of the *Chlorobi* (four) and *Chloroflexi* (three), *Acidobacteria* (one), *Armatimonadetes*/OP10 (one) and candidate phyla *Omnitrophica*/OP3 (one), *Microgenomates*/OP11 (two), WS6 (two) and *Parcubacteria*/OD1 (one). For convenience, we will hereafter refer to these 23 organisms by their identifiers listed in Table 1.

Analysis of the 16S rRNA genes from the draft genomes revealed that most of these organisms have been detected previously in other (partial nitritation) anammox systems, indicating that our work has broad relevance for understanding the community and processes in PNA systems (Fig. 2c and Supplementary Table 3). However, with the exception of AMX, AOB and NOB, none of the organisms whose genomes were obtained here have close relatives that are cultured or previously

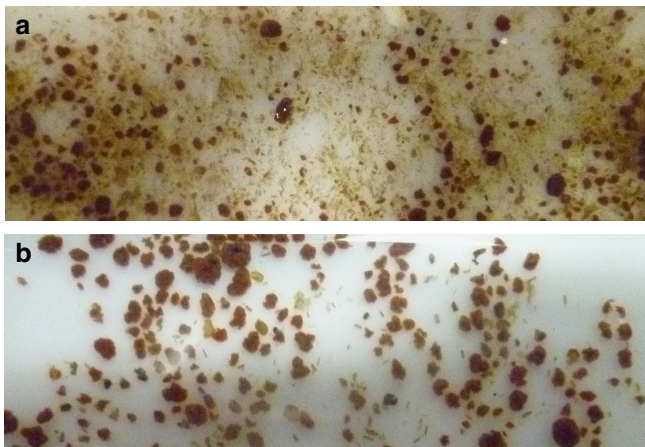


Figure 1 | Biomass from the Olburgen PNA reactor. (a) Untreated sample consisting of both flocculent and granular biomass. (b) Washed sample containing predominantly granular biomass.

Table 1 | Characteristics of the 23 draft genomes obtained in this study.

Bin ID	Phylum	Number of contigs	Draft genome size (Mbp)	Single copy marker genes (out of 111)	Features (CDS/rRNA/tRNA)	Accession number
AMX	<i>Planctomycetes</i>	86	3.9	104	3650/3/40	JZEK00000000
AOB	<i>Proteobacteria</i>	343	2.6	103	2614/3/32	JZQZ00000000
NOB	<i>Nitrospirae</i>	79	3.8	98	3743/3/31	JZQY00000000
CHB1	<i>Chlorobi</i>	24	2.5	105	2213/2/38	JYPE00000000
CHB2	<i>Chlorobi</i>	389	3.3	99	3324/2/27	LLZO00000000
CHB3	<i>Chlorobi</i>	54	2.5	101	2232/3/34	LLZP00000000
CHB4	<i>Chlorobi</i>	271	3.9	104	3593/3/39	LMYZ00000000
BCD1	<i>Bacteroidetes</i>	74	3.6	105	3000/4/30	JZQW00000000
BCD2	<i>Bacteroidetes</i>	80	3.2	104	2900/2/28	LNBW00000000
BCD3	<i>Bacteroidetes</i>	51	3.0	100	2802/2/37	LNBX00000000
BCD4	<i>Bacteroidetes</i>	72	2.1	90	2567/2/25	LNFB00000000
BCD5	<i>Bacteroidetes</i>	118	3.6	90	3876/3/34	LNFR00000000
CFX1	<i>Chloroflexi</i>	148	4.2	103	3871/3/42	JZRA00000000
CFX2	<i>Chloroflexi</i>	74	3.0	95	3502/3/41	LMZR00000000
CFX3	<i>Chloroflexi</i>	190	3.8	96	3617/2/39	LMZS00000000
OP3	<i>Omnitophica</i> (OP3)	181	4.0	100	3759/2/33	LMZT00000000
ACD	<i>Acidobacteria</i>	65	3.0	97	2915/3/40	LLEU00000000
ATM	<i>Armatimonadetes</i> (OP10)	224	2.6	91	2630/2/31	JZQX00000000
OD1	<i>Parcubacteria</i> (OD1)	31	0.9	94	973/2/34	JZEL00000000
WS6-1	WS6	7	1.4	91	1508/2/43	JYNZ00000000
WS6-2	WS6	32	1.0	84	1140/2/29	JYPD00000000
OP11-1	<i>Microgenomates</i> (OP11)	29	0.9	82	1148/2/44	LLEV00000000
OP11-2	<i>Microgenomates</i> (OP11)	47	0.9	83	1267/2/39	LMZU00000000

sequenced (16S rRNA gene identity ranging from 81 to 93% and from 81 to 90%, to cultured or sequenced organisms, respectively, see Fig. 2b and Supplementary Table 3), emphasizing the lack of knowledge on the microbial community of PNA systems.

Identification of distinct aerobic and anaerobic communities.

Our experimental design capitalized on the niche differentiation required for the coupling of aerobic and anaerobic processes in a single reactor, and provided insight into the location of the 23 organisms (Fig. 3a and Supplementary Table 2). The dense granules have a steep oxygen gradient, with a micro-aerobic outer layer surrounding a large anaerobic core⁹. As the flocculent biomass is both smaller and less dense than the granules, it is unlikely that the flocs harbour anaerobic pockets^{17,26}. Thus, the hypothesis was that the ‘washed granule’ fraction would be enriched for anaerobic organisms, whereas the untreated sample would contain a larger fraction of the micro-aerobic organisms in the system. Because of the presence of micro-aerobic niches on the granule surface, all organisms were expected to be present in both fractions. As we expected most organisms to be present in both fractions, and we expected that the granular biomass would be enriched for the community of interest for reactor performance, we decided not to extract and sequence DNA from the flocculent fraction separately.

This niche differentiation hypothesis was supported by the higher abundance of the aerobic organisms AOB and NOB in the untreated sample and the higher abundance of the anaerobic AMX in the washed granules fraction (Fig. 3a). In addition, most of the organisms capable of nitrate respiration (Fig. 3b) and the hydrogenase-encoding organisms (CHB2, CFX1 and OP3) were present in the anaerobic fraction, with the exception of CFX3, which could not be confidently assigned. For the interpretation of our results, we assume that the outer layer of the granules and the flocs face similar micro-aerobic (0–3 mg l⁻¹ dissolved O₂) conditions and, therefore, that the micro-aerobic organisms in the reactor may have a niche both in the flocs and on the granule surface.

Potential for nitrogen conversions in the PNA system. Most organic carbon and phosphorus is removed from the wastewater in two upflow anaerobic sludge bed reactors and a struvite precipitation reactor, respectively, before the influent enters the Olburgen PNA reactor²². The concentrations of ammonium, nitrate and chemical oxygen demand (COD) in the influent and effluent over a 4-month period around the time of sampling are provided and indicate the reactor performance was stable (Table 2). The influent is rich in ammonium (200–320 mg l⁻¹ NH₄⁺-N) and contains residual (recalcitrant) organic carbon (150–225 mg l⁻¹ COD) and dissolved carbon dioxide (CO₂). Moreover, the reactor is aerated continuously, leading to a micro-aerobic environment (0–3 mg l⁻¹ dissolved O₂). Part of the ammonium flowing into the reactor is converted to nitrite, which is detectable at all times (typical range 1–11 mg l⁻¹ NO₂⁻-N). When nitrite builds up to levels higher than 11 mg l⁻¹ NO₂⁻-N, part of the spent air is recirculated, reducing the oxygen concentration and, consequently, the nitrite concentration in the system²².

As the PNA reactor is an ammonium-driven ecosystem, we first evaluated the genomes for marker genes encoding key enzymes relevant to the nitrogen cycle (Fig. 3b). The marker genes used were ammonium monooxygenase (*amo*) and hydroxylamine oxidoreductase (*hao*) for ammonium oxidation^{27,28}, nitrate reductase (*nar* and *nap*) and nitrate oxidoreductase (*nxr*) for interconversion of nitrite and nitrate^{29,30}, nitrite reductase (*nirK* and *nirS*), nitric oxide reductase (*norBC* and *norZ*) and nitrous oxide reductase (*nos*) for denitrification³⁰, pentaheme nitrite reductase (*nrf*) for respiratory ammonification³¹, and hydrazine synthase (*hzs*) and hydrazine dehydrogenase (*hdh*) for anammox metabolism^{6,32}.

In the aerobic community, nitrification is the predominant nitrogen metabolism. AOB and NOB encoded the key enzymes for ammonia (*amo* and *hao*) and nitrite oxidation (*nxr*), respectively (Fig. 3b). AOB, NOB, BCD4, CFX2 and OD1 encode a copper-containing nitrite reductase (*nirK*), possibly used for detoxification, to cope with fluctuating nitrite levels in the reactor. All *Bacteroidetes* species, BCD1–5, encoded a nitrous oxide

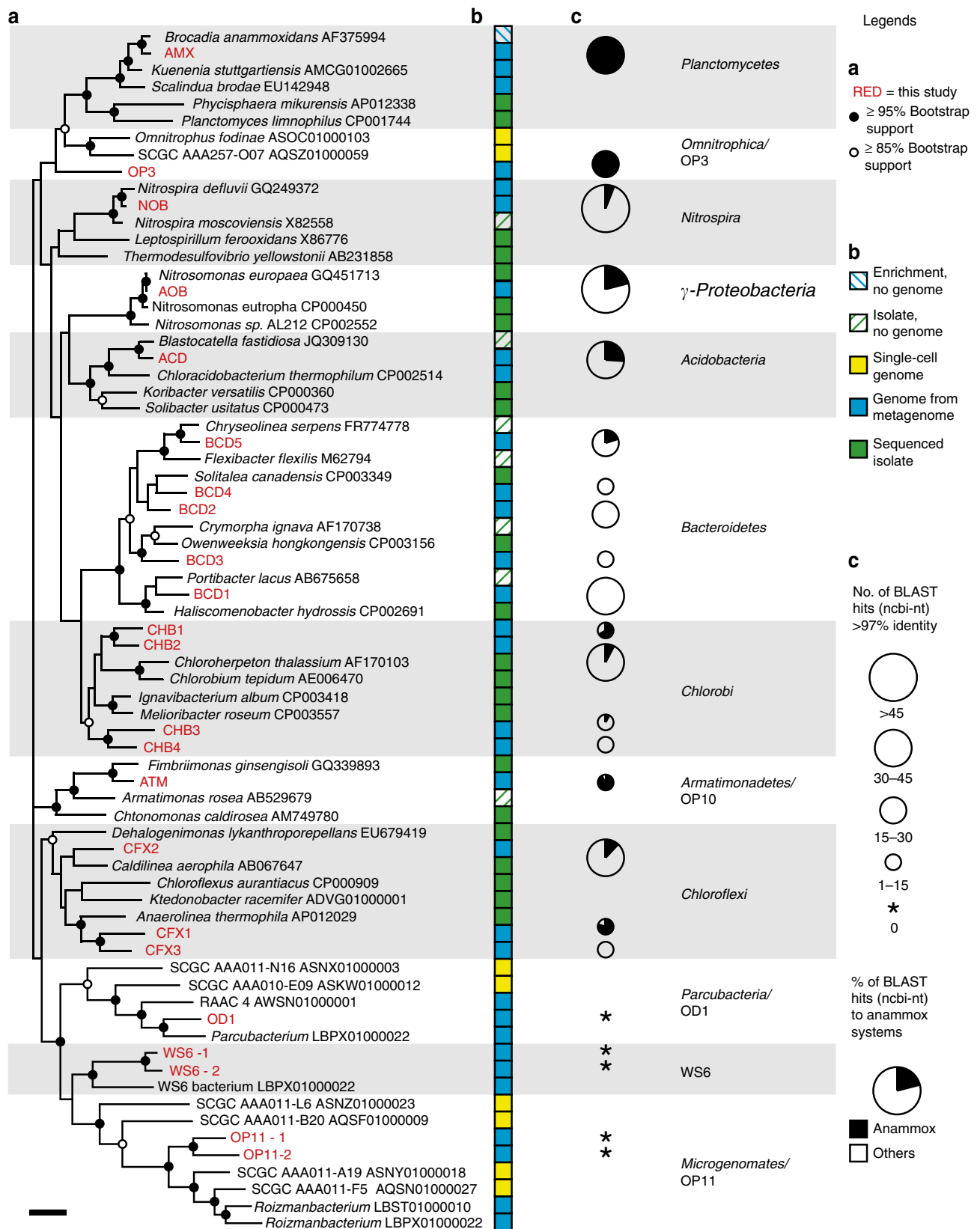


Figure 2 | Phylogeny and previous detection of the 23 dominant organisms. (a) 16S rRNA gene phylogeny of the obtained genomes and selected related organisms. Organisms in red were obtained in this study. Phylum level lineages (and the class *Gammaproteobacteria*) are indicated by background shading. Scale bar represents 10% sequence divergence. (b) Availability of isolate and/or genome for each organism in the tree. (c) Previous detection of the organisms obtained in this study using BLAST against the NCBI-nt and their occurrence in other aerobic or anaerobic anammox systems. Data are available in Supplementary Table 3.

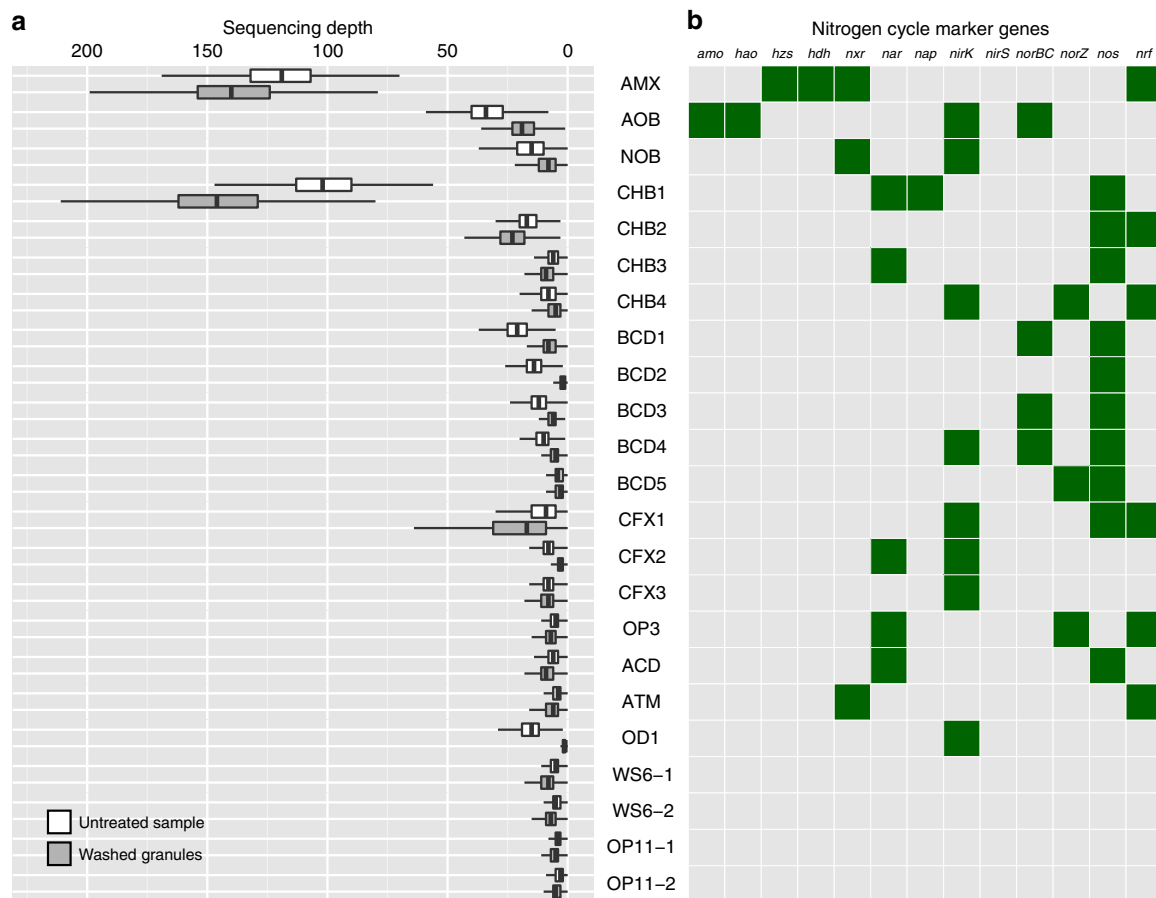


Figure 3 | Abundance of the 23 dominant organisms in washed and untreated sample and their potential for nitrogen cycling. (a) Box plots of the per-base sequencing depth of each genome, representing the average of the organic extraction data from both sampling points. Light grey indicates untreated sample and dark grey indicates washed granules. Data are available in Supplementary Table 6 (b) The presence/absence analysis of key nitrogen cycle genes from the draft genomes. Green indicates the presence and light grey indicates the absence. *amo*, ammonium monooxygenase; *hao*, hydroxylamine oxidoreductase; *hdh*, hydrazine dehydrogenase; *hzs*, hydrazine synthase; *nap*, periplasmic nitrate reductase; *nar*, cytoplasmic nitrate reductase; *nirK*, copper containing nitrite reductase; *nirS*, cytochrome *cd*, nitrite reductase; *norB*, cytochrome *c*-dependent nitric oxide reductase; *norZ*, quinol-dependent nitric oxide reductase; *nos*, nitrous oxide reductase; *nrf*, pentaheme nitrite reductase; *nrx*, nitrite oxidoreductase. Although a gene cluster with high homology to *Nitrospira/Nitrospina nrx* is present in ATM, this organism is unlikely to be a chemolithoautotrophic nitrite oxidizing bacterium.

reductase. Although nitrous oxide is not an intermediate of the anammox process⁶, it can either be produced chemically¹⁰ or by other members of the community (Fig. 3b). Whether the nitrous oxide reductases are expressed and functional and thus play a role in mitigation of nitrous oxide emissions from this PNA reactor remains to be investigated.

In the anaerobic community, the genome of AMX contained the core genes for hydrazine metabolism (*hzs* and *hdh*) and the gene cluster for nitrite oxidation, all essential for the anammox process. The potential for nitrate respiration is more widespread, as CHB1, CHB3, OP3 and ACD encode a nitrate reductase. Interestingly, none of these organisms encode a known nitrite reductase, suggesting they extrude the formed nitrite, potentially allowing cyclic feeding (Fig. 4 and discussed below). In addition to NOB and AMX, the organisms known to oxidize nitrite to nitrate^{5,33}, ATM encoded a gene cluster closely related to nitrite oxidoreductases of the *Nitrospira/Nitrospina*/anammox clade³⁴ (Supplementary Fig. 2). It remains to be investigated whether this organism is indeed capable of nitrite oxidation, as the ATM genome lacks genes required for chemolithoautotrophic growth. Interestingly, organisms closely related to ATM have almost exclusively been detected in anammox systems (Fig. 2c and Supplementary Table 3).

None of the genomes retrieved encoded a complete denitrification pathway from nitrate (or nitrite) to dinitrogen gas (Fig. 3b), suggesting that partial denitrification and exchange/transfer of nitrogen cycle intermediates could play an important role in the system. A similar fragmentation of the denitrification pathway was also observed in genome-resolved metagenome studies of aquifer and estuary sediments^{35,36}, illustrating the added value of this type of genome-resolved analysis over methods only based on gene or pathway presence.

Aerobic respiration and fermentation and carbon fixation. In addition to the conversion of nitrogen compounds, organisms can play important roles in the system via removal of organic matter, maintaining granule integrity, or by providing growth substrates for other community members. Besides nitrifiers AOB and NOB, most of the aerobic community consisted of aerobic heterotrophs (BCD1-5 and CHB4). Inside the granule core, fermentative organisms (CHB2, CFX1 and CFX3) and denitrifiers (CHB1, CHB3, ACD and OP3) potentially remove organic carbon and nitrate. A known pathway for carbon fixation could not be identified in any of these organisms. Even though PNA systems are thought to be autotrophic⁸, only AOB, NOB, AMX and CFX2 seem to be capable of fixing carbon through one of the described

pathways. In addition, the majority of the microbial community was found to be auxotroph for one or more amino acids, as determined by the absence of at least two genes from the biosynthetic pathway. Although the requirements for organic carbon could be satisfied through the influent, it seems likely to be that the amino acid requirements are met through primary production by autotrophs in the reactor.

Table 2 | Performance of the Olburgen PNA reactor around the time of sampling.

Week nr (2013)	COD (mg l ⁻¹)		NH ₄ -N (mg l ⁻¹)		NO ₃ -N (mg l ⁻¹)		Ammonium-N removed (%)
	In	Out	In	Out	In	Out	
36	220	125	310	25	0.3	60	91.9
37	225	100	280	15	0.3	50	94.6
38	200	<100	290	20	0.3	50	93.1
39	200	<100	260	20	0.4	40	92.3
40	175	170	215	15	0.3	30	93.0
41	250	100	275	20	0.3	40	92.7
42	165	<100	250	15	2	30	94.0
43	200	<100	220	15	0.3	30	93.2
44	225	<100	260	16	0.4	40	93.8
45	210	115	260	15	0.8	40	94.2
46	180	110	265	15	0.6	45	94.3
47	240	<100	260	15	0.6	45	94.2
48	150	<100	240	15	0.6	40	93.8
49	185	<100	320	20	0.5	55	93.8
50	175	<100	250	11	0.5	45	95.6
51	155	<100	280	15	0.4	35	94.6
52	200	<100	300	12	0.6	30	96.0
Average	197.4	#	266.8	16.4	0.5	41.5	93.8
s.d.	28.9	#	28.6	3.5	0.4	9.0	1.1

The week when the samples were taken is shaded in grey.

#As the detection limit is 100 mg l⁻¹ no average was calculated.

Integrating inferred niches and activities in a system model. Based on the functions derived from the assembled draft genome sequences, we propose an ecological model of the nitrogen cycle reactions catalysed by the microbial biomass in the Olburgen PNA reactor (Fig. 4). In the flocculent biomass that is dominated by nitrifiers (AOB and NOB) and *Bacteroidetes sp.* (BCD1-5), the main processes are the aerobic oxidation of organic carbon to CO₂ and complete nitrification of ammonium to nitrate (Fig. 4). The formed CO₂ either escapes to the atmosphere or is used by the primary producers (AOB, NOB and CFX2) for growth and biosynthesis (Fig. 4). A small amount of nitrate may be reduced to nitrite coupled to the oxidation of organic matter by CFX2 (not shown in Fig. 4).

Similar to the flocs, the outer layer of the granules is also exposed to oxygen. We expect that similar processes are occurring there, although there will be more competition for the formed nitrite on the granule surface (Fig. 4). Most nitrite is used by AMX to oxidize ammonium but a part is oxidized to nitrate by NOB and AMX. Formed nitrate can be reduced to nitrite, using either organic matter or hydrogen as the electron donor, by CHB1/CHB3/ACD and OP3, respectively, making additional nitrite available for AMX (Fig. 4). The hydrogen required for autotrophic nitrate reduction can be formed through fermentation of organic matter by CHB2 and CFX1. The organic matter required for both nitrate reduction and fermentation can come from the substantial amount of electron donor, measured as COD, in the influent²⁸ (Table 2), dead organic matter in the granule or extracellular granule matrix synthesized by autotrophs.

To what extent nitrite respiration contributes to the nitrite removal in the system remains to be investigated. It was clear that many organisms in the Olburgen PNA reactor have the potential to reduce nitrite to nitric oxide (Fig. 3b), but only AOB encodes both nitrite and nitric oxide reductases to further reduce the formed NO to nitrous oxide (N₂O). This indicates that detoxification, rather than respiration, of nitrite could be the main purpose of the encoded nitrite reductases, releasing the

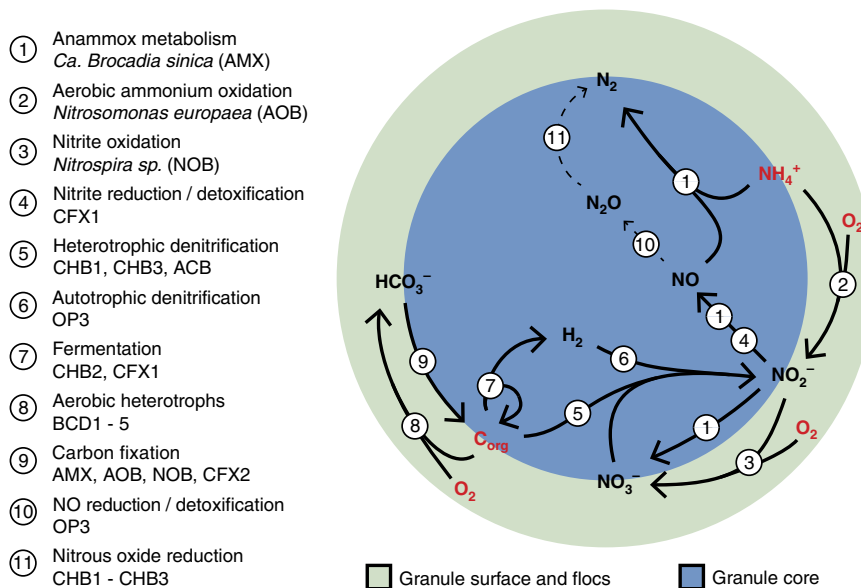


Figure 4 | Schematic overview of nitrogen conversions in the Olburgen PNA reactor. Metabolites indicated in red, ammonium (NH₄⁺), organic carbon (C-org) and molecular oxygen (O₂) are supplied in the influent and through aeration/stirring, respectively. In the flocs and on the granule surface C-org is oxidized to carbon dioxide (CO₂), which can be fixed by the autotrophs in the system. In the flocs and on the granule surface, NH₄⁺ is oxidized to nitrite (NO₂⁻). The formed NO₂⁻ is either reduced to nitric oxide (NO) or further oxidized to nitrate (NO₃⁻). NO₃⁻ that is formed in the granules can be reduced again in the anaerobic core, either with C-org or molecular hydrogen (H₂) as electron donor. H₂ can be formed through fermentation of organic carbon by CHB2, and CFX1. This cyclic feeding will result in additional C-org removal from the system. NO formed from NO₂⁻ will be combined with NH₄⁺ and converted to dinitrogen gas (N₂) by AMX.

formed NO into solution where it will be removed by the aeration. Alternatively, AMX might metabolize the NO to N₂ (ref. 37). In line with the latter, a previous study showed that NO and N₂O emission in the Olburgen PNA reactor dropped when oxygen became limiting³⁸. However, it should be noted that the effect of dissolved oxygen on nitrous oxide emissions is not well understood.

Our proposed ecological model offers an explanation for the effective performance of this PNA reactor and is consistent with the performance data of the reactor (Table 2). It also provides an ecosystem role for previously understudied organisms in this bioreactor. However, it should be noted that our model is based on genome content and differential abundance of the genomes across the biomass fractions, and awaits further validation using, for example, metatranscriptomics and metaproteomics, or enrichment/isolation of key organisms from this reactor. We note that one of the critical functions in this reactor, ammonium oxidation to nitrite, is carried out by an organism represented by only 2.9% of the reads, illustrating the need for activity-based follow-up work. Previous studies, such as the work by Ishii *et al.*³⁹ on extracellular electron transfer in a lab-scale bioreactor or by Stewart *et al.*⁴⁰ in the Eastern Tropical South Pacific oxygen minimum zone, have shown that the correlation between organism abundance and activity can be weak. The study by Ishii *et al.*³⁹ further illustrates the power of metatranscriptomics in discovery of novel gene products with important roles in community functioning. We expect that our results will provide a valuable groundwork for such future studies.

Nitrite loop enhances PNA performance. The presence of NOB in PNA systems is often viewed as detrimental to optimal N-removal, as these organisms compete with both types of ammonium oxidizers for their electron acceptors and produce nitrate, lowering the overall nitrogen removal efficiency. In the Olburgen system, NOB were nearly as abundant as AOB (2.3 and 2.9% of the reads respectively), without noticeable negative effect on reactor performance. Nitrate accumulates in the effluent (Table 2), but this is the amount expected based on the stoichiometry of the anammox process, which produces 0.3 mole of nitrate per mole of nitrite reduced⁵. Thus, excess nitrate accumulation in the system might be mitigated by the presence of (partial) denitrifiers that catalyse the conversion of nitrate back to nitrite. Such cyclic feeding was proposed previously to explain a high NOB:AOB ratio in lab-scale, aerobic sequencing batch reactors with granular sludge, where it was termed the 'nitrite loop'⁴¹. *Chlorobi* species CHB1 is probably the biggest contributor to this process. It is the second most abundant organism in the system and preferentially present in the granules (Fig. 3a). An organism with an identical 16S rRNA gene sequence was previously detected in other anammox systems by clone libraries or amplicon sequencing, where it was mislabelled as a *Planctomycete*^{16,20}. Interestingly, this organism was identified as the most dominant organism in a recent amplicon study of a different full-scale PNA reactor²⁰.

A closer look at the metabolic capabilities of CHB1 indicated that it is probably a heterotrophic denitrifier encoding an anion exporter of the *tauE* family at the same genomic locus as the nitrate reductase. Although the only characterized member of the extensive *tauE* family is a sulfite exporter, we predict a role in nitrite extrusion based on its genomic co-localization with the nitrate reductase. Consistent with this prediction, homologues of this protein are also present in CHB3 OP3 and ACD, which also encode a nitrate reductase but lack a nitrite reductase. As many other microorganisms could perform partial denitrification to recycle nitrate to nitrite, this alone is unlikely to account for the

abundance of CHB1 in this PNA reactor and similar systems. The organism could be maintaining granule integrity, as was previously suggested for another full-scale PNA reactor²⁰ or possibly supply AMX with an essential nutrient.

Candidate phyla OP11 and OD1 and WS6. In addition to the organisms described above, our work yielded high-quality draft genomes of organisms from the candidate phyla OD1, OP11 and WS6 (refs 42–44). These phyla belong to the 'candidate phyla radiation' (CPR) proposed by Brown *et al.*⁴⁵, who recently reported on nearly 800 draft genomes of CPR organisms. The CPR genomes reported here have 16S rRNA genes 78–84% identical to the closest relative described by Brown *et al.*⁴⁵, which further illustrates the large diversity of the CPR (Fig. 2a and Supplementary Table 3). As observed previously for CPR organisms^{45–47}, the genomes are very small (0.9–1.4 Mb, Table 1) and their known metabolic and biosynthetic capabilities are limited.

For the OP11 and WS6 organisms, their predicted location in the anaerobic granule core supports a fermentative lifestyle (Fig. 3a), as was suggested for related organisms^{23,24,46}. In contrast, the OD1 organism was strongly depleted in the washed granules. A possible explanation for this could be a parasitic or symbiotic relationship with BCD2, which was equally abundant in the system and similarly depleted in the granules (Fig. 3a). In recent times, an organism from the related TM7 lineage was co-cultivated as (parasitic) epibiont of a strain of *Actinomyces odontolyticus*⁴⁸, supporting this hypothesis. Further work to assess a potential relation between OD1 and BCD2 is required.

Discussion

Using a metagenomic assembly approach, we retrieved 23 draft genomes from a full-scale PNA reactor, accounting for the majority of the sequenced DNA. Our work on the PNA reactor presents the first system-wide metagenomic characterization of a full-scale engineered system. At the lab-scale, an excellent previous study reported the system-wide characterization of a terephthalate degrading culture, using a combination of single-cell genomics and shotgun metagenomics⁴⁹.

In addition to the obtained genomes, we could assign most of the organisms to niches in the reactor based on our experimental design and genome content. This allowed us to infer an ecological model of the microbial ecosystem, segregated into micro-aerobic and anaerobic compartments (Fig. 4). Our results provide a potential role for the dominant members of the microbial community, which we have integrated in a genome-based model of the Olburgen PNA reactor ecosystem.

In our model, NOBs are not detrimental to the system. Rather than causing build-up of excess nitrate, the nitrate they form can serve as electron acceptor for the degradation of organic matter and the oxidation of fermentation products, including hydrogen, in the granule core. The nitrate formed by NOB and AMX could thus partly be recycled and made available as nitrite for anammox bacteria, further improving effluent quality.

The flocculent biomass contains a large fraction of the nitrifying bacteria in the reactor. The physical separation of nitrifiers in the flocs and AMX in the granule core probably prohibits effective coupling of both processes. Therefore, a stronger selection towards granular biomass, where nitrifiers were present in the outer layer, might improve overall system performance and reduce nitrate in the effluent.

Finally, the presence of organisms closely related to the ones we have described here in other (partial nitrification) anammox systems suggests that our study is of broad relevance for the understanding of PNA systems. Our results contribute to the

understanding of PNA systems and highlight the need for further work on the microbial ecology of these innovative, sustainable and increasingly important wastewater treatment systems.

Methods

Sampling. The described system is a full-scale (600 m³) PNA reactor at the end of the main line of a wastewater treatment plant treating potato-processing wastewater at 35 °C (Olburgen, The Netherlands)²². Samples (10 l), consisting of granular and flocculent biomass, were taken from two points in the reactor to assure stirring was sufficient for homogenous biomass distribution, at 1.4 and 3.8 m from the base, on 6 November 2013 (week 45). From both the 10 l samples two representative 50 ml subsamples were taken. The granules of one subsample were washed three times with 1 × PBS, removing the flocculent biomass. Washing was done by pouring off the liquid from 50 ml reactor sample, adding 1 × PBS to a final volume of 50 ml, vigorously shaking and allowing the granular biomass to settle for 3 s before pouring off the liquid and the suspended flocculent biomass. The other subsample was processed without wash steps. All four resulting fractions were homogenized using a potter homogenizer and 1 ml of the homogenized sample was used for each DNA extraction.

DNA extraction. DNA extraction from each of the four fractions was performed using two different methods, resulting in eight different samples. Two different DNA extraction methods were used to aid in genome binning, based on the method previously published by Albertsen *et al.*²⁴. In addition, the use of multiple DNA extraction methods might mitigate the effect of extraction bias of each method. DNA extraction using the PowerSoil kit (MoBio Laboratories Inc., Carlsbad, USA) was performed according to the manufacturer's instructions. Organic solvent extraction using CTAB/Phenol/Chloroform was done using a protocol modified from Zhou *et al.*⁵⁰. Briefly, the samples were pelleted and resuspended in 675 µl CTAB extraction buffer (10 g l⁻¹ CTAB, 100 mM Tris, 100 mM EDTA, 100 mM sodium phosphate, 1.5 M NaCl pH 8), 50 µl lysozyme (10 mg ml⁻¹) and 30 µl RNase A (10 mg ml⁻¹), and incubated at 37 °C for 30 min. Subsequently, 50 µl proteinase K (10 mg ml⁻¹) was added and the samples were incubated at 37 °C for 30 min. Next, 150 µl of 10% SDS was added and the samples were incubated at 65 °C for 2 h. After cell lysis, DNA was recovered using phenol/chloroform extraction and isopropanol precipitation, and resuspended in diethylpyrocarbonate-treated nuclease-free water.

Metagenome sequencing. All kits described in this paragraph were obtained from Life Technologies (Carlsbad, CA, USA). The library preparation procedure described below was performed separately for each of the eight samples obtained after DNA extraction. DNA was sheared for 5 min using the IonXpress Plus Fragment Library Kit, following the manufacturer's instructions, and barcoded using the IonXpress barcode adapters. Further library preparation was performed using the Ion Plus Fragment Library Kit following manufacturer's instructions, with size selection using an E-gel 2% agarose gel. Emulsion PCR was done using the Ion PGM Template OT2 400 kit and sequencing was performed on an IonTorrent PGM using the Ion PGM 400 bp sequencing kit and an Ion 318v2 chip. For each run, two barcoded libraries were pooled and a total of six runs were performed. The libraries that were pooled had been treated the same way (for example, washed granules where DNA was extracted using the PowerSoil kit), but originated from the two different sampling points. The resulting 30 million reads were imported into the CLC Genomics Workbench (v6.0.4, CLCbio, Aarhus, Denmark) and read ends were trimmed using the integrated trimming algorithm. In addition, after end trimming all reads below 100 bp were discarded, resulting in 8.2 Gbp of sequencing data in single reads (Supplementary Table 1).

Metagenome assembly. All read data were cross-assembled, that is, the data sets were combined and the 27.2 million reads were assembled *de novo* together, to obtain the maximum amount of biologically relevant information on contigs with an average coverage above 15 ×. Assembly was performed using the *de novo* assembler incorporated in the CLC Genomics Server software (v6.0.4, CLCbio), using a word size of 35 and a bubble size of 5,000. This assembly resulted in 93,035 contigs larger than 1,000 bp, totalling 272 million bp. Reads (21.2 million reads; 78% of the data, 6.4 Gbp) were incorporated in the assembled contigs. A total of 8,396 of these contigs were longer than 5,000 bp.

Binning. To bin the metagenomic contigs into draft genomes, we used a multi-step approach employing a mix of available methods and custom scripts. Custom scripts are available at www.github.com/dspeth/bioinfo_scripts.

First, tetranucleotide frequencies of contigs longer than 5,000 bp were used to generate an emergent self-organizing map (ESOM), on which all contigs were subsequently projected^{51,52}. Using this map, 32,630 contigs were classified into 27 clusters (Supplementary Fig. 3). The contigs in these clusters were extracted using ESOM_bin_parser.pl. Every clustered contig was assigned a cluster number, which was used to visualize the clustered contigs during the manual binning procedure described below.

In addition, coverage, length and GC content of all contigs was obtained using the script `fasta_to_gc_cov_length_tab.pl` and tetramer content was calculated using `calc_kmerfreq.pl` (www.github.com/MadsAlbertsen/miscperlscripts)²⁴. For all contigs the average coverage in each subsample was determined using the read mapper of the CLC genomics workbench (v7.0.4, CLCbio), with mismatch penalty 2, insertion/deletion penalty 3 and an 80% identity over 50% of the read requirement.

All contig information (full sequence included) was loaded in R and ratios for both differential coverage between DNA isolations and sample treatment were calculated. All data used to extract the 23 genome bins are available on figshare (https://figshare.com/articles/Olburgen_PNA_genome_binning_source_data/1612256). Binning was done manually, based on visualization of the data using the `ggplot2` package. Clustering information from ESOM was used as visual aid in bin identification. Bins were extracted and refined based on differential coverage, differential coverage ratios, overall coverage, GC content and tetramer composition. The importance of each metric differed per bin as, for example, AMX could be binned primarily using coverage, whereas, for example, CFX1 sequence composition provided a stronger signal. An overview of the binning procedure used is available at https://github.com/dspeth/bioinfo_scripts/tree/master/binning. All commands run to extract the 23 genome bins are available as Supplementary Note 1. Examples of plots generated during the binning procedure are shown in Supplementary Figs 4–6.

After binning, reads were mapped on the binned contigs using `bowtie2` (ref. 53) and the data of each bin were reassembled with SPAdes, using the original CLC contigs as 'trusted contig'⁵⁴.

Bin completeness check and refining and annotation. The taxonomic affiliation and completeness of the obtained genome bins were assessed using CheckM⁵⁵. After the completeness estimate, several bins were manually refined by addition of previously unbinned contigs and CheckM was run on the data again. Reported completeness estimates (Table 1 and Supplementary Fig. 2) were based on the 111 essential single-copy genes proposed by Dupont *et al.*⁵⁶. Hidden Markov models (HMMs) of the proteins encoded by these genes (combined in essential.hmm) were downloaded from github.com/MadsAlbertsen/mmgenome/tree/master/scripts. CheckM was used to assess their presence and number of copies in each bin, followed by a manual check for the five bins belonging to candidate phyla OD1, WS6 and OP11.

We first validated that all genomes used the standard genetic code using the FACIL webserver⁵⁷. Next, genome annotation was done using Prokka (v1.10), with the `-c` flag removed from the prodigal command (line 649) to include partial open reading frames at contig ends^{58–62}. A set of trusted nitrogen cycle protein sequences (Supplementary Table 4) and custom whole-genome databases based on the taxonomic affiliation of the draft genomes (Supplementary Table 5) were made for annotation with Prokka.

After initial annotation the draft genomes were corrected for persistent frameshifts, common in IonTorrent data⁶³, using `iontorrent_indel_correcter.pl` (https://github.com/dspeth/bioinfo_scripts/tree/master/iontorrent_errors) followed by a round of manual curation. After error correction, the genomes were reannotated using Prokka.

Raw data and assembled, annotated draft genomes are available in GenBank under BioProject accession number PRJNA274364 and the accession numbers listed below and in Table 1.

Data interpretation. The gene complement of all 23 genomes was analysed for nitrogen cycle genes using text searches of the annotation and sequence searches with BLAST against the contig sequences, to detect misannotated genes or genes missed in annotation⁶⁴. To assess general metabolic capabilities and biosynthetic pathways, annotation searches were combined with analysis in Artemis⁶⁵, BLAST searches and mapping of the protein complement of the draft genomes on Kyoto Encyclopedia of Genes and Genomes pathway maps using KAAS⁶⁶.

Genomic data of each organism were interpreted in view of its location in the biomass, inferred from enrichment or depletion of the sequences from the granule fraction. To confidently assign each draft genome to either the flocs/granule surface niche (micro-aerobic) or the granule core niche (anaerobic), the per-base sequencing depth was calculated using SAMtools⁶⁷ on a BAM file of reads mapping on the assembled contigs and exported from the CLCgenomics Workbench (version 8.5.1). To account for the inherently lower number of reads from the untreated sample mapping to the assembled contigs (Supplementary Table 2), we subsampled the reads obtained from the washed granule fraction to 80% of their total. After this correction, an equal number of reads mapped to the combined assembled contigs (Supplementary Table 6). The distributions of sequencing depth per base were compared using the *t*-test and the Mann–Whitney *U*-test, and visualized as boxplots using R.

16S rRNA gene analysis. Phylogenetic affiliation of the bins was initially determined using CheckM⁵⁵. However, to link our draft genomes to previous (and future) 16S rRNA gene-based studies, it is essential that all bins contain a 16S rRNA gene. Eleven draft genome bins (CHB2-4, BCD2-5, CFX2-3, OP11-2 and WS6-2) lacked a 16S rRNA gene, most probably due to co-assembly of 16S reads

from different organisms into the most abundant bin within the lineage. For these bins, (near) full-length 16S rRNA gene sequences were reconstructed by mapping all reads against the Silva SSU-NR database (release 115) and extracting reads per phylum. The obtained sets were then assembled *de novo* using the CLC genomics server (v6.0.4) to obtain seed sequences, which were iteratively extended through cycles of mapping and *de novo* assembly with the 'CLCserver_16S_extension' shell script (www.github.com/dspeth/bioinfo_scripts). In case of eight bins (CHB2-4, BCD3, BCD5, CFX3, OP11-2 and WS6-2) this extension approach led to overlap with contigs from the metagenome assembly previously assigned to a bin, linking the 16S rRNA gene to the bin. The 16S rRNA genes in the remaining three bins (BCD2, BCD4 and CFX2) were subsequently linked to bins based on comparison of the 16S rRNA and the CheckM phylogenies. This bin assignment was subsequently checked using the differential coverage profile across the eight samples.

The assembled sequences were imported in ARB and aligned against the Silva reference alignment using the integrated aligner⁶⁸. The alignments were manually inspected and refined and 44 reference sequences from the Silva database were chosen based on phylogeny of the 23 16S rRNA gene sequences obtained from our data (Fig. 2a). The 67 aligned sequences were exported from ARB and a phylogenetic tree was constructed using FastTree2 (ref. 69) and bootstrapped using SeqBoot from the PHYLIP package⁷⁰ and the script CompareToBootstrap.pl (<http://meta.microbesonline.org/fasttree/treecmp.html>).

Assembled 16S sequences were used for BLAST searches against the NCBI-nr database (December 2014), to assess the previous detection of the identified organisms (at >97% 16S rRNA gene identity) and their co-occurrence with anammox bacteria (Fig. 2c and Supplementary Table 3).

References

- Daims, H., Taylor, M. W. & Wagner, M. Wastewater treatment: a model system for microbial ecology. *Trends Biotechnol.* **24**, 483–489 (2006).
- Kartal, B., Kuenen, J. G. & van Loosdrecht, M. C. M. Sewage treatment with anammox. *Science* **328**, 702–703 (2010).
- Zhu, G. *et al.* Biological removal of nitrogen from wastewater. *Rev. Environ. Contam. Toxicol.* **192**, 159–195 (2008).
- Abma, W. R. *et al.* Full-scale granular sludge anammox process. *Wat. Sci. Technol.* **55**, 27–33 (2007).
- Strous, M. *et al.* Missing lithotroph identified as new planctomycete. *Nature* **400**, 446–449 (1999).
- Kartal, B. *et al.* Molecular mechanism of anaerobic ammonium oxidation. *Nature* **479**, 127–130 (2011).
- Sliemers, A. O., Third, K. A., Abma, W. R., Kuenen, J. G. & Jetten, M. S. M. CANON and anammox in a gas-lift reactor. *FEMS Microbiol. Lett.* **218**, 339–344 (2003).
- Tsushima, I., Ogasawara, Y., Kindaichi, T., Satoh, H. & Okabe, S. Development of high-rate anaerobic ammonium-oxidizing (anammox) biofilm reactors. *Water Res.* **41**, 1623–1634 (2007).
- Nielsen, M. *et al.* Kinetics, diffusional limitation and microscale distribution of chemistry and organisms in a CANON reactor. *FEMS Microbiol. Ecol.* **51**, 247–256 (2004).
- Kampschreur, M. J., Temmink, H., Kleerebezem, R., Jetten, M. S. M. & van Loosdrecht, M. C. M. Nitrous oxide emission during wastewater treatment. *Water Res.* **43**, 4093–4103 (2009).
- Siegrist, H., Salzgeber, D., Eugster, J. & Joss, A. Anammox brings WWTP closer to energy autarky due to increased biogas production and reduced aeration energy for N-removal. *Water Sci. Technol.* **57**, 383–388 (2008).
- Lackner, S. *et al.* Full-scale partial nitrification/anammox experiences—an application survey. *Water Res.* **55**, 292–303 (2014).
- Cho, S., Takahashi, Y., Fujii, N., Yamada, Y. & Satoh, H. Nitrogen removal performance and microbial community analysis of an anaerobic up-flow granular bed anammox reactor. *Chemosphere* **78**, 1129–1135 (2010).
- Hu, B.-L. *et al.* Identification and quantification of anammox bacteria in eight nitrogen removal reactors. *Water Res.* **44**, 5014–5020 (2010).
- Li, X.-R. *et al.* The bacterial diversity in an anaerobic ammonium-oxidizing (anammox) reactor community. *Syst. Appl. Microbiol.* **32**, 278–289 (2009).
- Park, H., Rosenthal, A., Ramalingam, K., Fillos, J. & Chandran, K. Linking community profiles, gene expression and N-removal in anammox bioreactors treating municipal anaerobic digestion reject water. *Environ. Sci. Technol.* **44**, 6110–6116 (2010).
- Vlaeminck, S. E. *et al.* Aggregate size and architecture determine microbial activity balance for one-stage partial nitrification and anammox. *Appl. Environ. Microbiol.* **76**, 900–909 (2010).
- Gonzalez-Gil, G., Sougrat, R., Behzad, A. R., Lens, P. N. L. & Saikaly, P. E. Microbial community composition and ultrastructure of granules from a full-scale anammox reactor. *Microb. Ecol.* **70**, 118–131 (2015).
- Gonzalez-Martinez, A. *et al.* Bacterial community structure of a lab-scale anammox membrane bioreactor. *Biotechnol. Prog.* **31**, 186–193 (2014).
- Gonzalez-Martinez, A. *et al.* Microbial community analysis of a full-scale DEMON bioreactor. *Bioprocess. Biosyst. Eng.* **38**, 499–508 (2015).
- Chu, Z.-R. *et al.* Microbial characterization of aggregates within a one-stage nitrification–anammox system using high-throughput amplicon sequencing. *Chem. Eng. J.* **262**, 41–48 (2015).
- Abma, W. R., Driessen, W., Haarhuis, R. & van Loosdrecht, M. C. M. Upgrading of sewage treatment plant by sustainable and cost-effective separate treatment of industrial wastewater. *Wat. Sci. Technol.* **61**, 1715–1722 (2010).
- Wrighton, K. C. *et al.* Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**, 1661–1665 (2012).
- Albertsen, M. *et al.* Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* **31**, 533–538 (2013).
- Nielsen, H. B. *et al.* Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32**, 822–828 (2014).
- Winkler, M. K. H. *et al.* Microbial diversity differences within aerobic granular sludge and activated sludge flocs. *Appl. Microbiol. Biotechnol.* **97**, 7447–7458 (2013).
- Rees, M. K. Hydroxylamine metabolism of *Nitrosomonas europaea*. I. Purification of hydroxylamine oxidase. *Biochemistry* **7**, 353–366 (1968).
- Moir, J. W., Crossman, L. C., Spiro, S. & Richardson, D. J. The purification of ammonia monooxygenase from *Paracoccus denitrificans*. *FEBS Lett.* **387**, 71–74 (1996).
- Meincke, M., Bock, E., Kastrau, D. & Kroneck, P. M. Nitrite oxidoreductase from *Nitrobacter hamburgensis*: redox centers and their catalytic role. *Arch. Microbiol.* **158**, 127–131 (1992).
- Zumft, W. G. Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* **61**, 533–616 (1997).
- Cole, J. Nitrate reduction to ammonia by enteric bacteria: redundancy, or a strategy for survival during oxygen starvation? *FEMS Microbiol. Lett.* **136**, 1–11 (1996).
- Strous, M. *et al.* Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440**, 790–794 (2006).
- Watson, S. W., Bock, E., Valois, F. W., Waterbury, J. B. & Schlosser, U. *Nitrospira marina* gen. nov. sp. nov.: a chemolithotrophic nitrite-oxidizing bacterium. *Arch. Microbiol.* **144**, 1–7 (1986).
- Lücker, S., Nowka, B., Rattei, T., Spieck, E. & Daims, H. The genome of *Nitrospina gracilis* illuminates the metabolism and evolution of the major marine nitrite oxidizer. *Front. Microbiol.* doi:10.3389/fmicb.2013.00027 (2013).
- 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 (2015).
- Baker, B. J., Lazar, C. S., Teske, A. P. & Dick, G. J. Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome*. doi:10.1186/s40168-015-0077-6 (2015).
- Kartal, B., Tan, N. C. G., van de Biezen, E., Kampschreur, M. J., van Loosdrecht, M. C. M. & Jetten, M. S. M. Effect of nitric oxide on anammox bacteria. *Appl. Environ. Microbiol.* **76**, 6304–6306 (2010).
- Kampschreur, M. J. *et al.* Emission of nitrous oxide and nitric oxide from a full-scale single-stage nitrification-anammox reactor. *Wat. Sci. Technol.* **60**, 3211–3217 (2009).
- Ishii, S. *et al.* A novel metatranscriptomic approach to identify gene expression dynamics during extracellular electron transfer. *Nat. Commun.* **4**, 1601 (2013).
- Stewart, F. J., Ulloa, O. & DeLong, E. F. Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environ. Microbiol.* **14**, 23–40 (2012).
- Winkler, M. K. H., Bassin, J. P., Kleerebezem, R., Sorokin, D. Y. & van Loosdrecht, M. C. M. Unravelling the reasons for disproportion in the ratio of AOB and NOB in aerobic granular sludge. *Appl. Microbiol. Biotechnol.* **94**, 1657–1666 (2012).
- Hugenholtz, P., Pitulle, C. C., Hershberger, K. L. K. & Pace, N. R. Novel division level bacterial diversity in a Yellowstone hot spring. *J. Bacteriol.* **180**, 366–376 (1997).
- Dojka, M. A., Hugenholtz, P., Haack, S. K. & Pace, N. R. Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Appl. Environ. Microbiol.* **64**, 3869–3877 (1998).
- Harris, J. K., Kelley, S. T. & Pace, N. R. New perspective on uncultured bacterial phylogenetic division OP11. *Appl. Environ. Microbiol.* **70**, 845–849 (2004).
- Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211 (2015).
- Kantor, R. S. *et al.* Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. *MBio* **4**, e00708–e00713 (2013).
- Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).
- He, X. *et al.* Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proc. Natl Acad. Sci. USA* **112**, 244–249 (2015).
- Nobu, M. K. *et al.* Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. *ISME J.* **9**, 1710–1722 (2015).

50. Zhou, J., Bruns, M. A. & Tiedje, J. M. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**, 316–322 (1996).
51. Ultsch, A. & Mörchen, F. *ESOM-Maps: Tools for Clustering, Visualization, and Classification with Emergent SOM*. Technical Report No. 46. (Dept. of Mathematics and Computer Science, University of Marburg, Germany, 2005).
52. Dick, G. J. *et al.* Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* **10**, R85 (2009).
53. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
54. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comp. Biol.* **19**, 455–477 (2012).
55. 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).
56. Dupont, C. L. *et al.* Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J.* **6**, 1186–1199 (2012).
57. Dutilh, B. E. *et al.* FACIL: fast and accurate genetic code inference and logo. *Bioinformatics* **27**, 1929–1933 (2011).
58. Laslett, D. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **32**, 11–16 (2004).
59. Bland, C. *et al.* CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**, 209 (2007).
60. Nawrocki, E. P., Kolbe, D. L. & Eddy, S. R. Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**, 1335–1337 (2009).
61. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
62. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
63. Bragg, L. M., Stone, G., Butler, M. K., Hugenholtz, P. & Tyson, G. W. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Comput. Biol.* **9**, e1003031 (2013).
64. McGinnis, S. & Madden, T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* **32**, W20–W25 (2004).
65. Carver, T., Berriman, M., Tivey, A. & Patel, C. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* **24**, 2672–2676 (2008).
66. 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–W185 (2007).
67. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
68. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
69. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2009).
70. Felsenstein, J. PHYLIP - Phylogeny Inference Package (version 3.2). *Cladistics* **5**, 163–166 (1989).

Acknowledgements

We thank Bernadet Otten for assistance with sample collection, Waterstromen b.v. for providing the performance data of the PNA reactor and Boran Kartal for helpful discussions and critically reading the manuscript. D.R.S. was supported by BE-Basic grant fp07.002.01. B.E.D. was supported by the Netherlands Organization for Scientific Research (NWO) Vidi grant 864.14.004 and CAPES/BRASIL. M.S.M.J. was supported by ERC AG anammox 232937 and SIAM gravitation 2400202.

Author contributions

D.R.S. and M.S.M.J. designed the project. D.R.S., M.H.i.t.Z. and S.G.C. performed the sampling and sequencing. D.R.S., B.E.D. and M.S.M.J. analysed the data and wrote the manuscript.

Additional information

Accession codes: This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accessions JZEK000000000, JZQZ000000000, JZQY000000000, JYPE000000000, LLZO000000000, LLZP000000000, LMYZ000000000, JZQW000000000, LNBW000000000, LNBX000000000, LNFQ000000000, LNFRO000000000, JZRA000000000, LMZR000000000, LMZS000000000, LMZT000000000, LLEU000000000, JZQX000000000, JZEL000000000, JYNZ000000000, JYPD000000000, LLEV000000000 and LMZU000000000. The versions described in this study are version JZEK010000000, JZQZ010000000, JZQY010000000, JYPE010000000, LLZO010000000, LLZP010000000, LMYZ010000000, JZQW010000000, LNBW010000000, LNBX010000000, LNFQ010000000, LNFRO100000000, JZRA010000000, LMZR010000000, LMZS010000000, LMZT010000000, LLEU010000000, JZQX010000000, JZEL010000000, JYNZ010000000, JYPD010000000, LLEV010000000 and LMZU010000000.

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npublishing.nature.com/reprintsandpermissions/>

How to cite this article: Speth, D. R. *et al.* Genome-based microbial ecology of anammox granules in a full-scale wastewater treatment system. *Nat. Commun.* **7**:11172 doi: 10.1038/ncomms11172 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>