New insights into functional divergence and adaptive evolution of uncultured bacteria in anammox community by complete genome-centric analysis

Yi-Cheng Wang[a,1], Yanping Mao[b,1], Hui-Min Fu[a,c], Jin Wang[a], Xun Weng[a], Zi-Hao Liu[a], Xiao-Wei Xu[a], Peng Yan[a], Fang Fang[a], Jin-Song Guo[a], Yu Shen[c], You-Peng Chen[a,*]

[a] Key Laboratory of the Three Gorges Reservoir Region's Eco-Environments of MOE, Chongqing University, Chongqing 400045, China

[b] College of Chemistry and Environmental Engineering, Shenzhen University, Shenzhen 518071, Guangdong, China

[c] National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing 400067, China

[*] Corresponding author: You-Peng Chen; Key Laboratory of the Three Gorges Reservoir Region's Eco-Environments of MOE, Chongqing University, Chongqing 400045, China; E-mail address: ypchen@cqu.edu.cn.

[1] These authors contributed equally to this work.

Abstract

Anaerobic ammonium-oxidation (anammox) bacteria play a crucial role in global nitrogen cycling and wastewater nitrogen removal, but they share symbiotic relationships with various other microorganisms. No pure culture is available for anammox bacteria so far. Although shotgun metagenomics based on short reads has been widely used in anammox research, metagenome-assembled genomes (MAGs) are often discontinuous and highly contaminated, which limits in-depth analyses of anammox communities. Here, for the first time, we performed Pacific Biosciences high-fidelity (HiFi) long-read sequencing on the anammox granule sludge sample from a lab-scale bioreactor, and obtained 30 accurate and complete metagenome-assembled genomes (cMAGs). These cMAGs were obtained by selecting high-quality circular contigs from initial assemblies of long reads generated by HiFi sequencing, eliminating the need for Illumina short reads, binning, and reassembly. One new anammox species and species from three novel families were found in this anammox community. cMAG-centric analysis revealed divergences in general and nitrogen metabolism among members of the anammox community. Furthermore, we identified mobile genetic elements (MGEs) and putative horizontal gene transfer (HGT) events within these cMAGs to explore the adaptive evolution of the community. The results suggest that MGEs and HGT events, particularly transposons containing *tnpA* in anammox bacteria, might play important roles in the adaptive evolution of this anammox community. The cMAGs generated in the present study could be used to establish of a comprehensive database for anammox bacteria and associated microorganisms. Our findings highlight the advantages of HiFi sequencing for the studies of complex mixed cultures such as anammox communities and advance our understanding of anammox communities.

Keywords: Anammox community; HiFi metagenomic sequencing; Adaptive evolution; Mobile genetic elements; Horizontal gene transfer

1. Introduction

Anaerobic ammonium-oxidation (anammox) plays a key role in the global nitrogen cycle and is a cost-effective technique for treating ammonium-rich wastewaters due to no external carbon source requirement, lower oxygen consumption and sludge production (Peeters and van Niftrik, 2019; Adams et al., 2022). Approximately 120 anammox wastewater treatment plants are currently operating worldwide (Sheng et al., 2020). Anammox bacteria affiliated with the phylum Planctomycetota (Planctomycetes) are responsible for catalyzing anammox process, in which anammox bacteria utilize nitrite as electron acceptor to oxidize ammonium to dinitrogen gas (N_2) under anoxic conditions (Kartal et al., 2013). Although anammox bacteria were discovered more than 20 years ago, and substantial efforts have been made to isolate them, yet pure cultures of these bacteria are still unavailable to date (Peeters and van Niftrik, 2019; Kuenen, 2020). It is well known that anammox bacteria hardly survive alone and share symbiotic relationships with various other microorganisms affiliated with the phyla Acidobacteriota, Armatimonadota, Bacteroidota, Chloroflexota, Proteobacteria, Patescibacteria, etc. (Bhattacharjee et al., 2017; Lawson et al., 2017; Feng et al., 2018; Zhao et al., 2018a). The co-existing bacteria can provide anammox bacteria with growth factors or stable nitrite under specific C/N ratios, and prevent the inhibition of anammox bacteria by organic matter (Zhao et al., 2018b; Cao et al., 2021). Therefore, comprehensive analysis of the diversity, functional divergence, and adaptive evolution of anammox communities can contribute to advance our understanding of anammox systems and promote the widespread application of this energy-efficient treatment process.

Shotgun metagenomics employing next-generation sequencing (NGS) has been widely used in studies of anammox systems, and the genomes of different anammox bacteria and associated bacteria have been successfully retrieved from samples from

various habitats (Speth et al., 2016; Lawson et al., 2017; Suarez et al., 2022; Yang et al., 2022). However, due to the short reads and PCR requirement of NGS represented by the Illumina platform, leading to some shortcomings such as fragmented assembly and high genome contamination (Meziti et al., 2021). The recovered metagenome-assembled genomes (MAGs) generated by NGS are generally discontinuous because of repetitive, mobile, and conserved sequences, which limits their application in further analyses (Yuan et al., 2015; Sangwan et al., 2016; Meziti et al., 2019; Maguire et al., 2020). Long-read third-generation sequencing (TGS) developed by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) is a promising approach for addressing the above problems. The complete circular genomes of anammox bacteria from highly enriched cultures of *Candidatus* Kuenenia stuttgartiensis (> 87% of the total biomass) were previous recovered using only long reads generated by PacBio RS II sequencing (Frank et al., 2018; Ding and Adrian, 2020). Additionally, Liu et al. (2020) employed a combination of Illumina and ONT sequencing to develop an iterative hybrid assembly (IHA) method, which enabled the retrieval of eight circular genomes from one sample of a partial nitrification anammox reactor. However, nucleotide accuracy and assembly bubbles were not reported for these eight circular genomes. Consequently, a simpler and more efficient approach for recovering numerous accurate circular genomes from anammox systems is currently lacking. Furthermore, whether TGS can be employed alone to retrieve accurate and complete prokaryotic genomes, beyond those of dominant species, from anammox communities remains unclear.

Recently, PacBio high-fidelity (HiFi) sequencing, which can yield high-accuracy long reads (accuracy > 99.5% and average read length of 10–25 kb) (Hon et al., 2020), has become popular for the assembly of plant and animal genomes (Hon et al., 2020; Nurk et al., 2022), as well as the analysis of complex microbiomes (from human gut, sheep fecal, chicken cecum, and saline lake sediment samples) (Bickhart et al., 2022; Feng et al., 2022; Kim et al., 2022; Tao et al., 2023). HiFi sequencing has been shown to greatly improve the quality of metagenome assembly (Tao et al., 2023). Most

importantly, HiFi sequencing enables the acquisition of accurate and complete prokaryotic genomes from complex microbiomes, without the need for Illumina short reads, binning, and reassembly. Kim et al. (2022) assembled the HiFi reads derived from five human gut samples, and obtained 102 complete metagenome-assembly genomes (cMAGs). One of these cMAGs is almost identical to the genome of pure culture, with an average nucleotide identity (ANI) value exceeding 99.99%. Taken together, HiFi sequencing provides a highly suitable approach for the accurate and comprehensive analysis of environmental microbiomes, such as anammox microbiomes. The utilization of HiFi sequencing enables the retrieval of more reliable cMAGs compared with MAGs. These cMAGs can serve as a valuable resource for subsequent downstream analyses, including the identification of novel species, exploration of genes associated with specific pathways (Parks et al., 2017; Galambos et al., 2019), investigation of functional divergence, and study of the adaptive evolutionary mechanisms of prokaryotes (Douglas and Langille, 2019).

To address the aforementioned research gaps and facilitate in-depth analysis of anammox communities, HiFi sequencing were performed on the granular sludge sample obtained from a lab-scale anammox bioreactor for the first time. The long-read metagenome obtained was *de novo* assembled, and non-redundant 30 cMAGs, including a genome of a novel anammox species and three genomes from novel families, were filtered out without any binning and reassembly processes. Using these reliable cMAGs, functional divergence and adaptive evolution within the anammox community were further explored. Overall, the results of this study provide a valuable workflow for analyzing complex aquatic microbiomes as well as new insights into bacterial interactions and evolution within the anammox community.

2. Materials and methods

2.1. Sample collection and DNA extraction

The anammox granular sludge samples were collected in quintuplicate from a lab-

126 scale expanded granular sludge blanket (EGSB) reactor (**Fig. S1**), which has been
127 operated for two years. The composition of the culture medium is shown in **Table S1**
128 and **Table S2**. Residual sludge from a municipal wastewater treatment plant was used
129 as the seed sludge. The EGSB reactor was operated at room temperature (the water bath
130 portion was maintained at 35°C) and equipped with an effluent reflux system. Total
131 genomic DNA was extracted using the E.Z.N.A.® Soil DNA Kit (Omega Bio-Tek,
132 USA). The purity and quantity of genomic DNA were assessed using the NanoDrop
133 2000 (ThermoFisher Scientific, USA) and the Qubit 3.0 Fluorometer (Invitrogen, USA)
134 respectively (**Table S3**), and its integrity was verified using agarose gel electrophoresis
135 (**Fig. S2**). The DNA sample with a high molecular weight and sufficient quantity was
136 used in subsequent procedures.

137 **2.2. Metagenomic sequencing**

138 Metagenomic sequencing libraries were prepared and sequenced by Biozeron
139 Biotechnology Co. Ltd. (Shanghai, China). For PacBio HiFi sequencing, 5 μg of DNA
140 was used to construct a SMRTbell library with the SMRTbell Prep Kit (Pacific
141 Biosciences, USA). Damaged double-stranded DNA in the initial DNA sample was
142 repaired using the PreCR® Repair Mix Kit (New England Biolabs, USA) before library
143 construction. The repaired DNA was then size-selected to obtain molecules larger than
144 3 kb. The SMRTbell library was sequenced on the PacBio Sequel IIe platform using a
145 single SMRT 8M cell. HiFi reads were generated with the CCS module of SMRT Link
146 v11.0 (Pacific Biosciences, USA). For Illumina sequencing, the sequencing library was
147 constructed with a fragment length of ~450 bp and sequenced using the PE150 strategy
148 on the Illumina NovaSeq 6000 platform. Basic information of both the HiFi and
149 Illumina reads is shown in **Table S4** and **Table S5**, respectively.

150 **2.3. *De novo* assembly and quality assessment**

151 *De novo* assembly of HiFi reads was performed using three different assemblers:
152 hifiasm-meta v0.3-r063.2 (Feng et al., 2022), Flye v2.9.1-b1780 (Kolmogorov et al.,

153 2020), and Canu v2.2 (Nurk et al., 2020). Both the hifiasm-meta and Flye assembly
154 were conducted using the recommended PacBio HiFi meta mode. The Canu assembly
155 was conducted using the PacBio HiFi mode with recommended parameters. The
156 assembly quality was assessed using QUAST v5.2.0 (Gurevich et al., 2013), and the
157 results are shown in **Table S6**.

158 **2.4. Filtering, taxonomic classification, and functional annotation of cMAGs**

159 The HiFi cMAGs were filtered from initial assemblies using an adapted workflow
160 published before (Kim et al., 2022). The workflow is briefly described as follows: (1)
161 the circular contigs without assembly bubbles (assessed by BubbleGun v1.1.6
162 (Dabbaghie et al., 2022) as required) and repeats were filtered from the three assemblies;
163 (2) contigs with sequence lengths shorter than 100 kb were removed; (3) the
164 completeness and taxonomic classification of remaining contigs were assessed using
165 GTDB-Tk v2.1.1 (Parks et al., 2022), and contigs with bacterial marker genes (bac120)
166 less than 80 were removed (at this step, no archaea genome was found); (4) the SSU
167 rRNA of the remaining contigs was predicted using barrnap v0.9
168 (https://github.com/tseemann/barrnap), and contigs missing 5S, 16S or 23S rRNA were
169 removed; (5) the tRNA of remaining contigs was identified using tRNAscan-SE v2.0.11
170 (Chan et al., 2021), and contigs with tRNA types fewer than 18 were filtered out; and
171 (6) ANI values of contigs were calculated using FastANI v1.33 (Jain et al., 2018).
172 Redundant contigs were those with ANI values ≥ 99% and alignment percentages ≥
173 95%, and the contigs with the maximal sum of ANI values were selected as the
174 representative sequences. The contigs obtained through above six steps were considered
175 non-redundant cMAGs. In addition, PlasClass v0.1 (Pellow et al., 2020) was used to
176 determine whether the filtered circular contig represented a plasmid or not. The cMAG
177 abundances represented by transcripts per million (TPM) were calculated using salmon
178 v1.10.1 (Patro et al., 2017), by mapping Illumina reads onto these genomes.

179 Gene prediction of cMAGs was performed using Prodigal v2.6.3 (Hyatt et al., 2010),
180 and the predicted genes were comprehensively annotated using MEBS v1.2 (De Anda

181 et al., 2017), eggNOG database v5.0.2 (e-value < 1e-5) (Huerta-Cepas et al., 2019),
182 NCycDB (e-value < 1e-5, identity > 40%, and coverage > 40%) (Tu et al., 2019), and
183 KofamKOALA (Aramaki et al., 2020). The annotation results based on eggNOG
184 database included the functional annotation of the Clusters of Orthologous Groups
185 proteins (COGs) (Galperin et al., 2015) and CAZy (Lombard et al., 2014) databases.
186 The gene abundances represented by TPM were also calculated using salmon v1.10.1
187 (Patro et al., 2017).

188 **2.5. Identification of mobile genetic elements (MGEs) and horizontal gene transfer**
189 **(HGT)**

190 MGEs of cMAGs were identified by alignment (--escore = 1e-6, --pidentvalue = 80,
191 and –queryscore = 75) against mobileOG-db beatrix-1.6 (Brown et al., 2022). The HGT
192 events among all cMAGs were identified using MetaCHIP v1.10.12 (Song et al., 2019)
193 with default parameters. All putative transferred genes were further annotated by
194 alignment (e-value < 1e-5) against both the eggNOG database v5.0.2 (Huerta-Cepas et
195 al., 2019) and NCBI non-redundant (NR) database.

196 **2.6. Comparative genomic analyses**

197 Reference genomes (**Table S7**) of 26 anammox bacteria in eight genera (*Ca.*
198 Anammoxibacter, *Ca.* Anammoxoglobus, *Ca.* Anammoximicrobium, *Ca.* Brocadia, *Ca.*
199 Jettenia, *Ca.* Kuenenia, *Ca.* Loosdrechtia, and *Ca.* Scalindua) were downloaded from
200 both the NCBI and EMBL-EBI databases. JSpeciesWS (Richter et al., 2015) was used
201 to calculate the ANI values (based on blast) between cMAGs representing anammox
202 bacteria and other reference genomes of *Ca.* Jettenia. DNA-DNA hybridization (DDH)
203 *in silico* was performed using GGDC v3.0 (Meier-Kolthoff et al., 2022) with
204 recommended parameters. All reference anammox bacterial genomes were also
205 comprehensively annotated like cMAGs.

3. Results

3.1. Overview of cMAGs

A total of 541 circular contigs were obtained through HiFi sequencing and assembly (**Fig. 1)**. These contigs were filtered based on biological features and ANI values, yielding 30 non-redundant contigs known as cMAGs (**Fig. 1)**. As shown in **Table S8**, the predicted values by PlasClass for all cMAGs were far lower than the threshold of 0.5, thus none of them represented plasmids. The number of cMAGs from the assembly of hifiasm-meta, Canu and Flye was 19, 7, and 4, respectively. The size of these cMAGs ranged from 0.6 to 4.4 Mb, and the number of coding sequences (CDSs) ranged from 629 to 4,388 (**Table 1**). According to the taxonomic classification by GTDB-Tk, only 16 cMAGs were assigned to known species (**Fig. 2**). Three cMAGs could not be assigned to known families, and four cMAGs could not be assigned to known genera. These cMAGs represented nine phyla, with Patescibacteria having the highest number of cMAGs. Notably, more than 75% and 50% of Patescibacteria genomes were not assigned at the species and genus level, respectively. The genomes of Patescibacteria bacteria were very small, with far fewer CDSs than the genomes of bacteria affiliated with other phyla (**Table 1**). The phyla Actinobacteriota, Armatimonadota, Bdellovibrionota, and Proteobacteria each had only one member, but the abundances of cMAG.1 (Actinobacteriota) and cMAG.30 (Proteobacteria) were not low compared with other cMAGs. cMAG.27, cMAG.1, and cMAG.25 were the most abundant in the community of 30 members, and the abundances of nine species (cMAG.10, cMAG.17, cMAG.5, cMAG.26, cMAG.18, cMAG.19, cMAG.2, cMAG.23, and cMAG.13) were less than one-fiftieth of that of cMAG.27. No pure cultures are available for any of the 16 species assigned at the species level at present, indicating that all cMAGs might represent uncultured species.

Three anammox bacteria, namely cMAG.27, cMAG.28, and cMAG.29, were

identified and they were all assigned to the genus *Ca.* Jettenia. According to the classification by GTDB-Tk, cMAG.28 was unclassified at the species level, and cMAG.27 and cMAG.29 were assigned to the species *Ca.* Jettenia sp. AMX2 and *Ca.* Jettenia caeni, respectively. Therefore, we downloaded the 11 available *Ca.* Jettenia genomes from the NCBI database and calculated ANI values among the above three cMAGs and other *Ca.* Jettenia genomes. As shown in **Fig. S3**, ANI values between cMAG.29 and *Ca.* Jettenia caeni (GCA_000296795.1), and between cMAG.27 and *Ca.* Jettenia sp. AMX2 (GCA_900696655.1) were 99.70% and 99.45%, respectively. These results indicate that cMAG.29 and cMAG.27 are same species of *Ca.* Jettenia caeni and *Ca.* Jettenia sp. AMX2, respectively. The ANI value between cMAG.28 and *Ca.* Jettenia sp. AM49 (GCA_021650895.1) was 95.83%, which was just within the threshold of 95–96% species boundary (Richter and Rosselló-Móra, 2009; Palmer et al., 2020). Consequently, we calculated the DDH value between cMAG.28 and *Ca.* Jettenia sp. AM49 (GCA_021650895.1), and the resulting DDH value (66.60%) was less than the species boundary threshold of 70% (Meier-Kolthoff et al., 2013), indicating that cMAG.28 could represent a novel species of *Ca.* Jettenia different from *Ca.* Jettenia sp. AM49.

3.2. Functional divergence in general metabolism

Metabolic inference revealed that, aside from members of Patescibacteria, other bacteria in the anammox system might play a crucial role in nitrogen (N) and iron (Fe) cycles (**Fig. 3**). Additionally, members of phyla such as Planctomycetota and Chloroflexota were found to be important for the carbon (C) cycle in this anammox system. These inferences were further supported by the presence of important metabolic genes or pathways in their genome. For instance, members of the phyla Planctomycetota and Chloroflexota harbor a relatively high abundance of carbohydrate enzymes, including glycoside hydrolases and glycosyl transferases (**Fig. S4**).

In this community, there might be 13 species capable of autotrophic growth that utilize three different carbon fixation pathways. Conversely, no known carbon fixation

260 pathways could be identified in the remaining 17 species, suggesting that they might be
261 heterotrophic. In detail, all the members of the phyla Chloroflexota and
262 Planctomycetota (cMAG.27, cMAG.28, and cMAG.29) harbor the key genes encoding
263 CO dehydrogenase and acetyl-CoA synthase, which are essential for carbon fixation via
264 the Wood-Ljungdahl (WL) pathway. As for the only Proteobacteria bacterium
265 (cMAG.30), the presence of a nearly complete Arnon-Buchanan cycle indicates its
266 potential for autotrophic growth. Moreover, the presence of the complete Embden–
267 Meyerhof pathway (EMP) and the non-oxidative pentose phosphate pathway (PPP)
268 pathways in cMAG.30 reflect its capability for heterotrophic growth. Besides, the
269 presence of genes encoding cytochrome c oxidase (cbb_3-type), cytochrome bc_1
270 ubiquinol oxidase, and succinate dehydrogenase in cMAG.30 suggests its capability for
271 aerobic respiration. Three members of Patescibacteria (cMAG.18, cMAG.19, and
272 cMAG.26) were found to carry the RubisCO-encoding gene (*rbcL*), suggesting that
273 they might perform carbon fixation through the Calvin-Benson-Bassham cycle (CBB).
274 The presence of complete EMP and the non-oxidative PPP pathways in three *Ca.*
275 Jettenia species (cMAG.27, cMAG.28, and cMAG.29), as well as a complete
276 tricarboxylic acid cycle (TCA cycle) and *pdhA* encoding pyruvate dehydrogenase,
277 suggest their capabilities for fermentation.

278 **3.3. Functional divergence in nitrogen metabolism**

279 To further explore functional divergence in nitrogen metabolism of these bacteria, 30
280 cMAGs were comprehensively annotated using a curated nitrogen cycle database
281 (NCycDB). A total of 24 species were identified as potentially participating in key
282 nitrogen conversion processes, such as anammox, denitrification, nitrification, and
283 dissimilatory nitrate reduction to ammonia (DNRA) (**Fig. 4A** and **B**). Three members
284 of *Ca.* Jettenia harbor the full set of anammox genes, including *nirK*/*nirS*, *hzs*, and *hdh*,
285 but they all lack *nxrA* encoding nitrite oxidoreductase (**Fig. 4A**). In addition, they all
286 harbor *napA, narG*, and *nrfA*, suggesting their capacities for DNRA. Similarly, among
287 the 11 *Ca.* Jettenia species from other habitats, none of them carry the *nxrA* gene (**Fig.**

288 **S5**). Only three members, i.e. two members of *Ca.* Jettenia caeni and one member of
289 *Ca.* Jettenia sp. AMX2, harbor both *napA* and *narG*. Nearly all of these *Ca.* Jettenia
290 members (10 out of 11) may be capable of DNRA. Members belonging to the genera
291 *Ca.* Jettenia and *Ca.* Brocadia possess a richer repertoire of nitrogen metabolism marker
292 genes compared with anammox bacteria in other genera (**Fig. S5**).

293 A total of 20 non-anammox bacteria community members were found to carry
294 denitrification genes (**Fig. 4A**). However, none of them carry the full set of
295 denitrification genes, indicating a potential for cooperative denitrification. In addition,
296 seven members carry the key gene *nrfA* for DNRA, while only five of them harbor the
297 complete DNRA pathway. The Armatimonadota bacterium (cMAG.2) exclusively
298 harbors *nrfA*. According to gene abundance, the denitrification pathway might have not
299 outcompeted both the anammox and DNRA pathways (**Fig. 4B)**. Of note, the majority
300 of Patescibacteria members were not been identified with important known nitrogen
301 metabolism genes, suggesting that they might play other roles in the system. No
302 complete nitrification pathway was identified in any of these cMAGs, and the gene
303 encoding ammonia monooxygenase (*amo*) was not found.

304 **3.4. Metagenomic evidence for adaptive evolution**

305 To gain insights into the adaptive evolution of the anammox community, an analysis
306 was conducted to identify MGEs in all cMAGs using the latest MGE database
307 (mobileOG-db) with stringent parameters. Among these cMAGs, eight members were
308 found to carry MGEs, with a total of 63 MGEs identified (**Fig. 5A**). Notably, cMAG.4,
309 cMAG.7, cMAG.28, and cMAG.30 were found to harbor sequences related to phages.
310 The majority of the predicted MGEs were associated with integration or excision
311 processes. Of note, one of the anammox bacteria (cMAG.29) harbored a discrete
312 distribution of 28 transposons, all of which contained the transposase-encoding gene
313 *tnpA* (**Fig. S6**). Another anammox bacterium (cMAG.28) also harbors three transposons
314 that containing *tnpA*, while the remaining members (including cMAG.27) did not carry
315 *tnpA*. The same identification procedure was performed on the other 26 reference

genomes of anammox bacteria. Approximately half of the anammox bacteria carried *tnpA*, with gene counts no less than 18 in *Ca.* Jettenia caeni (only one) and all members of *Ca.* Kuenenia stuttgartiensis, and others carried no more than eight *tnpA* genes (**Fig. 5B**). Furthermore, the upstream and downstream genes of *tnpA* in cMAG.29 were found to be different (**Fig. 6**), indicating that they represent the simple transposon (insertion sequence) without additional functional genes.

In addition to MGE analysis, HGT events were identified on all cMAGs. A total of 50 putative HGT events along with their transfer direction were identified, involving 6 phyla and 14 species (**Fig. 7A**). Most of the putative HGT events occurred within the phylum Chloroflexota, accounting for 66% of the total. Putative HGT events within the phylum Bacteroidetes accounted for 34% of the total, with 24% assumed to occur with members of different phyla. At the species level, putative HGT events were the most abundant in one member of the genus *UBA12294* (cMAG.13), with 25 donors and 5 recipients. Interestingly, the identification by MetaCHIP showed that the denitrifying bacterium cMAG.13 donated 10 genes to another denitrifying bacterium, cMAG.16, and reciprocally, cMAG.16 donated 5 genes to cMAG.13. A putative HGT event was observed between an anammox bacterium (cMAG.27) and a denitrifying bacterium affiliated with Bacteroidota, showing that cMAG.27 might acquire a gene from cMAG.8.

The donor and recipient genes for putative HGT events were identified. A total of 26 genes (> 50% of the total) in both donated and received genes could be annotated using the COG database (**Fig. 7B** and **C**). The donated genes are primarily responsible for energy production and conversion (C, 4 genes), posttranslational modification, protein turnover, chaperones (O, 4 genes), while four genes were annotated with unknown functions. For example, the Bacteroidota bacterium (genus *UBA2336*) cMAG.9 might acquire four genes (*sucD*, *fumC*, *atpA*, and *atpD*) from another Bacteroidota bacterium cMAG.4. It should be noted that the anammox bacterium cMAG.27 might acquire a gene encoding IS200 family transposase. No transfer genes directly related to nitrogen

removal were identified. However, nearly half of the genes in both donors and recipients were either unannotated or had unknown functions; thus, additional investigation is required.

4. Discussion

In this study, a total of 30 complete circular prokaryotic genomes were successfully recovered from an anammox community using HiFi sequencing. We made slight adjustments to the filtering criteria compared with a previous study (Kim et al., 2022) to include circular contigs that did not meet all the proposed criteria but exhibited other desirable features. Since the accuracy of HiFi sequencing was extremely high (average quality = Q29.5) and the circular contig with assembly errors had been removed, the substandard circular contigs represented either chromosomes or circular plasmids. According to the prediction by PlasClass, these substandard contigs were not circular plasmid sequences. More importantly, most of these contigs were assigned to Patescibacteria, which are also referred to as candidate phyla radiation bacteria and are known to have much smaller genomes compared with other prokaryotes (Luef et al., 2015; Castelle and Banfield, 2018; Tian et al., 2020). They might have fewer marker genes than other conventional bacteria. These substandard contigs were therefore retained for downstream analysis. In strict terms, some of the recovered cMAGs from this study or previous studies (Liu et al., 2020; Kim et al., 2022) might represent complete chromosomes rather than complete genomes, as the existence of associated MGEs (e.g., plasmids) remains uncertain. For complex mixed cultures, such as anammox communities, accurately resolving the link of MGEs to their host complete chromosomes poses a major challenge. To address this problem, we propose that a combination of the recently developed high-throughput single-microbe sequencing (Microbe-seq) (Zheng et al., 2022) with HiFi sequencing would offer an effective strategy.

370 Although hifiasm-meta did not generate the highest number of circular contigs, it
371 yielded cMAGs that accounted for 63% of the total. This is consistent with previous
372 findings and highlights the advantages of hifiasm-meta for the assembly of HiFi reads
373 (Kim et al., 2022). These cMAGs represent anammox bacteria and associated bacteria,
374 including those with low abundance, and ~47% of them may represent novel species,
375 including a novel species of anammox bacteria. This suggests that a considerable
376 number of anammox-associated bacteria remain poorly characterized. The reliability of
377 the recovered cMAGs was supported by the high ANI values with conspecific genomes
378 from previous studies. Our approach outperformed the IHA method (Liu et al., 2020)
379 in terms of using fewer long reads (26.7 Gb vs. 69.4 Gb) and no need for short reads
380 while ensuring the accuracy of circular genome reconstruction, demonstrating the
381 potential for HiFi sequencing to accurately recover prokaryotic genomes from complex
382 environmental microbiomes.

383 cMAGs provide comprehensive and accurate information on the metabolic potential
384 and genetic interactions within the anammox community. Unlike previous MAGs
385 (Lawson et al., 2017; Liu et al., 2020; Kallistova et al., 2022) that were affected by
386 discontinuity and contamination, these cMAGs offer a more complete picture of the
387 community. The anammox system was thought to be autotrophic due to the absence of
388 external organic carbon source, but there were still many heterotrophs in this system.
389 Actually, it is very common that anammox bacteria share symbiotic relationships with
390 heterotrophic bacteria (Kindaichi et al., 2012; Lawson et al., 2017; Zhao et al., 2018b).
391 In light of our culture conditions (without external organic carbon source), these co-
392 existing heterotrophic bacteria might be responsible for granular aggregation, supplying
393 growth factors to anammox bacteria, and consuming organic matter.

394 Approximately 70% of the species in the community were potentially engaged in
395 nitrogen removal processes, either through anammox or denitrification pathways. The
396 remaining 30% might not be directly involved in such nitrogen conversion processes.
397 Nevertheless, some of them could perform carbon fixation, which provides organic

matter to other co-existing bacteria. These members of Patescibacteria might also play roles in the degradation of poly-N-acetylglucosamine produced by anammox bacteria (Hosokawa et al., 2021). Three anammox bacteria affiliated with *Ca.* Jettenia could conduct DNRA and carbon fixation through the WL pathway, which is consistent with the results of previous studies (Strous et al., 2006; Lawson et al., 2017; Mardanov et al., 2019). Intriguingly, our results indicate that these three anammox bacteria were capable of respiratory ammonification with oxidation acetate, and similar results were also observed for the two other anammox bacteria affiliated with *Ca.* Jettenia (Ali et al., 2015; Mardanov et al., 2019). The DNRA capability of anammox bacteria could be a valuable strategy for survival, as they would no longer dependent on other microorganisms for the supply of ammonium and nitrite (Kartal et al., 2007). Furthermore, not all anammox bacteria are capable of carrying out the DNRA pathway, and it could be speculated that these anammox bacteria with the DNRA pathway might be better adapted to environments with a low concentration of ammonium and a high concentration of nitrite or nitrate.

Adaptive evolution plays a crucial role in the adaptation of bacteria to different habitats, and both MGEs and HGT contribute to this process (Brito, 2021; Arnold et al., 2022; Shi et al., 2022). We identified abundant MGEs and putative HGT events within the anammox community. Certain members of the community showed evidence of past phage infections, which led to mutations in their genomes. It is well known that transposons can provide a pool of pre-evolved genes that help bacteria adapt quickly to their environment (Dahlberg and Hermansson, 1995). The transposons containing the *tnpA* gene were prevalent in some anammox bacteria, indicating they might play important roles in their environmental adaptation. As mentioned above, anammox bacteria are not available for pure culture to date; thus, molecular biology methods based on monoclones cannot be used for the identification of metabolic pathways or bioaugmentation. Many questions about the basic metabolic pathways of anammox bacteria have not yet been addressed (Peeters and van Niftrik, 2019; Kuenen, 2020),

challenges (such as the fast start-up and operation under non-optimal conditions) during the full-scale operation of anammox also require attention (Adams et al., 2020; Kouba et al., 2022), and this will require the development of methods for the genome editing of anammox bacteria. We believe that the transposon containing *tnpA* might be one of the breakthroughs for the genome editing of anammox bacteria in mixed cultures.

A considerable number of putative HGT events were predicted, suggesting that HGT might be another important strategy for promoting the adaptive evolution of bacteria within the anammox community. The transferred genes are involved in multiple metabolic pathways rather than antibiotic resistance, showing that these putative HGT events might contribute to the metabolic specialization and competitive advantage of these species (Seeleuthner et al., 2018; Wang et al., 2022). Particularly, the putative HGT events mainly occurred in denitrifying bacteria in the phyla Chloroflexota and Bacteroidota. Similar findings were also observed in the previous study on a full-scale anammox system (Wang et al., 2022). It could be inferred that denitrifying bacteria might be very active in genetic interactions within anammox communities. The putative HGT event between cMAG.27 and cMAG.8 shows that HGT events could occur between anammox bacteria and co-existing bacteria affiliated with distantly related lineages. Moreover, the transferred gene is a transposase-coding gene, again suggesting that some specific transposons could be utilized for the genome editing of anammox bacteria.

5. Conclusions

Our study demonstrates that HiFi sequencing enables the reconstruction of accurate cMAGs from the complex anammox community without the need for Illumina short reads, binning, and reassembly. These cMAGs represent not only high-abundance species, but also low-abundance species in this anammox community. Additionally, these cMAGs provide valuable insights into microbial diversity, metabolic divergence,

and genetic interactions within the community, and they may lay a core foundation for a comprehensive database for anammox bacteria and associated microorganisms. The presence of MGEs and putative HGT events suggests ongoing evolution within the community, and they might play an important role in the domestication of the anammox culture. To our knowledge, this is the first study to describe the functional divergence and adaptive evolution of uncultured bacteria in an anammox community at the complete genome level. Further studies are warranted to confirm the functional divergence and genetic interactions within anammox communities, employing a combination of HiFi sequencing and other techniques.

Data availability

The 30 cMAGs generated in this study have been deposited in the GenBank under the accession number PRJNA983097. The raw sequencing data have been deposited in the Sequence Read Archive under the accession number PRJNA981667 and PRJNA981689.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (21876016, 52070031, and 42207022) and the Natural Science Foundation of Guangdong Province (2023A1515012019).

References

Adams M., J. Xie, A. w. J. Kabore, Y. Chang, J. Xie, M. Guo, C. Chen, 2022. Research advances in anammox granular sludge: A review. Crit. Rev. Environ. Sci. Technol. 52, 631-674.

Adams M., J. X. Xie, J. W. Xie, Y. F. Chang, M. L. Guo, C. J. Chen, T. C. Zhang, 2020. The effect of carrier addition on Anammox start-up and microbial community:

476 a review. Rev. Environ. Sci. Biotechnol. 19, 355-368.
477 Ali M., M. Oshiki, T. Awata, K. Isobe, Z. Kimura, H. Yoshikawa, D. Hira, T. Kindaichi,
478 H. Satoh, T. Fujii, S. Okabe, 2015. Physiological characterization of anaerobic
479 ammonium oxidizing bacterium "*Candidatus* Jettenia caeni". Environ.
480 Microbiol. 17, 2172-2189.
481 Aramaki T., R. Blanc-Mathieu, H. Endo, K. Ohkubo, M. Kanehisa, S. Goto, H. Ogata,
482 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and
483 adaptive score threshold. Bioinformatics 36, 2251-2252.
484 Arnold B. J., I. T. Huang, W. P. Hanage, 2022. Horizontal gene transfer and adaptive
485 evolution in bacteria. Nat. Rev. Microbiol. 20, 206-218.
486 Bhattacharjee A. S., S. Wu, C. E. Lawson, M. S. M. Jetten, V. Kapoor, J. W. S. Domingo,
487 K. D. McMahon, D. R. Noguera, R. Goel, 2017. Whole-community
488 metagenomics in two different anammox configurations: Process performance
489 and community structure. Environ. Sci. Technol. 51, 4317-4327.
490 Bickhart D. M., M. Kolmogorov, E. Tseng, D. M. Portik, A. Korobeynikov, I.
491 Tolstoganov, G. Uritskiy, I. Liachko, S. T. Sullivan, S. B. Shin, A. Zorea, V. P.
492 Andreu, K. Panke-Buisse, M. H. Medema, I. Mizrahi, P. A. Pevzner, T. P. L.
493 Smith, 2022. Generating lineage-resolved, complete metagenome-assembled
494 genomes from complex microbial communities. Nat. Biotechnol. 40, 711-719.
495 Brito I. L., 2021. Examining horizontal gene transfer in microbial communities. Nat.
496 Rev. Microbiol. 19, 442-453.
497 Brown C. L., J. Mullet, F. Hindi, J. E. Stoll, S. Gupta, M. Choi, I. Keenum, P. Vikesland,
498 A. Pruden, L. Zhang, 2022. mobileOG-db: a manually curated database of
499 protein families mediating the life cycle of bacterial mobile genetic elements.
500 Appl. Environ. Microbiol. 88, e0099122.
501 Cao S. B., R. Du, Y. Zhou, 2021. Coupling anammox with heterotrophic denitrification
502 for enhanced nitrogen removal: A review. Crit. Rev. Environ. Sci. Technol. 51,
503 2260-2293.
504 Castelle C. J., J. F. Banfield, 2018. Major new microbial groups expand diversity and
505 alter our understanding of the tree of life. Cell 172, 1181-1197.
506 Chan P. P., B. Y. Lin, A. J. Mak, T. M. Lowe, 2021. tRNAscan-SE 2.0: improved
507 detection and functional classification of transfer RNA genes. Nucleic Acids
508 Res. 49, 9077-9096.
509 Dabbaghie F., J. Ebler, T. Marschall, 2022. BubbleGun: enumerating bubbles and
510 superbubbles in genome graphs. Bioinformatics 38, 4217-4219.
511 Dahlberg C., M. Hermansson, 1995. Abundance of Tn3, Tn21, and Tn501 transposase
512 (*tnpA*) sequences in bacterial community DNA from marine environments. Appl.
513 Environ. Microbiol. 61, 3051-3056.
514 De Anda V., I. Zapata-Peñasco, A. C. Poot-Hernandez, L. E. Eguiarte, B. Contreras-
515 Moreira, V. Souza, 2017. MEBS, a software platform to evaluate large
516 (meta)genomic collections according to their metabolic machinery: unraveling
517 the sulfur cycle. GigaScience 6, 1-17.

Ding C., L. Adrian, 2020. Comparative genomics in "*Candidatus* Kuenenia stuttgartiensis" reveal high genomic plasticity in the overall genome structure, CRISPR loci and surface proteins. BMC Genomics 21, 851.

Douglas G. M., M. G. I. Langille, 2019. Current and promising approaches to identify horizontal gene transfer events in metagenomes. Genome Biol. Evol. 11, 2750-2766.

Feng X., H. Cheng, D. Portik, H. Li, 2022. Metagenome assembly of high-fidelity long reads with hifiasm-meta. Nat. Methods 19, 671-674.

Feng Y., Y. P. Zhao, Y. Z. Guo, S. T. Liu, 2018. Microbial transcript and metabolome analysis uncover discrepant metabolic pathways in autotrophic and mixotrophic anammox consortia. Water Res. 128, 402-411.

Frank J., S. Lücker, R. H. A. M. Vossen, M. S. M. Jetten, R. J. Hall, H. J. M. Op den Camp, S. Y. Anvar, 2018. Resolving the complete genome of *Kuenenia stuttgartiensis* from a membrane bioreactor enrichment using Single-Molecule Real-Time sequencing. Sci. Rep. 8, 4580.

Galambos D., R. E. Anderson, J. Reveillaud, J. A. Huber, 2019. Genome-resolved metagenomics and metatranscriptomics reveal niche differentiation in functionally redundant microbial communities at deep-sea hydrothermal vents. Environ. Microbiol. 21, 4395-4410.

Galperin M. Y., K. S. Makarova, Y. I. Wolf, E. V. Koonin, 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 43, D261-269.

Gurevich A., V. Saveliev, N. Vyahhi, G. Tesler, 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072-1075.

Hon T., K. Mars, G. Young, Y. C. Tsai, J. W. Karalius, J. M. Landolin, N. Maurer, D. Kudrna, M. A. Hardigan, C. C. Steiner, S. J. Knapp, D. Ware, B. Shapiro, P. Peluso, D. R. Rank, 2020. Highly accurate long-read HiFi sequencing data for five complex genomes. Sci. Data 7, 399.

Hosokawa S., K. Kuroda, T. Narihiro, Y. Aoi, N. Ozaki, A. Ohashi, T. Kindaichi, 2021. Cometabolism of the superphylum Patescibacteria with anammox bacteria in a long-term freshwater anammox column reactor. Water 13, 208.

Huerta-Cepas J., D. Szklarczyk, D. Heller, A. Hernández-Plaza, S. K. Forslund, H. Cook, D. R. Mende, I. Letunic, T. Rattei, L. J. Jensen, 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47, D309-D314.

Hyatt D., G. L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, L. J. Hauser, 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119.

Jain C., L. M. Rodriguez-R, A. M. Phillippy, K. T. Konstantinidis, S. Aluru, 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. 9, 5114.

Kallistova A., Y. Nikolaev, V. Grachev, A. Beletsky, E. Gruzdev, V. Kadnikov, A.

Dorofeev, J. Berestovskaya, A. Pelevina, I. Zekker, N. Ravin, N. Pimenov, A. Mardanov, 2022. New insight into the interspecies shift of anammox bacteria *Ca.* "Brocadia" and *Ca.* "Jettenia" in reactors fed with formate and folate. Front. Microbiol. 12, 802201.

Kartal B., N. M. de Almeida, W. J. Maalcke, H. J. M. Op den Camp, M. S. M. Jetten, J. T. Keltjens, 2013. How to make a living from anaerobic ammonium oxidation. FEMS Microbiol. Rev. 37, 428-461.

Kartal B., M. M. M. Kuypers, G. Lavik, J. Schalk, H. den Camp, M. S. M. Jetten, M. Strous, 2007. Anammox bacteria disguised as denitrifiers: nitrate reduction to dinitrogen gas via nitrite and ammonium. Environ. Microbiol. 9, 635-642.

Kim C. Y., J. Ma, I. Lee, 2022. HiFi metagenomic sequencing enables assembly of accurate and complete genomes from human gut microbiota. Nat. Commun. 13, 6367.

Kindaichi T., S. Yuri, N. Ozaki, A. Ohashi, 2012. Ecophysiological role and function of uncultured Chloroflexi in an anammox reactor. Water Sci. Technol. 66, 2556-2561.

Kolmogorov M., D. M. Bickhart, B. Behsaz, A. Gurevich, M. Rayko, S. B. Shin, K. Kuhn, J. Yuan, E. Polevikov, T. P. L. Smith, P. A. Pevzner, 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat. Methods 17, 1103-1110.

Kouba V., C. Bachmannova, T. Podzimek, P. Lipovova, M. C. M. van Loosdrecht, 2022. Physiology of anammox adaptation to low temperatures and promising biomarkers: A review. Bioresour. Technol. 349, 126847.

Kuenen J. G., 2020. Anammox and beyond. Environ. Microbiol. 22, 525-536.

Lawson C. E., S. Wu, A. S. Bhattacharjee, J. J. Hamilton, K. D. McMahon, R. Goel, D. R. Noguera, 2017. Metabolic network analysis reveals microbial community interactions in anammox granules. Nat. Commun. 8, 15416.

Liu L., Y. Wang, Y. Che, Y. Chen, Y. Xia, R. Luo, S. H. Cheng, C. Zheng, T. Zhang, 2020. High-quality bacterial genomes of a partial-nitritation/anammox system by an iterative hybrid assembly method. Microbiome 8, 155.

Lombard V., H. G. Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, D490-D495.

Luef B., K. R. Frischkorn, K. C. Wrighton, H.-Y. N. Holman, G. Birarda, B. C. Thomas, A. Singh, K. H. Williams, C. E. Siegerist, S. G. Tringe, K. H. Downing, L. R. Comolli, J. F. Banfield, 2015. Diverse uncultivated ultra-small bacterial cells in groundwater. Nat. Commun. 6, 6372.

Maguire F., B. F. Jia, K. L. Gray, W. Y. V. Lau, R. G. Beiko, F. S. L. Brinkman, 2020. Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic Islands. Microb. Genomics 6, 000436.

Mardanov A. V., A. V. Beletsky, N. V. Ravin, E. A. Botchkova, Y. V. Litti, A. N.

Nozhevnikova, 2019. Genome of a novel bacterium "*Candidatus* Jettenia ecosi" reconstructed from the metagenome of an anammox bioreactor. Front. Microbiol. 10, 2442.

Meier-Kolthoff J. P., A. F. Auch, H. P. Klenk, M. Goker, 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14, 60.

Meier-Kolthoff J. P., J. S. Carbasse, R. L. Peinado-Olarte, M. Göker, 2022. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res. 50, D801-D807.

Meziti A., L. M. Rodriguez-R, J. K. Hatt, A. Pena-Gonzalez, K. Levy, K. T. Konstantinidis, 2021. The reliability of metagenome-assembled genomes (MAGs) in representing natural populations: Insights from comparing MAGs against isolate genomes derived from the same fecal sample. Appl. Environ. Microbiol. 87, e02593-02520.

Meziti A., D. Tsementzi, L. M. Rodriguez-R, J. K. Hatt, H. Karayanni, K. A. Kormas, K. T. Konstantinidis, 2019. Quantifying the changes in genetic diversity within sequence-discrete bacterial populations across a spatial and temporal riverine gradient. ISME J. 13, 767-779.

Nurk S., S. Koren, A. Rhie, M. Rautiainen, A. V. Bzikadze, A. Mikheenko, M. R. Vollger, N. Altemose, L. Uralsky, A. Gershman, S. Aganezov, S. J. Hoyt, M. Diekhans, G. A. Logsdon, M. Alonge, S. E. Antonarakis, M. Borchers, G. G. Bouffard, S. Y. Brooks, G. V. Caldas, N.-C. Chen, H. Cheng, C.-S. Chin, W. Chow, L. G. de Lima, P. C. Dishuck, R. Durbin, T. Dvorkina, I. T. Fiddes, G. Formenti, R. S. Fulton, A. Fungtammasan, E. Garrison, P. G. S. Grady, T. A. Graves-Lindsay, I. M. Hall, N. F. Hansen, G. A. Hartley, M. Haukness, K. Howe, M. W. Hunkapiller, C. Jain, M. Jain, E. D. Jarvis, P. Kerpedjiev, M. Kirsche, M. Kolmogorov, J. Korlach, M. Kremitzki, H. Li, V. V. Maduro, T. Marschall, A. M. McCartney, J. McDaniel, D. E. Miller, J. C. Mullikin, E. W. Myers, N. D. Olson, B. Paten, P. Peluso, P. A. Pevzner, D. Porubsky, T. Potapova, E. I. Rogaev, J. A. Rosenfeld, S. L. Salzberg, V. A. Schneider, F. J. Sedlazeck, K. Shafin, C. J. Shew, A. Shumate, Y. Sims, A. F. A. Smit, D. C. Soto, I. Sovic, J. M. Storer, A. Streets, B. A. Sullivan, F. Thibaud-Nissen, J. Torrance, J. Wagner, B. P. Walenz, A. Wenger, J. M. D. Wood, C. Xiao, S. M. Yan, A. C. Young, S. Zarate, U. Surti, R. C. McCoy, M. Y. Dennis, I. A. Alexandrov, J. L. Gerton, R. J. O'Neill, W. Timp, J. M. Zook, M. C. Schatz, E. E. Eichler, K. H. Miga, A. M. Phillippy, 2022. The complete sequence of a human genome. Science 376, 44-53.

Nurk S., B. P. Walenz, A. Rhie, M. R. Vollger, G. A. Logsdon, R. Grothe, K. H. Miga, E. E. Eichler, A. M. Phillippy, S. Koren, 2020. HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. Genome Res. 30, 1291-1305.

Palmer M., E. T. Steenkamp, J. Blom, B. P. Hedlund, S. N. Venter, 2020. All ANIs are not created equal: implications for prokaryotic species boundaries and

integration of ANIs into polyphasic taxonomy. Int. J. Syst. Evol. Microbiol. 70, 2937-2948.

Parks D. H., M. Chuvochina, C. Rinke, A. J. Mussig, P. A. Chaumeil, P. Hugenholtz, 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. 50, D785-D794.

Parks D. H., C. Rinke, M. Chuvochina, P. A. Chaumeil, B. J. Woodcroft, P. N. Evans, P. Hugenholtz, G. W. Tyson, 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat. Microbiol. 2, 1533-1542.

Patro R., G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417-419.

Peeters S. H., L. van Niftrik, 2019. Trending topics and open questions in anaerobic ammonium oxidation. Curr. Opin. Chem. Biol. 49, 45-52.

Pellow D., I. Mizrahi, R. Shamir, 2020. PlasClass improves plasmid sequence classification. PLoS Comp. Biol. 16, e1007781.

Richter M., R. Rosselló-Móra, 2009. Shifting the genomic gold standard for the prokaryotic species definition. PNAS 106, 19126-19131.

Richter M., R. Rosselló-Móra, F. Oliver Glöckner, J. Peplies, 2015. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 32, 929-931.

Sangwan N., F. F. Xia, J. A. Gilbert, 2016. Recovering complete and draft population genomes from metagenome datasets. Microbiome 4, 8.

Seeleuthner Y., S. Mondy, V. Lombard, Q. Carradec, E. Pelletier, M. Wessner, J. Leconte, J. F. Mangot, J. Poulain, K. Labadie, R. Logares, S. Sunagawa, V. de Berardinis, M. Salanoubat, C. Dimier, S. Kandels-Lewis, M. Picheral, S. Searson, S. Pesant, N. Poulton, R. Stepanauskas, P. Bork, C. Bowler, P. Hingamp, M. B. Sullivan, D. Iudicone, R. Massana, J. M. Aury, B. Henrissat, E. Karsenti, O. Jaillon, M. Sieracki, C. de Vargas, P. Wincker, C. Tara Oceans, 2018. Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. Nat. Commun. 9, 310.

Sheng L., Z. Lei, M. Dzakpasu, Y.-Y. Li, Q. Li, R. Chen, 2020. Application of the anammox-based process for nitrogen removal from anaerobic digestion effluent: A review of treatment performance, biochemical reactions, and impact factors. J. Water Process Eng. 38, 101595.

Shi A., F. Fan, J. R. Broach, 2022. Microbial adaptive evolution. J. Ind. Microbiol. Biotechnol. 49, kuab076.

Song W., B. Wemheuer, S. Zhang, K. Steensen, T. Thomas, 2019. MetaCHIP: community-level horizontal gene transfer identification through the combination of best-match and phylogenetic approaches. Microbiome 7, 36.

Speth D. R., M. H. in 't Zandt, S. Guerrero-Cruz, B. E. Dutilh, M. S. M. Jetten, 2016.

Genome-based microbial ecology of anammox granules in a full-scale wastewater treatment system. Nat. Commun. 7, 11172.

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

Suarez C., P. Dalcin Martins, M. S. M. Jetten, S. Karačić, B. M. Wilén, O. Modin, P. Hagelia, M. Hermansson, F. Persson, 2022. Metagenomic evidence of a novel family of anammox bacteria in a subsea environment. Environ. Microbiol. 24, 2348-2360.

Tao Y., F. Xun, C. Zhao, Z. Mao, B. Li, P. Xing, Q. L. Wu, 2023. Improved assembly of metagenome-assembled genomes and viruses in Tibetan saline lake sediment by HiFi metagenomic sequencing. Microbiol. Spectrum 11, e03328-03322.

Tian R., D. Ning, Z. He, P. Zhang, S. J. Spencer, S. Gao, W. Shi, L. Wu, Y. Zhang, Y. Yang, B. G. Adams, A. M. Rocha, B. L. Detienne, K. A. Lowe, D. C. Joyner, D. M. Klingeman, A. P. Arkin, M. W. Fields, T. C. Hazen, D. A. Stahl, E. J. Alm, J. Zhou, 2020. Small and mighty: adaptation of superphylum Patescibacteria to groundwater environment drives their genome simplicity. Microbiome 8, 51.

Tu Q., L. Lin, L. Cheng, Y. Deng, Z. He, 2019. NCycDB: a curated integrative database for fast and accurate metagenomic profiling of nitrogen cycling genes. Bioinformatics 35, 1040-1048.

Wang D., Y. Meng, F. Meng, 2022. Genome-centric metagenomics insights into functional divergence and horizontal gene transfer of denitrifying bacteria in anammox consortia. Water Res. 224, 119062.

Yang Y., Z. Lu, M. Azari, B. Kartal, H. Du, M. Cai, C. W. Herbold, X. Ding, M. Denecke, X. Li, M. Li, J.-D. Gu, 2022. Discovery of a new genus of anaerobic ammonium oxidizing bacteria with a mechanism for oxygen tolerance. Water Res. 226, 119165.

Yuan C., J. K. Lei, J. Cole, Y. N. Sun, 2015. Reconstructing 16S rRNA genes in metagenomic data. Bioinformatics 31, 35-43.

Zhao Y. P., Y. Feng, J. Q. Li, Y. Z. Guo, L. M. Chen, S. T. Liu, 2018a. Insight into the aggregation capacity of anammox consortia during reactor start-up. Environ. Sci. Technol. 52, 3685-3695.

Zhao Y. P., S. F. Liu, B. Jiang, Y. Feng, T. T. Zhu, H. C. Tao, X. Tang, S. T. Liu, 2018b. Genome-centered metagenomics analysis reveals the symbiotic organisms possessing ability to cross-feed with anammox bacteria in anammox consortia. Environ. Sci. Technol. 52, 11285-11296.

Zheng W., S. Zhao, Y. Yin, H. Zhang, D. M. Needham, E. D. Evans, C. L. Dai, P. J. Lu, E. J. Alm, D. A. Weitz, 2022. High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome. Science 376, eabm1483.

731 **Table 1.** Characteristics of the 30 cMAGs obtained in this study.

ID	Phylum (Genus)	Genome size (Mb)	GC (%)	CDS	GTDB-Tk marker gene	Types of tRNA	Features (5S/16S/23S)
cMAG.1	Actinobacteriota (Unclassified)	2.2	68.58	2,260	118	23	1/1/1
cMAG.2	Armatimonadota (H1-ARM1)	2.9	60.95	2,651	111	23	1/1/1
cMAG.3	Bacteroidota (OLB6)	3.0	49.64	2,427	119	22	2/2/2
cMAG.4	Bacteroidota (BCD5)	3.3	36.15	2,825	119	22	1/1/1
cMAG.5	Bacteroidota (UTCHB3)	3.4	42.13	2,623	118	23	1/1/1
cMAG.6	Bacteroidota (SURF-28)	3.7	34.47	3,351	118	22	1/1/1
cMAG.7	Bacteroidota (OLB8)	3.9	65.21	3,252	119	22	1/1/1
cMAG.8	Bacteroidota (PHOS-HE28)	4.0	41.08	3,258	119	22	2/2/2
cMAG.9	Bacteroidota (UBA2336)	4.1	43.59	3,752	117	23	2/2/2
cMAG.10	Bdellovibrionota (Unclassified)	2.3	55.04	2,089	93	18	1/1/1
cMAG.11	Chloroflexota (UBA7227)	3.5	59.60	3,200	116	24	1/1/1
cMAG.12	Chloroflexota (UBA7227)	3.6	59.76	3,268	116	24	1/1/1
cMAG.13	Chloroflexota (UBA12294)	3.9	53.51	3,590	111	23	1/1/1
cMAG.14	Chloroflexota (UBA12294)	4.0	53.65	3,738	110	23	1/1/1
cMAG.15	Chloroflexota (OLB14)	4.1	54.13	3,820	113	23	1/1/1
cMAG.16	Chloroflexota (OLB14)	4.4	53.56	4,388	115	24	1/1/1
cMAG.17	Desulfobacterota (UBA5799)	2.8	57.87	2,559	119	23	1/1/1
cMAG.18	Patescibacteria (Unclassified)	1.0	44.78	956	89	20	1/1/1
cMAG.19	Patescibacteria (Unclassified)	1.1	39.76	982	84	21	1/1/1
cMAG.20	Patescibacteria (Unclassified)	1.2	60.26	1,204	95	21	1/1/1

cMAG.21	Patescibacteria (OLB21)	1.2	38.42	1,073	93	21	1/1/1
cMAG.22	Patescibacteria (Unclassified)	0.6	52.17	629	82	21	1/1/1
cMAG.23	Patescibacteria (BOKC01)	0.7	40.37	698	84	22	1/1/2
cMAG.24	Patescibacteria (UBA12026)	0.8	34.42	947	80	20	1/1/1
cMAG.25	Patescibacteria (UBA919)	0.9	36.84	827	90	22	1/2/1
cMAG.26	Patescibacteria (Unclassified)	0.9	41.50	858	87	21	1/1/1
cMAG.27	Planctomycetota (*Ca.* Jettenia)	3.5	41.60	3,145	108	23	2/2/2
cMAG.28	Planctomycetota (*Ca.* Jettenia)	3.8	39.42	3,292	107	24	2/2/2
cMAG.29	Planctomycetota (*Ca.* Jettenia)	4.0	40.01	3,510	107	24	2/2/2
cMAG.30	Proteobacteria (QUBU01)	3.4	66.68	3,021	118	23	1/1/1

732

Figure legends

Fig. 1. Workflow for filtering the HiFi cMAGs from three assemblies by hifiasm-meta, Flye, and Canu.

Fig. 2. The phylogenetic placement and abundance of cMAGs recovered from anammox granular sludge. The phylogenetic tree was inferred based on the multiple sequence alignments of conserved marker genes from 30 cMAGs using the GTDB-Tk.

Fig. 3. The MEBS scores for the nitrogen (N), iron (Fe), oxygen (O), carbon (C) and sulfur (S) metabolic pathways of 30 cMAGs.

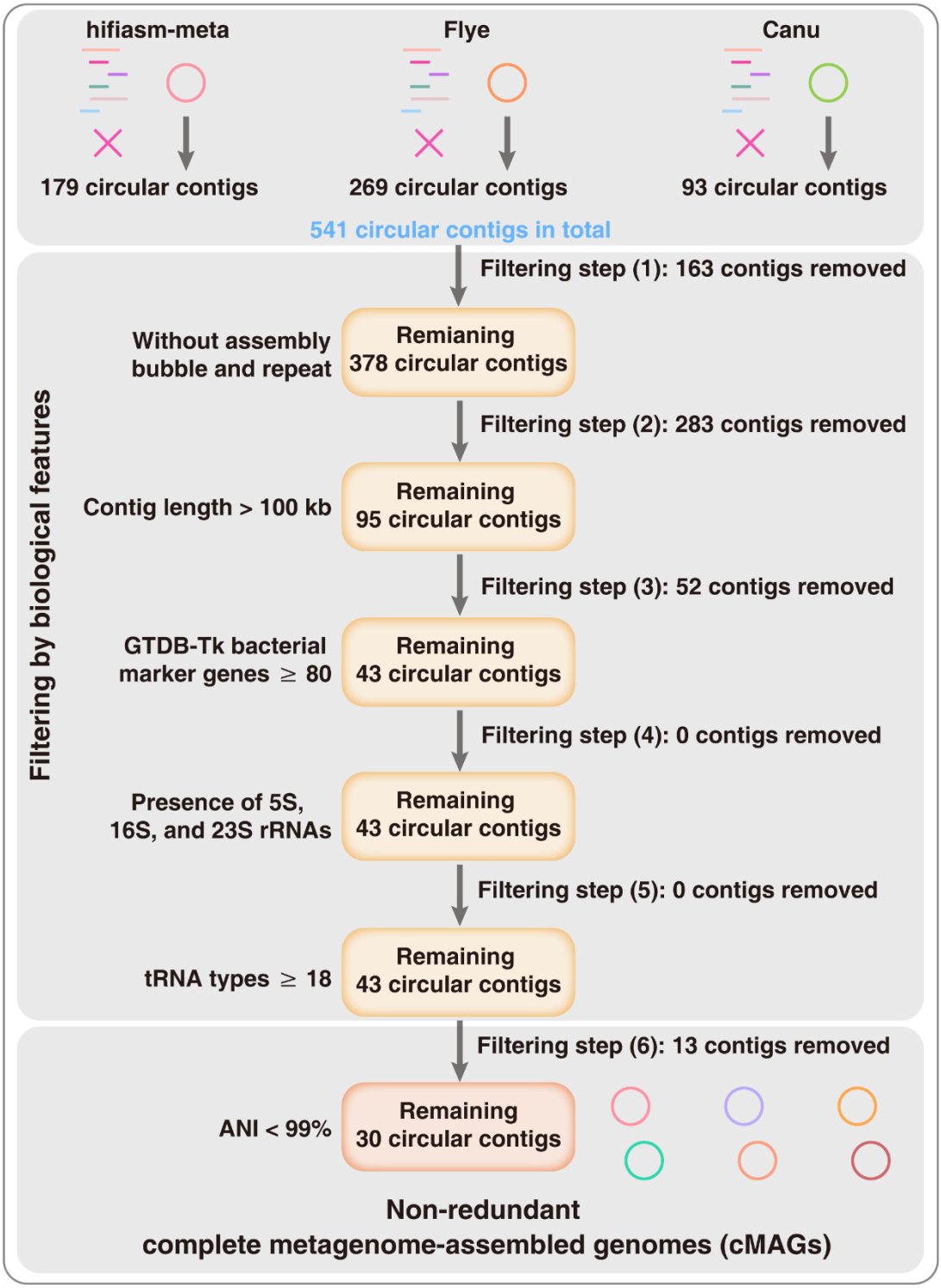
Fig. 4. The potential for nitrogen metabolism of 30 cMAGs. (**A**) Nitrogen metabolism marker genes. (**B**) Proposed pathway for nitrogen conversion in the anammox community. The number indicates the number of species with genes encoding a given step. Colored squares indicate gene abundances.

Fig. 5. Mobile genetic elements (MGEs) identified in the cMAGs and reference genomes. (**A**) MGEs identified in cMAGs. (**B**) The *tnpA* counts identified in two *Ca.* Jettenia species of this study and other anammox species.

Fig. 6. The predicted 28 gene clusters containing *tnpA* of cMAG.29.

Fig. 7. Network and functions of HGT genes of the anammox community. (**A**) HGT network of the anammox community; the direction and width of blue arrows indicate the direction of transfer and gene counts, respectively; the size of the node represents the count of connected edges. (**B**) and (**C**) Bubble size indicates the count of HGT genes

759 annotated with COG functional categories. C: Energy production and conversion; D:
760 Cell cycle control, cell division, chromosome partitioning; E: Amino acid transport and
761 metabolism; H: Coenzyme transport and metabolism; J: Translation, ribosomal
762 structure and biogenesis; L: Replication, recombination and repair; N: Cell motility; O:
763 Posttranslational modification, protein turnover; S: Function unknown; T: Signal
764 transduction mechanisms.


hifiasm-meta
Flye
Canu
179 circular contigs
269 circular contigs
93 circular contigs
541 circular contigs in total
Filtering step (1): 163 contigs removed
Filtering by biological features
Without assembly bubble and repeat
Remianing 378 circular contigs
Filtering step (2): 283 contigs removed
Contig length > 100 kb
Remaining 95 circular contigs
Filtering step (3): 52 contigs removed
GTDB-Tk bacterial marker genes ≥ 80
Remaining 43 circular contigs
Filtering step (4): 0 contigs removed
Presence of 5S, 16S, and 23S rRNAs
Remaining 43 circular contigs
Filtering step (5): 0 contigs removed
tRNA types ≥ 18
Remaining 43 circular contigs
Filtering step (6): 13 contigs removed
ANI < 99%
Remaining 30 circular contigs
Non-redundant complete metagenome-assembled genomes (cMAGs)

765

766 **Fig. 1**

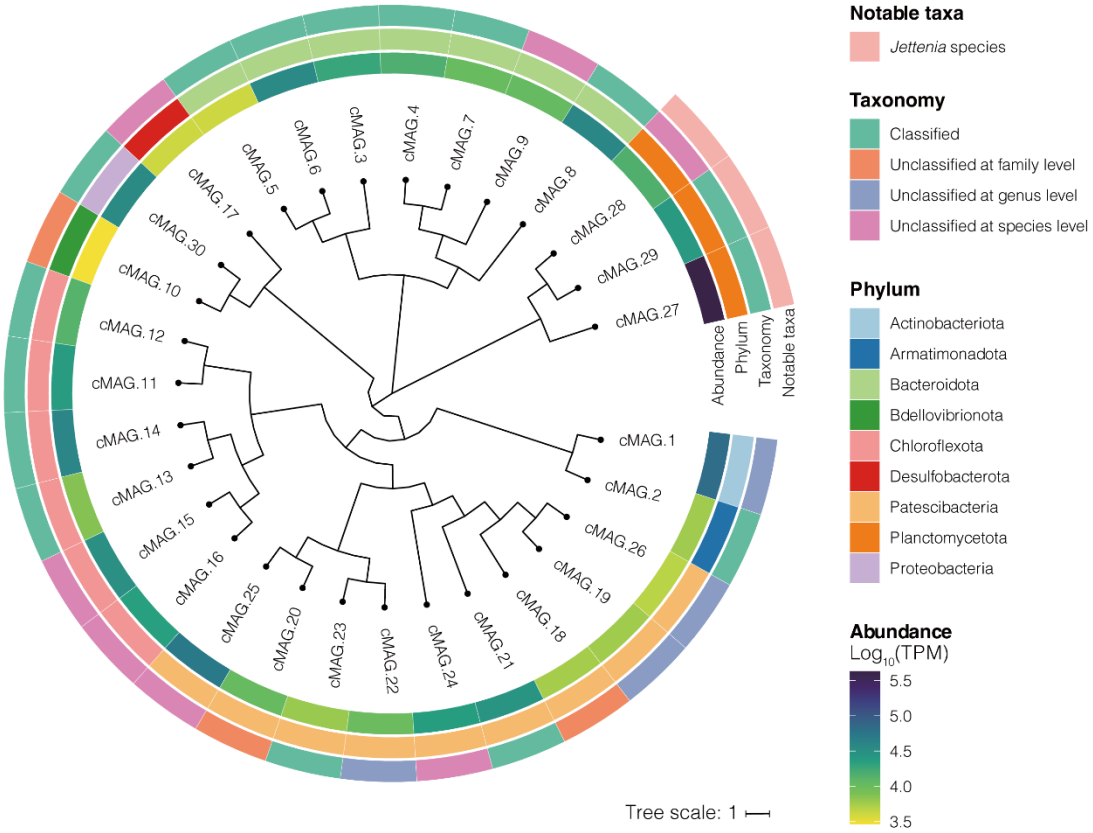

Notable taxa
Jettenia species
Taxonomy
Classified
Unclassified at family level
Unclassified at genus level
Unclassified at species level
Phylum
Actinobacteriota
Armatimonadota
Bacteroidota
Bdellovibrionota
Chloroflexota
Desulfobacterota
Patescibacteria
Planctomycetota
Proteobacteria
Abundance
Log10(TPM)
5.5
5.0
4.5
4.0
3.5
Abundance
Phylum
Taxonomy
Notable taxa
cMAG.1
cMAG.2
cMAG.26
cMAG.19
cMAG.18
cMAG.21
cMAG.24
cMAG.22
cMAG.23
cMAG.20
cMAG.25
cMAG.16
cMAG.15
cMAG.13
cMAG.14
cMAG.11
cMAG.12
cMAG.10
cMAG.30
cMAG.17
cMAG.5
cMAG.6
cMAG.3
cMAG.4
cMAG.7
cMAG.9
cMAG.8
cMAG.28
cMAG.29
cMAG.27
Tree scale: 1

Fig. 2

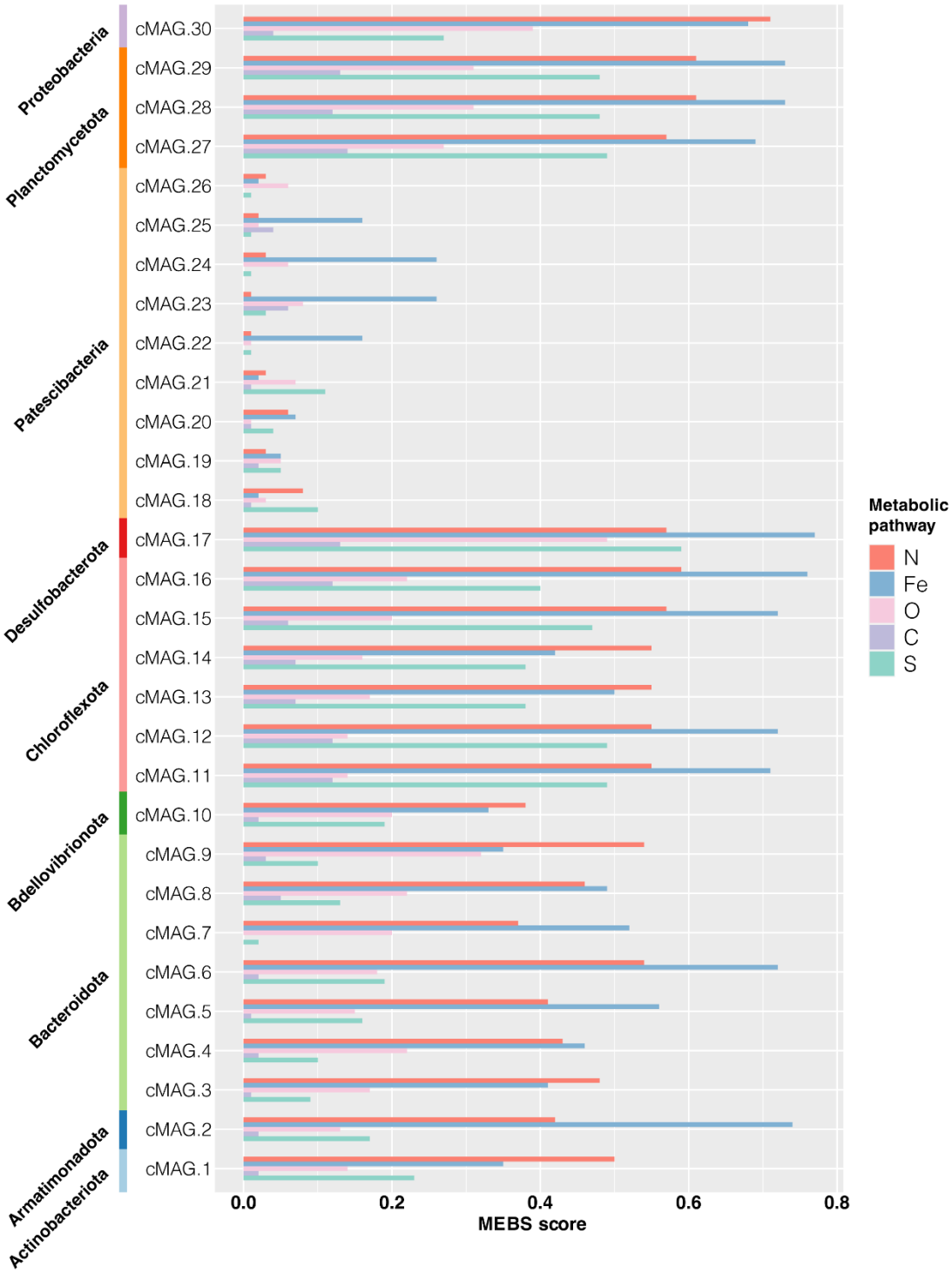

Proteobacteria
Planctomycetota
Patescibacteria
Desulfobacterota
Chloroflexota
Bdellovibrionota
Bacteroidota
Armatimonadota
Actinobacteriota
cMAG.30
cMAG.29
cMAG.28
cMAG.27
cMAG.26
cMAG.25
cMAG.24
cMAG.23
cMAG.22
cMAG.21
cMAG.20
cMAG.19
cMAG.18
cMAG.17
cMAG.16
cMAG.15
cMAG.14
cMAG.13
cMAG.12
cMAG.11
cMAG.10
cMAG.9
cMAG.8
cMAG.7
cMAG.6
cMAG.5
cMAG.4
cMAG.3
cMAG.2
cMAG.1
0.0
0.2
0.4
0.6
0.8
MEBS score
Metabolic pathway
N
Fe
O
C
S

Fig. 3

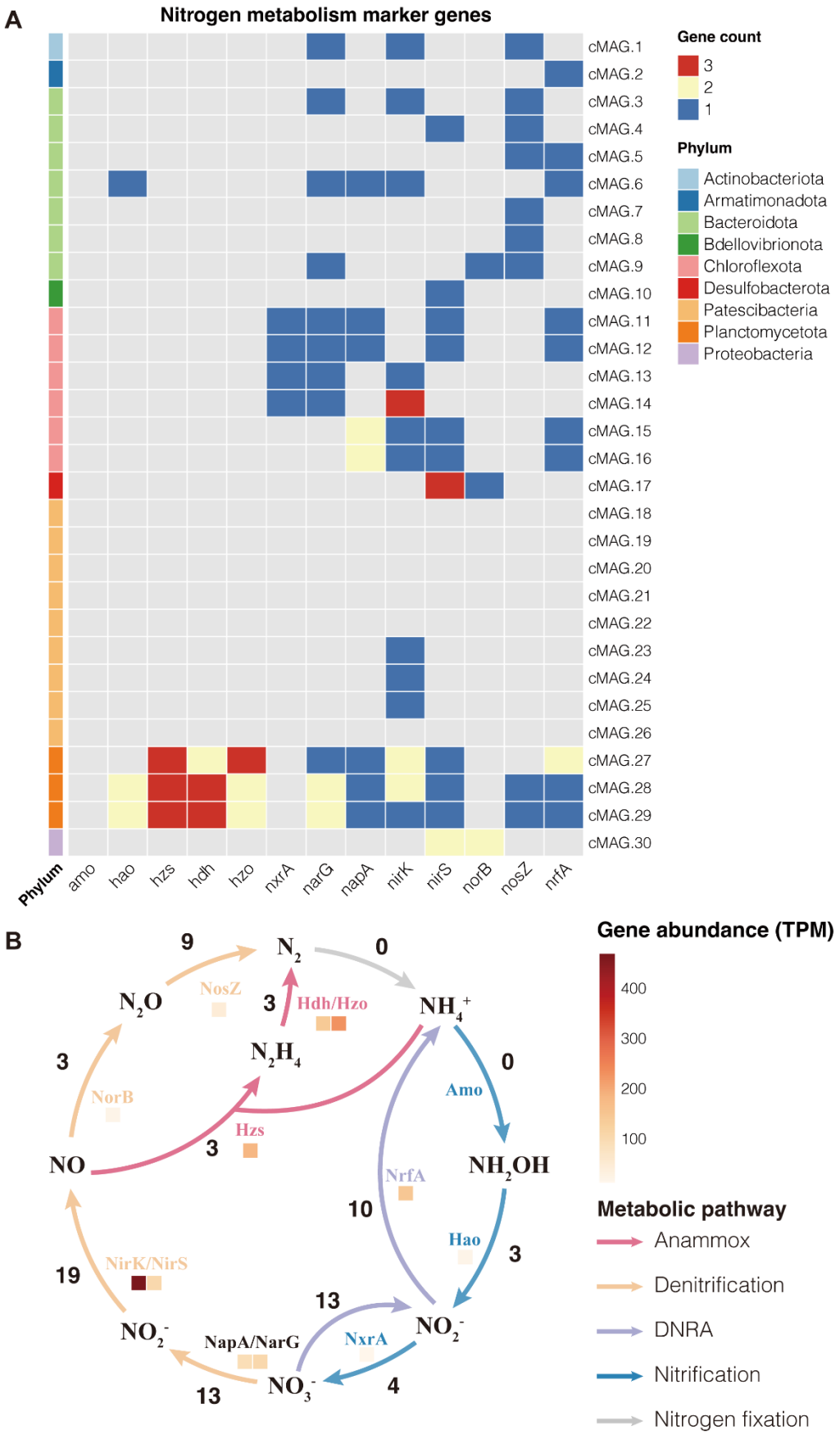

A
Nitrogen metabolism marker genes
cMAG.1
cMAG.2
cMAG.3
cMAG.4
cMAG.5
cMAG.6
cMAG.7
cMAG.8
cMAG.9
cMAG.10
cMAG.11
cMAG.12
cMAG.13
cMAG.14
cMAG.15
cMAG.16
cMAG.17
cMAG.18
cMAG.19
cMAG.20
cMAG.21
cMAG.22
cMAG.23
cMAG.24
cMAG.25
cMAG.26
cMAG.27
cMAG.28
cMAG.29
cMAG.30
Phylum
amo
hao
hzs
hdh
hzo
nxrA
narG
napA
nirK
nirS
norB
nosZ
nrfA
Gene count
3
2
1
Phylum
Actinobacteriota
Armatimonadota
Bacteroidota
Bdellovibrionota
Chloroflexota
Desulfobacterota
Patescibacteria
Planctomycetota
Proteobacteria
B
N2
N2O
N2H4
NH4+
NH2OH
NO
NO2-
NO3-
9
0
3
3
0
3
10
3
19
13
13
4
NosZ
Hdh/Hzo
NorB
Hzs
Amo
NrfA
Hao
NirK/NirS
NapA/NarG
NxrA
Gene abundance (TPM)
400
300
200
100
Metabolic pathway
Anammox
Denitrification
DNRA
Nitrification
Nitrogen fixation

Fig. 4

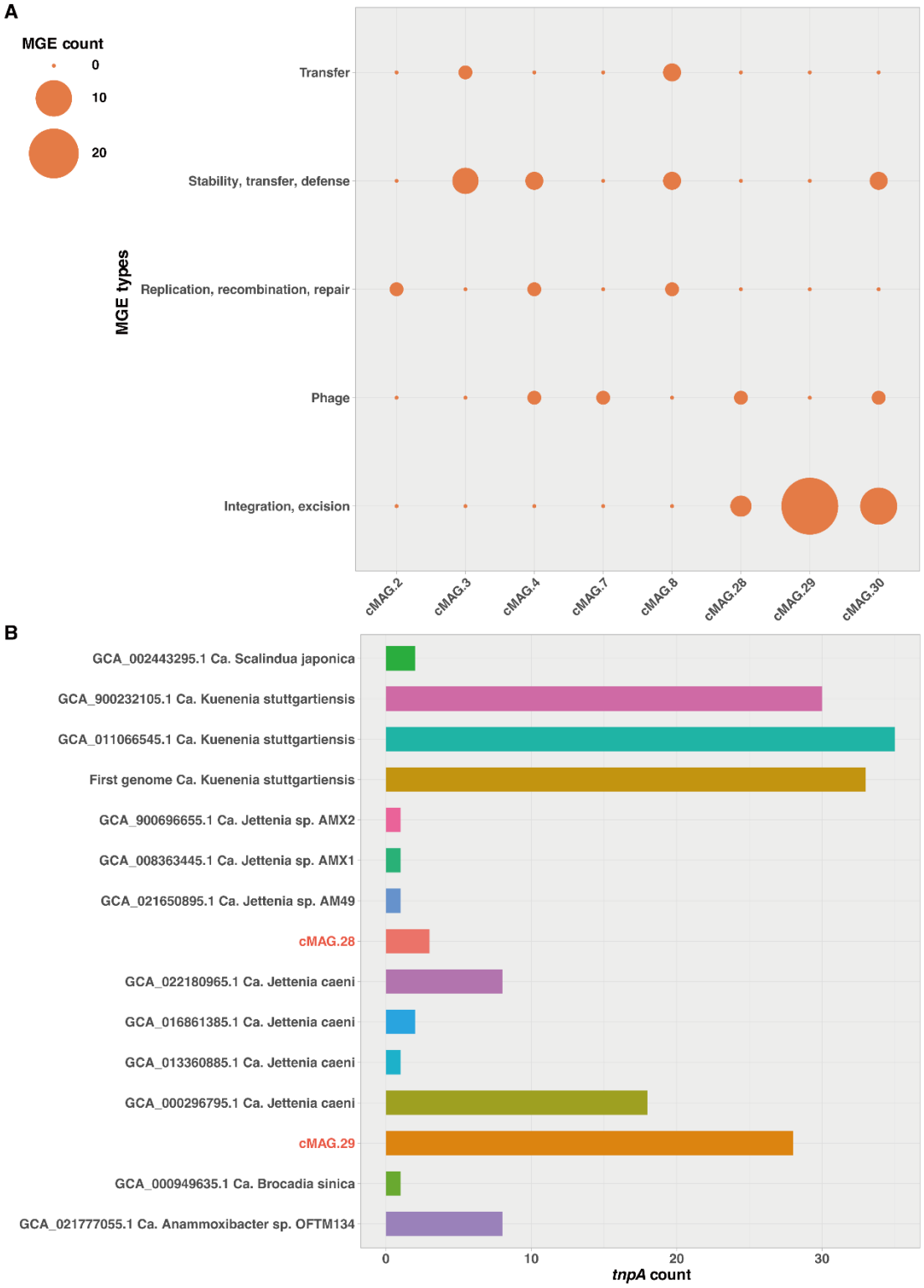

A
MGE count
0
10
20
MGE types
Transfer
Stability, transfer, defense
Replication, recombination, repair
Phage
Integration, excision
cMAG.2
cMAG.3
cMAG.4
cMAG.7
cMAG.8
cMAG.28
cMAG.29
cMAG.30
B
GCA_002443295.1 Ca. Scalindua japonica
GCA_900232105.1 Ca. Kuenenia stuttgartiensis
GCA_011066545.1 Ca. Kuenenia stuttgartiensis
First genome Ca. Kuenenia stuttgartiensis
GCA_900696655.1 Ca. Jettenia sp. AMX2
GCA_008363445.1 Ca. Jettenia sp. AMX1
GCA_021650895.1 Ca. Jettenia sp. AM49
cMAG.28
GCA_022180965.1 Ca. Jettenia caeni
GCA_016861385.1 Ca. Jettenia caeni
GCA_013360885.1 Ca. Jettenia caeni
GCA_000296795.1 Ca. Jettenia caeni
cMAG.29
GCA_000949635.1 Ca. Brocadia sinica
GCA_021777055.1 Ca. Anammoxibacter sp. OFTM134
0
10
20
30
tnpA count

Fig. 5

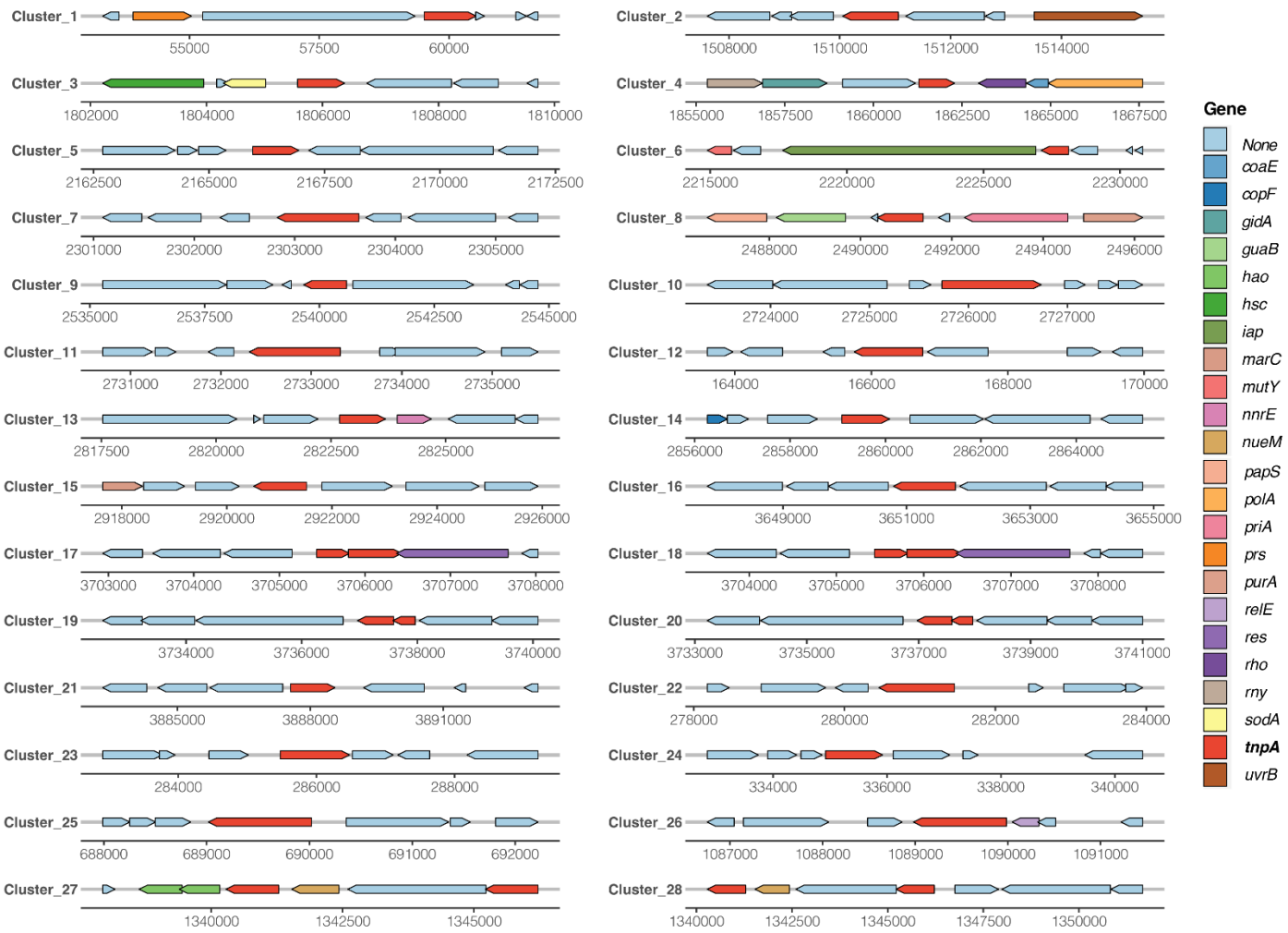

Cluster_1
55000 57500 60000
Cluster_2
1508000 1510000 1512000 1514000
Cluster_3
1802000 1804000 1806000 1808000 1810000
Cluster_4
1855000 1857500 1860000 1862500 1865000 1867500
Cluster_5
2162500 2165000 2167500 2170000 2172500
Cluster_6
2215000 2220000 2225000 2230000
Cluster_7
2301000 2302000 2303000 2304000 2305000
Cluster_8
2488000 2490000 2492000 2494000 2496000
Cluster_9
2535000 2537500 2540000 2542500 2545000
Cluster_10
2724000 2725000 2726000 2727000
Cluster_11
2731000 2732000 2733000 2734000 2735000
Cluster_12
164000 166000 168000 170000
Cluster_13
2817500 2820000 2822500 2825000
Cluster_14
2856000 2858000 2860000 2862000 2864000
Cluster_15
2918000 2920000 2922000 2924000 2926000
Cluster_16
3649000 3651000 3653000 3655000
Cluster_17
3703000 3704000 3705000 3706000 3707000 3708000
Cluster_18
3704000 3705000 3706000 3707000 3708000
Cluster_19
3734000 3736000 3738000 3740000
Cluster_20
3733000 3735000 3737000 3739000 3741000
Cluster_21
3885000 3888000 3891000
Cluster_22
278000 280000 282000 284000
Cluster_23
284000 286000 288000
Cluster_24
334000 336000 338000 340000
Cluster_25
688000 689000 690000 691000 692000
Cluster_26
1087000 1088000 1089000 1090000 1091000
Cluster_27
1340000 1342500 1345000
Cluster_28
1340000 1342500 1345000 1347500 1350000
Gene
None
coaE
copF
gidA
guaB
hao
hsc
iap
marC
mutY
nnrE
nueM
papS
polA
priA
prs
purA
relE
res
rho
rny
sodA
tnpA
uvrB

Fig. 6

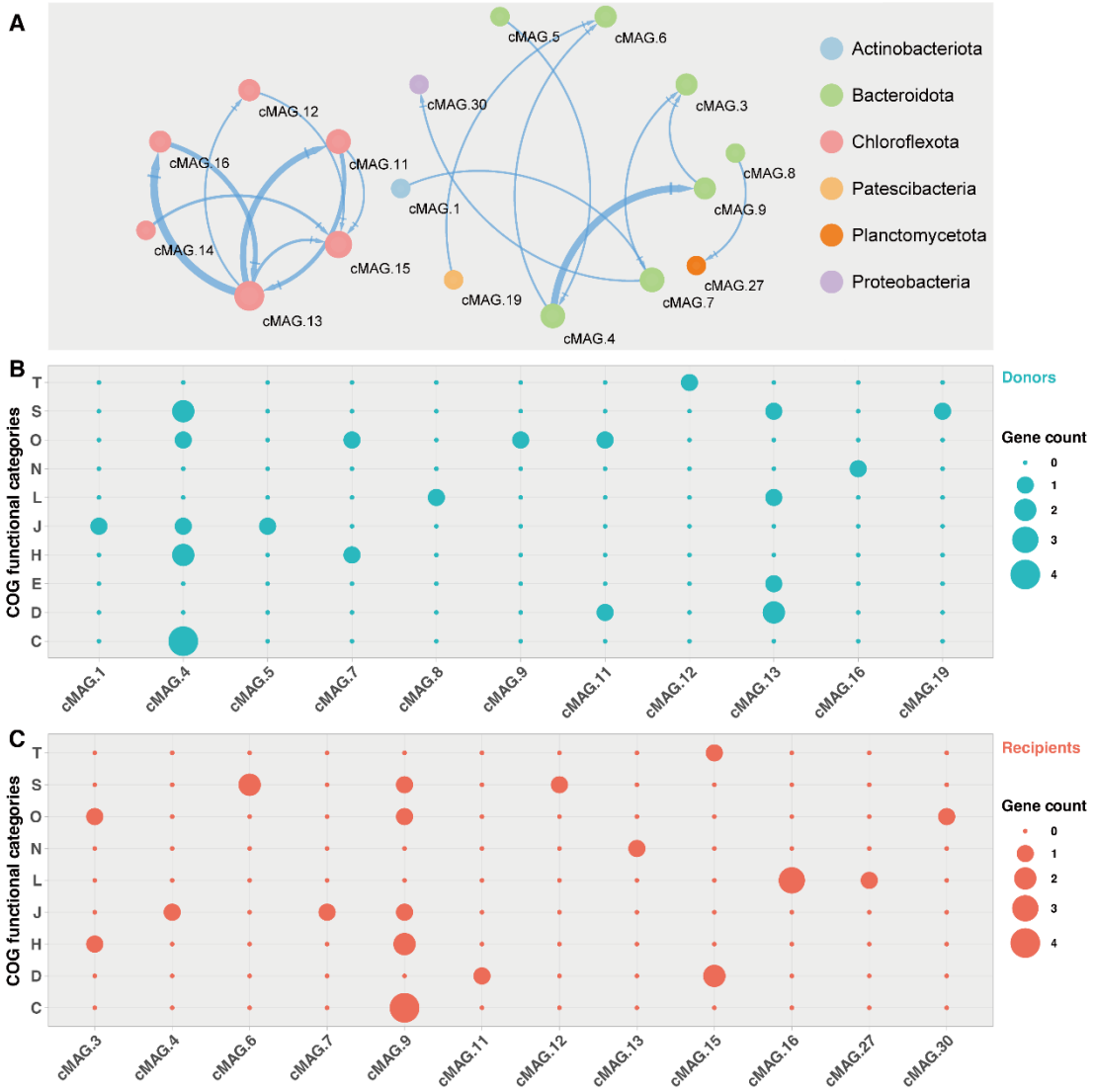

A
cMAG.5
cMAG.6
cMAG.12
cMAG.30
cMAG.3
cMAG.16
cMAG.11
cMAG.8
cMAG.1
cMAG.9
cMAG.14
cMAG.15
cMAG.27
cMAG.19
cMAG.7
cMAG.13
cMAG.4
Actinobacteriota
Bacteroidota
Chloroflexota
Patescibacteria
Planctomycetota
Proteobacteria
B
COG functional categories
T
S
O
N
L
J
H
E
D
C
cMAG.1
cMAG.4
cMAG.5
cMAG.7
cMAG.8
cMAG.9
cMAG.11
cMAG.12
cMAG.13
cMAG.16
cMAG.19
Donors
Gene count
0
1
2
3
4
C
COG functional categories
T
S
O
N
L
J
H
D
C
cMAG.3
cMAG.4
cMAG.6
cMAG.7
cMAG.9
cMAG.11
cMAG.12
cMAG.13
cMAG.15
cMAG.16
cMAG.27
cMAG.30
Recipients
Gene count
0
1
2
3
4

Fig. 7