ISME Communications, 2024, 4(1), ycae050
https://doi.org/10.1093/ismeco/ycae050
Advance access publication: 11 April 2024
Original Article

Unveiling the hidden diversity and functional role of Chloroflexota in full-scale wastewater treatment plants through genome-centric analyses

Patricia Bovio-Winkler[1,*], Angela Cabezas[2], Claudia Etchebehere[1]

[1]Microbial Ecology Laboratory, Department of Microbial Biochemistry and Genomic, Biological Research Institute "Clemente Estable", Avenida Italia 3318, 11600 Montevideo, CP, Uruguay
[2]Departamento de sostenibilidad ambiental, Instituto Tecnológico Regional Centro Sur, Universidad Tecnológica, Francisco Antonio Maciel s/n, 97000, Durazno, CP Uruguay

*Corresponding author: Patricia Bovio-Winkler, Microbial Ecology Laboratory, Department of Microbial Biochemistry and Genomics, Biological Research Institute "Clemente Estable", Av. Italia 3318. CP 11600 Montevideo, Uruguay. Email: pbovio@iibce.edu.uy

Abstract

The phylum Chloroflexota has been found to exhibit high abundance in the microbial communities from wastewater treatment plants (WWTPs) in both aerobic and anaerobic systems. However, its metabolic role has not been fully explored due to the lack of cultured isolates. To address this gap, we use publicly available metagenome datasets from both activated sludge (AS) and methanogenic (MET) full-scale wastewater treatment reactors to assembled genomes. Using this strategy, 264 dereplicated, medium- and high-quality metagenome-assembled genomes (MAGs) classified within Chloroflexota were obtained. Taxonomic classification revealed that AS and MET reactors harbored distinct Chloroflexota families. Nonetheless, the majority of the annotated MAGs (166 MAGs with >85% completeness and < 5% contamination) shared most of the metabolic potential features, including the ability to degrade simple sugars and complex polysaccharides, fatty acids and amino acids, as well as perform fermentation of different products. While Chloroflexota MAGs from MET reactors showed the potential for strict fermentation, MAGs from AS harbored the potential for facultatively aerobic metabolism. Metabolic reconstruction of Chloroflexota members from AS unveiled their versatile metabolism and suggested a primary role in hydrolysis, carbon removal and involvement in nitrogen cycling, thus establishing them as fundamental components of the ecosystem. Microbial reference genomes are essential resources for understanding the potential functional role of uncultured organisms in WWTPs. Our study provides a comprehensive genome catalog of Chloroflexota for future analyses aimed at elucidating their role in these ecosystems.

Keywords: Chloroflexota, methanogenic reactors, activated sludge, metagenome assembled genomes, meta-analysis

Introduction

Wastewater treatment systems are artificial ecosystems in which a microbial consortium degrades organic matter with the aim of cleaning the water. In this way, water with sufficient quality is obtained to be discharged into water courses without causing major environmental problems. The most conventional systems are activated sludge systems, which consist of an aerated reactor in which aerobic microorganisms degrade the organic matter [1]. For wastewater with a high organic matter concentration or for solid waste, it is more convenient to use anaerobic systems in which the organic matter is converted into methane by the action of a consortium of bacteria and archaea [2]. The advantage of anaerobic systems is that methane is obtained that can be used as fuel. Both systems are widely used on full scale wastewater treatment plants (WWTPs) and can be considered established technologies [1, 3]. However, the microbiology of these complex systems still presents several knowledge gaps. Despite the great difference between the two systems (aerobic and anaerobic) the phylum Chloroflexota have been detected in high abundance in the microbial communities from both methanogenic [4, 5] and activated sludge systems [6, 7]. Previous studies have indicated that Chloroflexota contributes to the formation of the filamentous matrix around which flocs and granules are formed [8, 9]. However, several authors have associated the overgrowth of some genera of Chloroflexota with bulking episodes and poor sludge–water separation, mainly in activated sludge systems but also in full-scale methanogenic reactors and lab-scale anammox reactors [10–14]. In activated sludge reactors, the biomass is flocculent in nature [15, 16]. Meanwhile, in granular sludge-based bioreactors, microbial biomass grows in the form of granules, which are small, self-immobilized, spherical, denser and more compact than activated sludge flocs [17]. In contrast, in anaerobic digester systems, the microbial biomass typically does not exhibit granulation but is instead dispersed [18, 19]. Bulking episodes occur when a significant portion of filaments extend beyond the confines of granules and flocs, protruding into the bulk water. In a recent study, most of the Chloroflexota members in activated sludge systems formed thin and short trichomes integrated into the floc structure, which are unlikely to form the typical inter-floc bridging that hinders the settling of activated sludge flocs

Received: 14 November 2023. **Revised:** 15 March 2024. **Accepted:** 8 April 2024

[20]. To understand under which conditions certain Chloroflexota members overgrow, it is necessary to isolate them in pure culture. However, despite reaching relative abundances of up to 35% in these ecosystems [21], Chloroflexota has long been considered a group of the yet-to-be-cultured microbes that are recalcitrant to cultivation and isolation [10, 22]. The most widely accepted hypothesis is that members of the Chloroflexota phylum are notoriously difficult to cultivate due to slow growth, particularly those belonging to the Anaerolineae class [23, 24]. Consequently, they are easily outcompeted by fast-growing heterotrophic bacteria.

Advancements in metagenome data analysis have made it possible to assemble genome sequences, providing access to detailed information about the metabolic potentials of these organisms [21, 25, 26]. Based on the physiology of isolates, assembled genomes from metagenomes and in situ characterization [20, 26], Chloroflexota members have been proposed to primarily function as heterotrophic and facultative anaerobic bacteria [20]. For instance, they are capable of hydrolyzing complex organic matter, fermenting carbohydrates and amino acids, and degrading cellular debris [4, 7, 8, 22, 26–32]. It has also been suggested that Chloroflexota members may play a role in the nitrogen cycle, thereby improving the nitrogen removal performance in anammox bioreactors [33, 34] and activated sludge systems [20, 21].

Due to the lack of pure cultures, microbial reference genomes are essential resources for expanding the phylogenomic representation of Chloroflexota and comprehend their functional role in WWTPs. This has direct implications for reactor performance, particularly in preventing and controlling bulking problems caused by their overgrowth.

In our previous investigations, we studied the diversity of Chloroflexota by analyzing amplicon sequencing data obtained from 62 full-scale methanogenic reactors. Then, we explored the potential metabolic role of 17 Chloroflexota members through genome-centric metagenomic analyses using samples collected from a methanogenic reactor, an activated sludge reactor and anammox reactor [5, 21, 35]. The outcomes of these studies present exciting prospects for addressing questions related to the taxonomic composition and potential metabolic functions of Chloroflexota members. This is accomplished with the added robustness of determining these aspects through a meta-analysis of metagenome-assembled genomes (MAGs).

In this study, we conducted an analysis of 264 medium- and high-quality metagenome-assembled genomes of Chloroflexota, retrieved from metagenomic public data from activated sludge and methanogenic full-scale reactor. The objective of our research was to address the following questions: Is the Chloroflexota taxonomic composition the same in activated sludge and methanogenic reactors? Does the metabolic potential of Chloroflexota differ between both systems? Which carbon compound degradation pathways do they have?

Materials and methods

Collection of public metagenomic data and metagenome-assembled genomes.

This meta-analysis included sequences from 18 published papers and data from the present work, which used shotgun sequencing to survey the microbial community in full-scale activated sludge (AS) and methanogenic (MET) reactors. We compiled the data from 87 full-scale reactors distributed in 15 countries, comprising 45 AS and 42 MET reactors (36 digesters and 6 upflow anaerobic sludge blanket (UASB) reactors) (Table 1, Supplementary Data 1). For the analysis, digesters and UASB type reactors were grouped into MET reactors.

Metagenome-assembled genomes (MAGs) from 11 studies were available in the National Center for Biotechnology Information (NCBI), Sequence Read Archive (SRA), and/or European Nucleotide Archive (ENA) databases. Metagenomic assembly and contig binning were conducted for two reactors from the current study and 16 reactors from seven additional studies (Table 1, Supplementary Data 1) [21, 36–42] since MAGs had not been previously assembled or were not available in public databases. All shotgun sequencing datasets were generated using the Illumina platform with paired-end sequencing strategy. For samples taken from both reactors in the present study we followed the protocols for sampling, DNA extraction and metagenome sequencing of samples described in a previous publication [21]. Briefly, total DNA was extracted from samples from two full-scale reactors located in Uruguay: a full-scale Internal Circulation (IC) methanogenic reactor treating effluent from a brewery (FNC) and a full-scale activated sludge reactor treating oil and grease (CO). For DNA extraction, sludge samples were thawed and centrifuged (5 min, 10 000 g). Approximately 0.35 g of wet pellet was used for DNA extraction with the ZR Soil Microbe DNA MiniPrepTM kit (Zymo Research) according to the manufacturer's instructions. The quality of the extracted genomic DNA was determined by 1% agarose gel electrophoresis (Nucleic Acid Stain, GoodViewTM, Beijing) and stored at −20°C until further use.

Metagenomic assembly and contigs binning

Default parameters were used with all software, unless otherwise specified. The overall quality of metagenomes reads from the present work and for the seven studies was assessed using FastQC (v0.11.8) [43]. Basic statistics of shotgun sequence statistics (number of reads, read length, dataset size) of all metagenomes used for assembly in this study are provided in Table S1. Reads were then trimmed using Trimmomatic (v0.39) [44] to remove adapters and bases below a quality score of 25 (HEADCROP:10, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:24, MINLEN:75). For each work, clean metagenomic reads were assembled into contigs using MEGAHIT v1.2.9 [45] (—k-min 21, —k-max 123, —k-step 4, —min-contig-length 1000). The contigs were then binned (—maxbin2, —concoct, —metabat2) and refined (–c 50, –x 10) using metaWRAP v1.3.2 [46]. To increase the completion of the bins, and reduce contamination, metaWRAP reassemble_bins module (–c 50 -x 10 options) was used. The resulting MAGs were analyzed along MAGs from the other 11 studies (Table 1). All MAGs were identified using GTDB-Tk v2.3.0 [47] and the Genome Taxonomy Database (GTDB) R214 [48] (Supplementary Data 1). The sequence statistics of the genomes were estimated using QUAST [49]. The completeness and contamination were estimated using CheckM2 (v1.0.1) [50].

To ensure stricter genome quality control, we selected Chloroflexota MAGs (retrieved from databases and assembled in the present work) with a completeness over 50%, a contamination level of less than 5%, a genome quality score (QS) greater than 50 (defined as the estimated completeness of a genome minus five times its estimated contamination) and free of chimerism as determined by GUNC [51]. We used a CSS of 0 for all genomes, a CSS closer to a value of 0 indicates that a genome is free of contamination and all genes are assigned to the same taxonomy, whereas a CSS score closer to 1 indicates chimerism. As a result, 522 high- and medium-quality MAGs were retained ($MAGs_{QS}$). From this set of MAGs obtained, we selected for diversity and

Table 1. Summary of the data analyzed indicating, reactor type, country, number of reactors and the reference.

Reactor type	Country	Reactors	Reference
Aerobic activated sludge reactors (45 reactors)	Uruguay	1	This work
	Uruguay	1	Bovio-Winkler et al. 2023 [5]
	Singapore	1	Haryono et al. 2021 [52]
	China	1	Liang et al. 2021[a] [42]
	Argentina	1	Pérez et al. 2019 [53]
	Germany	2	Schneider et al. 2021 [54]
	Denmark	23	Singleton et al. 2021 [55]
	Hong Kong	1	Wang et al. 2021 [56]
	Denmark	1	Ye et al. 2020 [57]
	China	8	
	USA	1	
	Argentina	1	
	Slovenia	1	
	Singapore	1	
	Taiwan	1	Yang et al. 2020[a] [45]
Methanogenic digesters (36 reactors)	Sweden	4	Brandt et al. 2020 [58]
	Germany	2	
	Denmark	1	Campanaro et al. 2020 [59]
	Sweden	1	
	Spain	2	
	Denmark	2	Hao et al. 2020[a] [60]
	China	2	Li et al. 2022[a] [41]
	China	6	Ma et al. 2021 [61]
	France	1	Puig-Castellví et al. 2021[a] [43]
	UK	8	Raguideau et al. 2021 [62]
	China	5	Fan et al. 2022[a] [46]
	Germany	2	Schneider et al. 2021 [54]
Methanogenic UASB type reactor (6 reactors)	Uruguay	1	This work
	Uruguay	1	Bovio-Winkler et al. 2023 [5]
	China	2	Liang et al. 2021[a] [42]
	Japan	1	Park et al. 2020[a] [63]
	Taiwan	1	Yang et al. 2020[a] [45]

[a]Indicates in which works we performs the assembly and binning.

metabolism analysis (see further details below) the highest number of MAGs possible for each case. For diversity and phylogenomic analysis, we used dereplicated MAGs (completeness and redundancy score equal to or greater than 50). Meanwhile, for metabolism analysis, dereplicated MAGs with a completeness greater than 85% were utilized, as there is evidence suggesting that the completeness of genomes significantly influences the recovered functional signal [64]. We dereplicated the $MAGs_{QS}$ at 95% average nucleotide identity (ANI) using dRep v3.2.2 [65] with the parameters -comp 50 -con 10 -sa 0.95 to identify the representative species in the genome tree (where ANI refers to alignment of at least 95% similarity for any section of the genome spanning a minimum of 10% of the bin length). These $MAGs_{dRep}$ were used for phylogenomic and taxonomic analyses. The 16S rRNA gene sequences of Chloroflexota $MAGs_{dRep}$ were identified using barrnap v0.993 (https://github.com/tseemann/barrnap).

Phylogenomic analysis

To determine the phylogenetic position of the 264 $MAGs_{dRep}$ (and 96 reference genomes retrieved from NCBI in February 2021) a phylogenomic tree based on concatenated alignments of 120 single copy marker genes was constructed using FastTree v2.1.11 (JTT, SH support values) applying the JTT protein substitution model for tree inference. From 264 $MAGs_{dRep}$, 16 MAGs were renamed according to [20]. Newick format tree file was uploaded to iTOL v6, a web-based tool for annotating and editing trees [66].

Gene annotation

Gene annotation was performed using $MAGs_{dRep}$ with more than 85% completeness and less than 5% contamination resulting in 166 MAGs (MAG_{sannot}, 102 MAGs for AS and 64 MAGs for MET reactors). The annotation was carried out with the "annotate" function of EnrichM v0.6.5 (https://github. com/geronimp/enrichM), using the Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthologies (KOs) [67] and Carbohydrate Active enzyme (CAZy) database [68]. EnrichM's "classify" function was used to calculate the completeness of KEGG modules. "Modules" are groups of genes organized by steps in a metabolic pathway as defined by KEGG. Only KEGG modules with 100% completeness in at least one MAG were kept in the downstream analyses.

Statistical analysis and data visualization

Statistical analyses and graphs were performed in R [69] using the following packages: vegan [70], ampvis2 [71], ggplot2 [72] and pheatmap [73]. Shapiro–Wilk's test was used to determine whether the estimated genome size (Mb) and coverage were

normally distributed. Since the estimated genome size was not normally distributed, a Kruskal-Wallis test (used to test equality of means when data is not normally distributed) was used to compare the estimated genome size (Mb) between AS and MET reactors. To determine significant differences in the coverage of Chloroflexota between AS and MET reactors, the Mann–Whitney U test (to test equality of means, Bonferroni corrected) were applied using Past (v3.21) [74]. Hierarchical cluster analysis (Euclidean distance) was used to cluster the CAZy families and KEGG pathways.

Results and discussion

Chloroflexota metagenome assembled-genomes dataset

To explore the diversity and potential metabolic role of the phylum Chloroflexota in full-scale WWTPs, we compiled publicly available data from 18 studies comprising 87 reactors distributed in 15 countries (Fig. 1A, Table 1, Supplementary Data 1). The reactors were divided into AS (45 reactors) and MET (42 reactors) (Table 1). After applying the quality score of >50 and dereplication, we retained 264 $MAGs_{dRep}$ (206 for AS and 108 for MET) (Supplementary Data 1). $MAGs_{dRep}$ were estimated to be 91.1% (median) complete, 2.8% (median) contaminated, with an estimated genome size of 4.5 Mb (median) and a GC content of 57.4% (Fig. 1b). The estimated genome size (Mb) in $MAGs_{dRep}$ from AS was significantly larger than in MET reactors (Kruskal–Wallis P-value <0.05) (Fig. 1C). Some ecological factors are correlated with genome size [75]. For example, oxygen is known to promote larger genome sizes [76]; there is a negative correlation between genome size and temperature [77, 78] and species with larger genome-sizes may dominate environments where resources are scarce but diverse [79–82]. Activated sludge reactors are oxygen-rich ecosystems, most of them treat sewage containing variable compounds and operational temperatures are lower than in digesters (Supplementary Data 1). For that reason, species from Chloroflexota in these environments may possess a more diverse set of genes and larger genome sizes. In contrast, MET reactors, which operate in anaerobic conditions, have higher temperatures and treat substrates such as sludge, manure, and other materials with less variability and high concentration of organic matter [83]. These conditions favor the presence of small genomes.

Different Chloroflexota communities dominate aerobic and methanogenic reactors

The Chloroflexota community taxonomic composition determined using $MAGs_{dRep}$ (n = 264) revealed that the class Anaerolineae represented 81% and 84% of these $MAGs_{dRep}$ in AS and MET reactors, respectively (Fig. 2A). These results were in accordance with earlier reports in WWTPs [5, 84]. Within Anaerolineae class, $MAGs_{dRep}$ belonging to Anaerolineales (29.8%), Promineofilales (21.5%) and Caldilineales (9.3%) orders were the most abundant in AS reactors (Fig. 2B). Meanwhile, Anaerolineales (50.4%) and Aggregatilineales (8.4%) were the most abundant orders in MET reactors. Dehalococcoidia class was the second most abundant class in AS (10.2%) and MET (13.4%) reactors (Fig. 2a). Within Dehalococcoidia class, Tipidiformales was only present in AS reactors while Dehalococcoidales only in MET reactors. Coverage values were collected for 54% of the $MAGs_{QS}$ (coverage was used as a proxy for the relative abundance of MAGs in complex communities), as this is a value frequently reported for MAGs in most studies (Supplementary Data 1), unlike relative abundance. The $MAGs_{QS}$ from MET showed higher coverage than those from AS reactors (normal distribution Shapiro–Wilk P-value = 0.00049, Mann–Whitney U test (Bonferroni corrected) P-value = 0.0004924), indicating a greater abundance of Chloroflexota in methanogenic systems (Fig. S1) There are no reports comparing the abundance of Chloroflexota genomes between AS and MET reactors, thus our results would be of interest from both ecological and process perspectives. In our previous study, amplicon sequencing results showed that methanogenic reactors presented a higher relative abundance of Chloroflexota than activated sludge reactors [21]. Most of the families and genera of Chloroflexota were not shared between AS and MET reactors (Fig. 2C and D). These results were clearly depicted in the phylogenomic tree constructed using $MAGs_{dRep}$ (Fig. 3), where distinct clusters at the family level were formed based on AS or MET reactors. In a recent work by Petriglieri [20], 29 newly described species of Chloroflexota were designated as *Candidatus*. We incorporated 21 of these species covering all Chloroflexota genera studied in this work into our dataset and phylogenomic tree. Additionally, we recovered 16 MAGs that met the criteria for high-quality (HQ) draft MAGs (Supplementary Table 2), in accordance with minimum information about a MAG (MIMAG) standard [85] ($MAGs_{MIMAG}$). These $MAGs_{MIMAG}$ are within clusters without representative species (isolates or Candidatus membes) (Fig. 3 clusters A-O), and were not classified at the genus level (3 MAGs) or species level (16 MAGs) using the GTDBtk database. This suggest that they are new species, highlighting the importance of the genomics analyses of Chloroflexota. Microbial reference genomes are essential resources for understanding the functional role of specific organisms in the different ecosystems. However, according to our results, an estimated 68% of Chloroflexota species analyzed in the present work lack a reference genome (Supplementary Data 1).

Potential role of Chloroflexota as hydrolytic and heterotrophic bacteria

To confirm the potential capacity of Chloroflexota to recycle soluble microbial products by acting as hydrolytic bacteria, we performed the annotation of $MAGs_{annot}$ (166 MAGs) using Carbohydrate-active Enzymes (CAZy) database. Overall, all $MAGs_{annot}$ were enriched on average in 8.6 classes (ranging from 0 to 25 of 1414) of glycoside hydrolases (GH), 4.9 classes (ranging from 0 to 11 of 801) of glycosyltransferases (GT) and, 3 classes (ranging from 0 to 6 of 487) of carbohydrate esterases (CE) (Fig. S2). This evidence suggests that polysaccharide degradation may represent the most widespread functional activities in the phylum Chloroflexota (Figs 4 and S3, Supplementary Data 2). These CAZy enzymes are capable of breaking down various substrates, including starch (GH77), lignocellulose (GH1, GH2, CE1), xylan polymers (CE7), lignin (AA3) and glycogen (endoglucanase and beta-glucosidase, blgX and/or blgB).

The differences in classes of CAZy enzymes between $MAGs_{AS}$ and $MAGs_{MET}$ were limited (MAGs from activated sludge reactors: $MAGs_{AS}$; MAGs from methanogenic reactors: $MAGs_{MET}$). Among $MAGs_{AS}$, AA3 (ligninolytic enzymes), GH5 (endo-type cellulases), GH36 (α-Galactosidase), and CE3 (acetylxylan esterases) were more prevalent compared to $MAGs_{MET}$ (Fig. 4). In contrast, $MAGs_{MET}$ showed a higher prevalence of GH25, which is associated with cell wall degradation, such as lysozyme.

The presence of cellulolytic enzymes in genomes from MET reactors (fed with primary sludge from AS reactors) and AS reactors (treating municipal wastewater) can be attributed to their high cellulose content, which originates mainly from toilet paper and constitutes about 35% of the suspended solids in the influent [86, 87]. Furthermore, in our set of $MAGs_{annot}$, several MET reactors

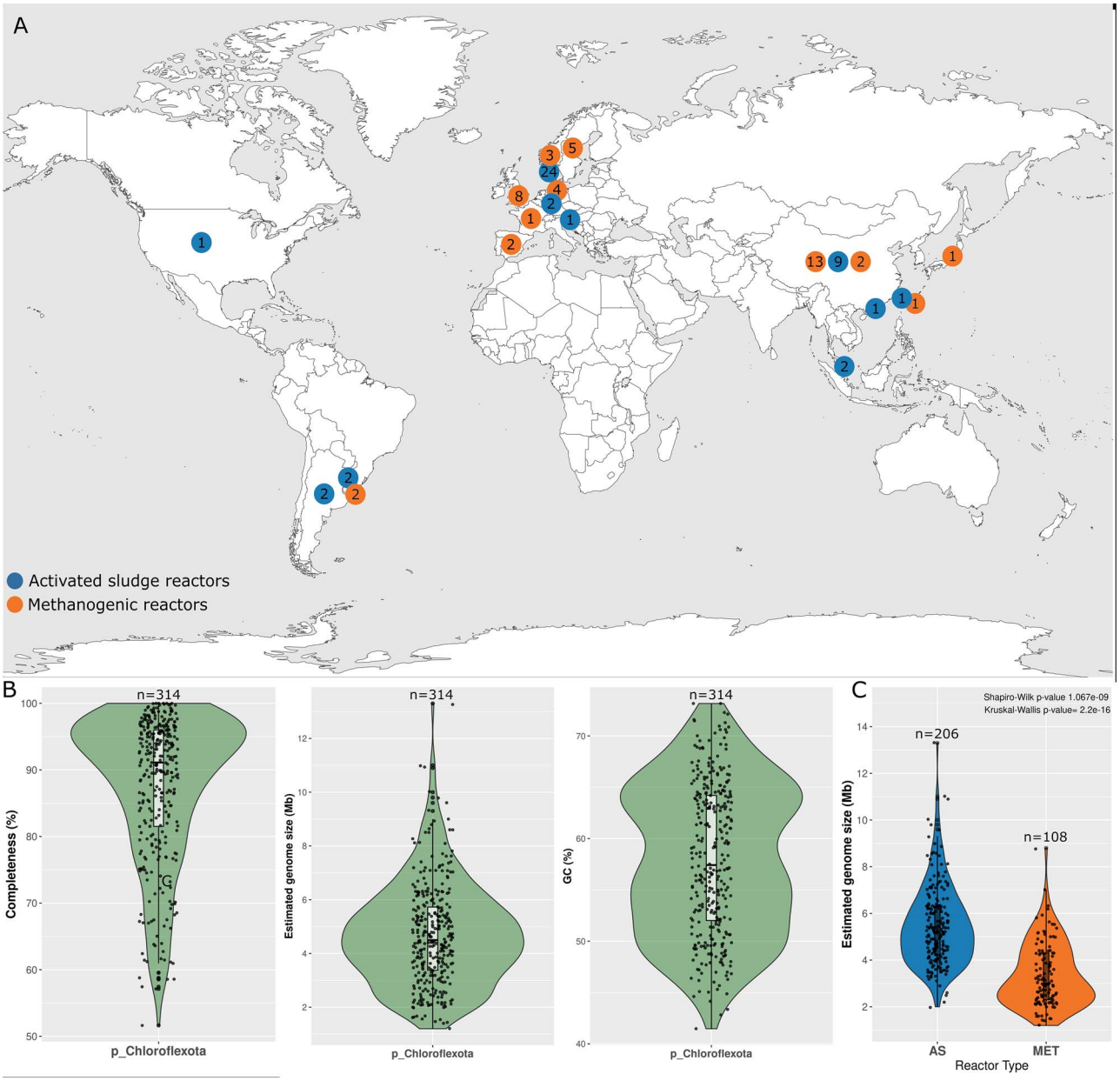

Figure 1. A) Geographic distribution of the 87 reactors. The numbers inside the circles indicate the number of reactors in each country. B) Violin plots showing the estimated completeness (%), estimated genome size (Mb) and GC (%) content for the representative species of Chloroflexota ($MAGs_{dRep}$ = 264) obtained from all WWTPs. C) Violin plots showing the comparison of the estimated genome size (Mb) between $MAGs_{dRep}$ from AS and MET reactors (Shapiro–Wilk normality test <0.05, non-parametric test Kruskal–Wallis *P*-value <0.05).

treated plant material. The cellulolytic potential of Chloroflexota was previously reported in Ktedonobacteria lineage (presence of acetylxylan esterases belonging to the CE1) [88] and some Anaerolineae members (several GHs families) [89].

The common ability of Chloroflexota members to grow on starch (alpha-glucan polysaccharides) [30, 31] and cellulose [90] is important considering the bottlenecking polysaccharide hydrolysis step of anaerobic digestion. This is in accordance with previous in situ studies, which revealed high levels of surface associated hydrolytic enzymes and their involvement in the breakdown of complex organic compounds [8, 91].

The diverse repertoire of CAZyme genes provides the basis for a flexible carbohydrate metabolism within the microbial community [86]. Under carbon-deficient conditions that prevail in nutrient removal WWTPs, carbon and energy sources supporting further growth of Chloroflexota members may originate from sugars released from the hydrolysis of cellulose, exopolysaccharides and cellular detritus [20, 26].

Aerobic and anaerobic uptake of different substrates is a shared trait exhibited by members of Chloroflexota [8, 9]. This widespread characteristic, regardless of the reactor type, was confirmed in all $MAGs_{annot}$, as genes encoding for different transporters were identified (Fig. 5): ABC-2 type (93% $MAGs_{AS}$, 90% $MAGs_{MET}$), Peptide/nickel (79% $MAGs_{AS}$, 81% $MAGs_{MET}$), Lipopolysaccharide (75% $MAGs_{AS}$, 82% $MAGs_{MET}$), Multiple sugar (74% $MAGs_{AS}$, 81% $MAGs_{MET}$), Branched-chain amino acid (86% $MAGs_{AS}$, 63% $MAGs_{MET}$) and Polar amino acid (68% $MAGs_{AS}$, 79% $MAGs_{MET}$).

Furthermore, other transporters, such as Putative multiple sugar (57% $MAGs_{AS,}$ 29% $MAGs_{MET}$), Glucose/mannose (50%

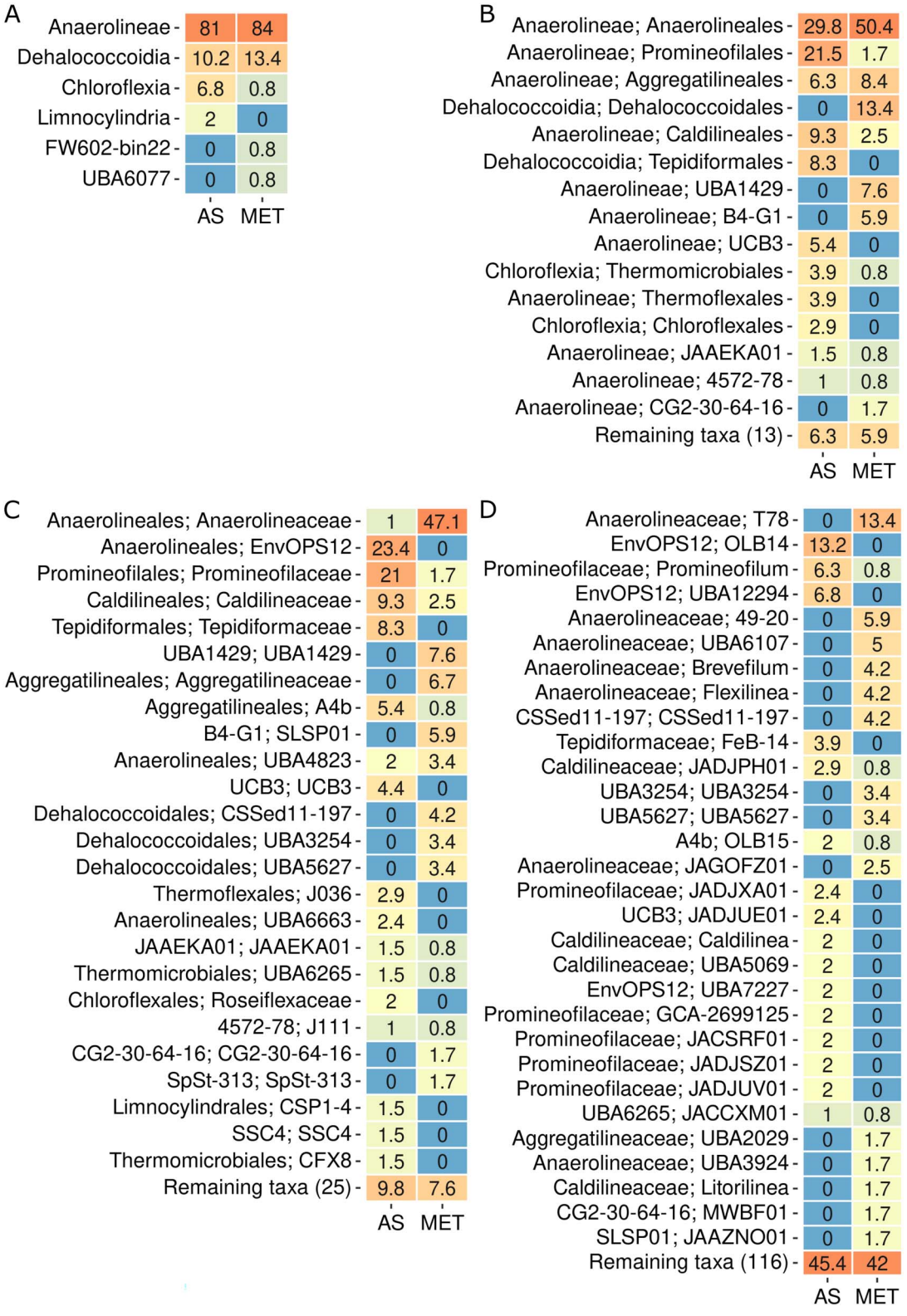

Figure 2. Heatmap showing the relative abundance of the Chloroflexota $MAGs_{dRep}$ in activated sludge (AS) and methanogenic reactors (MET) classification at A) class level, B) order level, C) family level and D) genus level.

$MAGs_{AS}$, 25% $MAGs_{MET}$), Ribose (49% $MAGs_{AS}$, 25% $MAGs_{MET}$), General L-amino acid (44% $MAGs_{AS}$, 22% $MAGs_{MET}$), Simple sugar (32% $MAGs_{AS}$, 16% $MAGs_{MET}$), Oligopeptide (23% $MAGs_{AS}$, 11% $MAGs_{MET}$), were more frequently identified in MAGs derived from activated sludge reactors. These findings indicated that the majority members of Chloroflexota could take up a wide range of sugars, fatty acids and amino acids as energy sources (Figs 5 and 6). This was demonstrated in previous experimental studies; wherein amino acid uptake was observed in both strictly anaerobic and facultatively aerobic Chloroflexota members

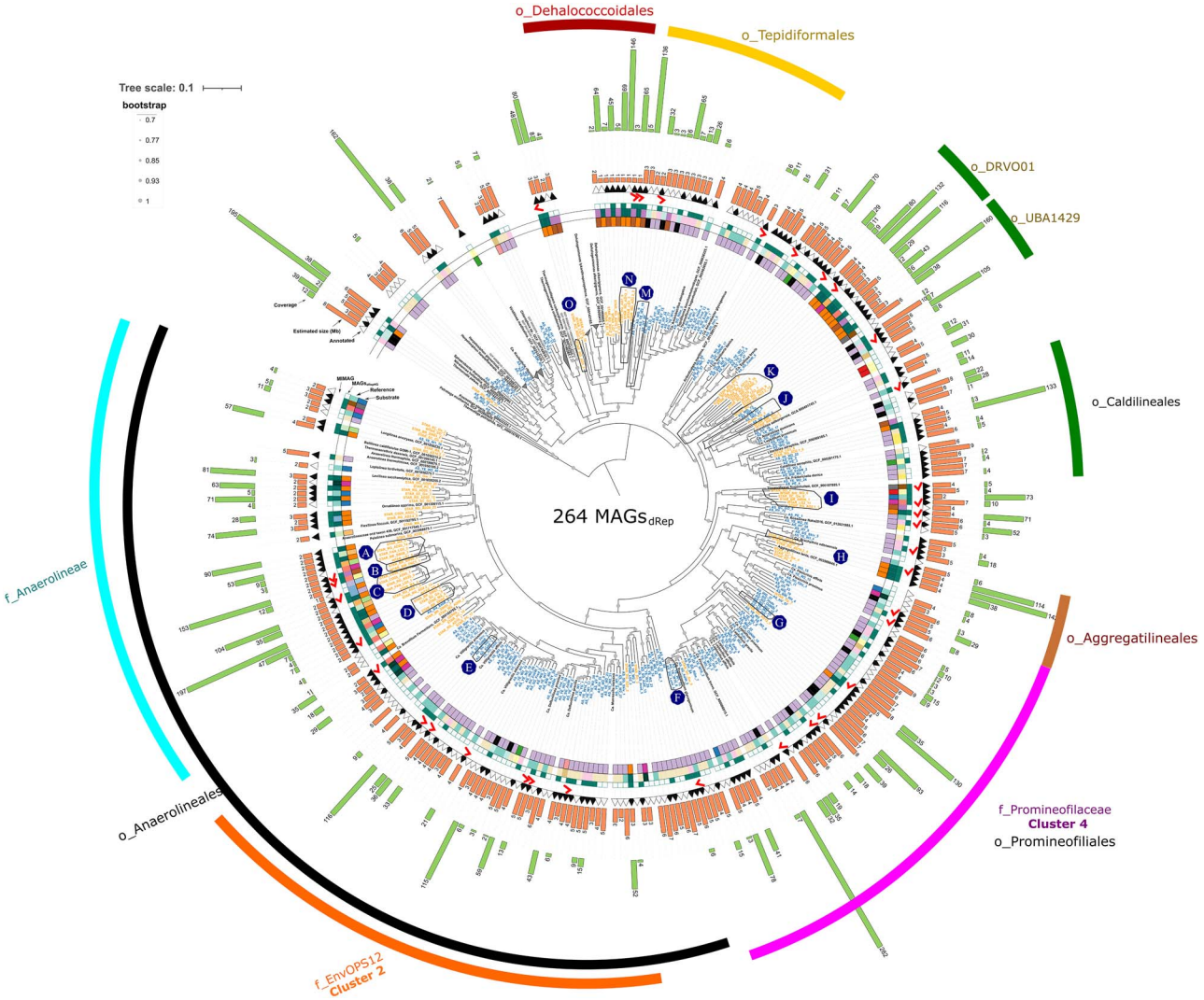

Figure 3. Phylogenomic tree of Chloroflexota $MAGs_{dRep}$ (n = 264) and reference genomes retrieved from NCBI (n = 96). Letters from A to O indicate clusters containing high-quality (HQ) draft MAGs with no representative cultures. Genomes from the phylum Thermotogota were used as outgroup to root the tree.

[22, 30]. Genes encoding beta-oxidation, responsible for the degradation of fatty acids and branched-chain amino acids, were identified in more MAGs from AS reactors (69% $MAGs_{AS}$, 8% $MAGs_{MET}$), indicating a prevalent characteristic among them. While Chloroflexota species do exhibit a clear preference for simple sugars, complex polymers and amino acids, species capable of consuming short and long fatty acids have also been identified [8, 9]. Therefore, it might represent an important metabolic route for some Chloroflexota species to acquire carbon and reducing equivalents (Figs 5 and 6).

Glycolysis pathway (Embden–Meyerhof-Parnas) was complete in 40% $MAGs_{AS}$ and 66% $MAGs_{MET}$ (Figs 5 and 6, Supplementary Data 3). The remaining $MAGs_{annot}$ could still achieve the glycolysis through a metabolic loop involving non-oxidative pentose phosphate pathway and the core module involving three-carbon compounds of glycolysis (75% $MAGs_{AS}$ and 73% $MAGs_{MET}$) (Figs 5 and 6) [92].

TCA cycle was found to be complete in 86% of $MAGs_{AS}$ (31% $MAGs_{MET}$) suggesting that terminal oxidation via the TCA serves as the primary energy source for members of Chloroflexota members in activated sludge systems (Figs 5 and 6). Additionally, the Glyoxylate cycle was identified in 31% of $MAGs_{AS}$ (2% $MAGs_{MET}$), suggesting the potential ability to utilize C2 compounds via the glyoxylate cycle for energy generation.

Furthermore, it was observed that 89% of $MAGs_{AS}$ and 24% of $MAGs_{MET}$ encoded acetyl coenzyme A synthetase (acsA), allowing for the conversion of acetate (a short-chain fatty acid) into acetyl-CoA. This finding is consistent with both isolated members and metatranscriptomic reports [33, 90]. This feature could be particularly advantageous in acetate-rich wastewater or in the absence of glucose [93]. It is worth noting that the acetate transporter gene (actP) was only present in 14% of $MAGs_{AS}$ and 5% of $MAGs_{MET}$, suggesting that these organisms may primarily rely on internal pools for the utilization of acetate as a carbon source [20]. Additionally, the majority of $MAGs_{annot}$ (89% $MAGs_{AS}$, 82% $MAGs_{MET}$) were found to possess genes encoding formate dehydrogenase (fdh), which could potentially be used to reduce formate generated during anaerobic fermentation. This feature has been previously confirmed in some members of Chloroflexota [26, 33].

The ability to ferment is widely distributed in Chloroflexota. Most $MAGs_{annot}$ encoded genes involved in producing at least

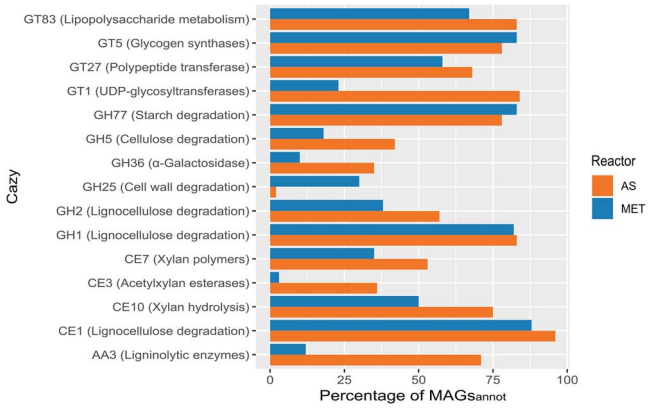

Figure 4. Analysis of the CAZy enzymes in Chloroflexota MAGs from AS and MET reactors. CAZy families found in more than 50% of the $MAGs_{annot}$ (annotated using the dbCAN2 carbohydrate-active enzyme (CAZy) domain HMM database). Glycoside hydrolase (GH), Glycosyl transferase (GT), carbohydrate esterase (CE), auxiliary activities (AA). Percentages of MAGs containing a Cazy family were calculated relative to the total number of MAGs in each reactor type (AS reactor vs MET reactors).

one fermentation product, such as ethanol (aldH, 40% $MAGs_{AS}$, 35% $MAGs_{MET}$), lactate (ldh, 33% $MAGs_{AS}$, 24% $MAGs_{MET}$), acetoin (ivlHI, 74% $MAGs_{AS}$, 47% $MAGs_{MET}$), acetate (yfiQ, 52% $MAGs_{AS}$, 6% $MAGs_{MET}$) and hydrogen (yfiQ, 89% $MAGs_{AS}$, 82% $MAGs_{MET}$) (Fig. 6, Fig. S4, Supplementary Data 3). This was in accordance with previous reports showing that, in a medium supplemented with glucose and yeast extract, an isolate classified within the Anaerolineae class produced acetate, hydrogen, lactate, succinate, formate, propionate and/or ethanol as the main end products of fermentation [30–32, 94]. Hence, these features indicate that most of these Chloroflexota could perform fermentation.

Chloroflexota from activated sludge systems have a facultative anaerobic metabolism

The majority of Chloroflexota members from activated sludge reactors exhibited a complete oxidative phosphorylation chain, including NADH:quinone oxidoreductase, succinate dehydrogenase, cytochrome c oxidase and/or F-type ATPase (94% $MAGs_{AS}$). In contrast, only 2% of $MAGs_{MET}$ showed this feature (Figs 5 and 6). These findings align with Anaerolineae members isolated from methanogenic reactors, which are typically described as obligate anaerobes [30–32, 90, 94]. However, it's worth noting that some Anaerolineae genomes have been found to contain genes for aerobic respiration [8, 9, 21, 95, 96]. It has been hypothesized that these complexes for aerobic respiration were likely acquired via horizontal gene transfer in some Chloroflexota members [97].

Considering that most of the Chloroflexota genomes detected in activated sludge reactors do not have isolated representatives in pure culture, the discovery of the first species capable of utilizing O_2 within the Anaerolineae class only occurred after the in situ characterization of Ca. Villigracilis [9]. A recent study using in situ characterization indicated that members of Chloroflexota retrieved from activated sludge reactors can utilize oxygen, N_2O and NO_2^- [20]. Our findings suggest that most of the Chloroflexota members in activated sludge reactors are capable of using O_2 as the final electron acceptor. Anoxic zones are likely common within large and denser flocs in activated sludge reactors due to limitations in oxygen diffusion [98]. The ability to ferment, found in most of the $MAGs_{AS}$, allows Chloroflexota to survive in these anoxic microniches. Moreover, enzymes related to protection against oxygen and/or reactive oxygen species, such as superoxide reductase/desulfoferrodoxin, superoxide dismutase, and catalase, were widely distributed in Chloroflexota phylum (present in 85% of AS $MAGs_{annot}$ and 95% of MET $MAGs_{annot}$) (Fig. 6, Supplementary Data 3). These findings suggest that they are well prepared for defense against reactive oxygen species.

The majority of $MAGs_{AS}$ have the potential to carry out the reduction of at least one nitrogen species (92% $MAGs_{AS}$, 24% $MAGs_{MET}$). This includes nitrate, with potential dissimilatory nitrate reduction to nitrite (narGHI, 16% $MAGs_{AS}$, 3% $MAGs_{MET}$), and the potential for dissimilatory nitrite reduction to ammonia (nrfAH, 46% $MAGs_{AS}$, 15% $MAGs_{MET}$). DNRA was present in 10% $MAGs_{AS}$ and 2% $MAGs_{MET}$. Genes for nitrite reduction to nitric oxide (nirK) were present in 46% $MAGs_{AS}$ and 10% $MAGs_{MET}$. Interestingly, 51% of the $MAGs_{AS}$ (3% $MAGs_{MET}$) contains a periplasmic nitrous oxide reductase (NosZ), indicating that nitrous oxide may also serve as a terminal electron acceptor. Therefore, Chloroflexota members could play a significant role in nitrite reduction and/or partial denitrification in activated sludge reactors (nirK, nosZ), suggesting a potential role in nitrogen removal from wastewater. The presence and use of genes related to the reduction of nitrogen species was reported by several studies [7, 20, 26, 33].

The ability to take up substrates under anoxic conditions as well as in presence of nitrate/nitrite was confirmed in members of Chloroflexota from activated sludge reactors, suggesting their character as facultative anaerobic chemoorganotrophs [9].

Why are members of Chloroflexota so successful in WWTPs?

The most widely accepted hypothesis is that members of the Chloroflexota phylum are notoriously difficult to cultivate due to slow growth, particularly those belonging to the Anaerolineae class [23, 24]. Consequently, they are easily outcompeted by fast-growing heterotrophic anaerobes. The slow growth of Chloroflexota members has also been demonstrated by the replication index of Chloroflexota MAGs across multiple samples from anaerobic digesters [99]. In this context, Chloroflexota was found to be within the 90% of the total MAGs with a dereplication index between 1.1 and 2, indicating slow growth. Only 10% of the total community had values between 2 and approximately 4, which can be considered "fast growing". On the other hand, isolated Anaerolineae members from methanogenic reactors require associations with other microbes (e.g. Archaea *Methanosaeta* spp.) for efficient growth [23].

Despite extensive efforts, only 14 species within Anaerolineae class have been successfully isolated from various environments, including anaerobic digesters, rice paddy soils, terrestrial aquifers, hydrothermal vents, and subseafloor sediment [30–32, 90, 94, 100–102]. The difficulty in isolating members of the Chloroflexota phylum in pure culture could also be due to the absence of key genes involved in B-vitamin biosynthesis are missing, as indicated by our results in most of the Chloroflexota MAGs (Supplementary Data 3). For instance, genes for thiamin (vitamin B1) biosynthesis (thiamine-phosphate synthase and thiamine- monophosphate kinase), biotin (vitamin B7) biosynthesis (adenosylmethionine-8-amino-7-oxononanoate ami-notransferase and biotin synthase) and adenosylcobalamin (vitamin B12) biosynthesis (cobalamin synthase and adenosylcobinamide-phosphate synthase) were absent as previously reported [21, 33]. This suggests that other microorganisms may support B-vitamin requirements for Chloroflexota community. Thus, the high abundances of Chloroflexota members

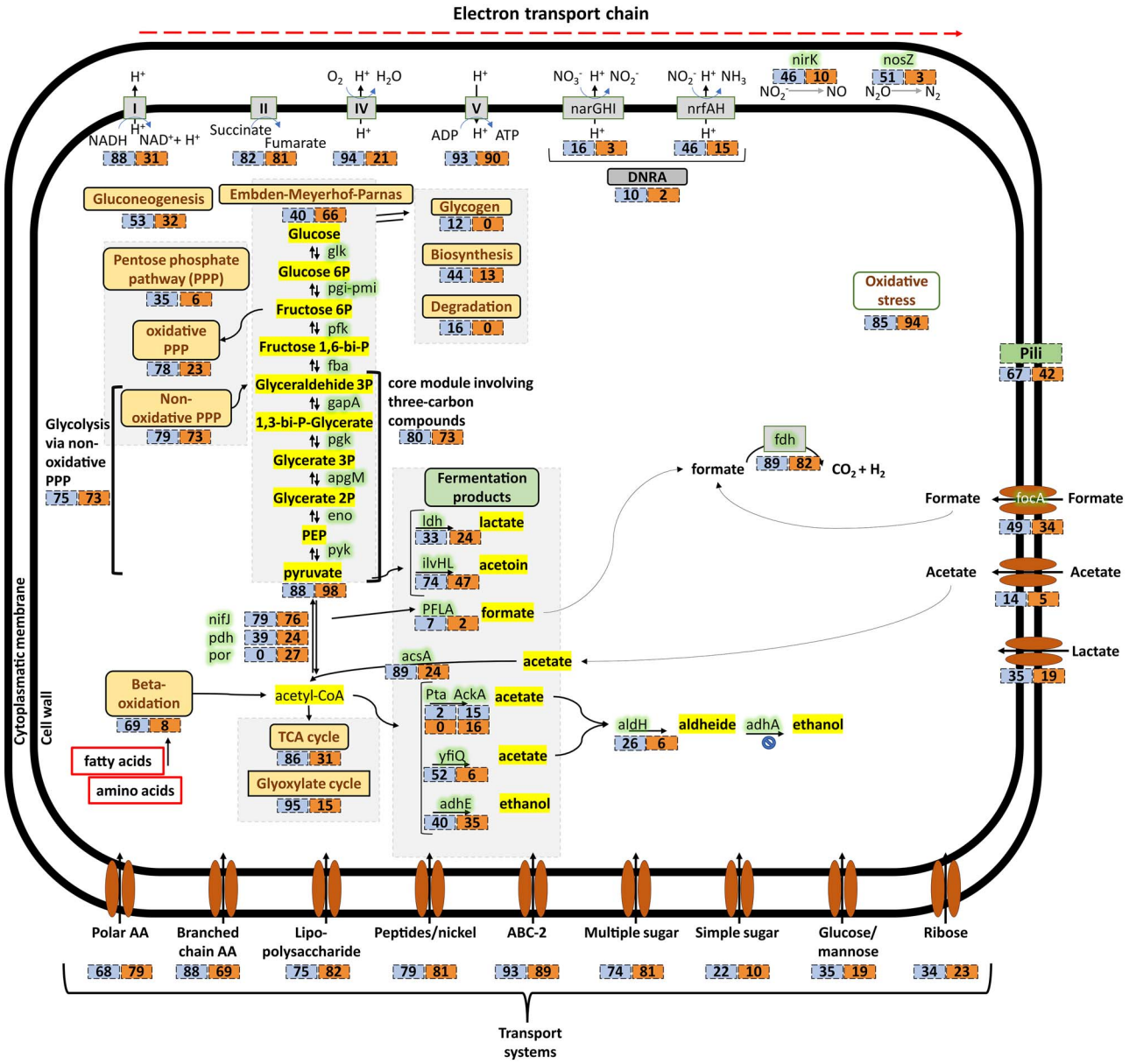

Figure 5. Metabolic model of the Chloroflexota $MAGs_{annot}$ in activated sludge and methanogenic reactors. The percentage of $MAGs_{annot}$ from AS and MET reactors appears in the boxes when the gene or metabolic pathway is present.

may suggest that their ecophysiology provides them with some competitive advantage over other bacterial populations, but they likely require vitamins and cofactors supplied by other microorganisms.

In our study, we demonstrate that all MAGs have a potential versatile metabolism related to the hydrolysis, fermentation and respiration of complex and simple organic compounds. This versatility enables these microorganisms to survive in the most diverse microbiomes, demonstrating their adaptability to different or changing conditions. On the other hand, it has been extensively proposed that Chloroflexota plays an important role in granule and floc formation. Evidence for this is their growth as filaments and the fact that some members showed cellular adhesiveness [89]. A complete set of genes for the pilus assembly (pilA, CpaB, CpaE, CpaF, TadB, TadC) which favor adhesiveness, was annotated for 67% $MAGs_{AS}$, 42% $MAGs_{MET}$ (Figs 5 and 6). As has been already noted, pili are often involved in facilitating adhesion and colonization in a wide variety of scenarios. Thus, these characteristics could represent a selective advantage for Chloroflexota evidenced by their high abundance in WWTPs.

Conclusions

In this study, we successfully addressed the initial questions through a comprehensive analysis of 264 genomes recovered from full-scale reactors.

Is the taxonomic composition of Chloroflexota the same in activated sludge and methanogenic reactors? Our findings suggest that the Chloroflexota taxonomic composition differs between activated sludge and methanogenic reactors. The Anaerolineae class was predominant in both systems, but with specific families in each.

Does the metabolic potential of Chloroflexota differ between both systems? Which carbon compound degradation pathways do they have? Genomes from both reactor types exhibit the potential to degrade complex organic matter and ferment a wide range of

Figure 6. Hierarchical clustering (Euclidean distance) of the presence/absence of metabolic pathways in the $MAGs_{annot}$. In the right part of the figure taxonomic assignment is shown at the order level.

substrates. Our results suggest that Chloroflexota species from MET reactors are strict fermenters, while species from AS reactors also possess genes for both aerobic and anaerobic respiration, potentially playing a crucial role in nitrogen removal.

Our study provides a robust analysis and contributes valuable insights into the diversity and metabolic potential of the Chloroflexota phylum within WWTPs. This compilation of genomes serves as a valuable resource for generating hypotheses in future studies and as a starting point for targeted cultivation of previously uncultivated Chloroflexota members.

However, it is important to note that due to missing metadata and limited statistical power, establishing significant associations between diversity or genes of Chloroflexota and reactor performance remains challenging. Therefore, well-designed studies that incorporate experimental approaches are necessary for a comprehensive understanding of Chloroflexota's impact on reactor functionality. Future research utilizing metagenomic and metatranscriptomic approaches will further validate genomic predictions, advancing our knowledge of Chloroflexota physiology and its influence on reactor responses.

Acknowledgements

The authors gratefully acknowledge support of the Clemente Estable Biological Research Institute (IIBCE), the Program of Basic Science Development of Uruguay (PEDECIBA-Chemistry).

Supplementary material

Supplementary material is available at *ISME Communications* online.

Funding

The Uruguayan National System of Researchers and National Agency for Research and Innovation (SNI-ANII), Projects: FCE_3_2020_1_162450 and FCE_1_2021_1_166934.

Conflicts of interest

The authors declare no conflict of interest.

Data availability

The raw metagenome sequences and MAG_{sdRep} generated in the present work have been deposited in the NCBI database with BioProject accession number PRJNA1037517. Scripts used in this study are available at Figshare (https://doi.org/10.6084/m9.figshare.24480880.v). Metagenomes and MAGs from published papers can be downloaded using the accession number associated with each respective work.

References

1. Sancho I, Lopez-Palau S, Arespacochaga N *et al.* New concepts on carbon redirection in wastewater treatment plants: a review. *Sci Total Environ* 2019;**647**:1373–84. https://doi.org/10.1016/j.scitotenv.2018.08.070
2. Verstraete W, Morgan-Sagastume F, Aiyuk S *et al.* Anaerobic digestion as a core technology in sustainable management of organic matter. *Water Sci Technol* 2005;**52**:59–66. https://doi.org/10.2166/wst.2005.0498 Available from. https://iwaponline.com/wst/article-pdf/52/1-2/59/433904/59.pdf
3. Zamri MFMA, Hasmady S, Akhiar A *et al.* A comprehensive review on anaerobic digestion of organic fraction of municipal

solid waste. *Renew Sust Energ Rev* 2021;**137**:110637. https://doi.org/10.1016/j.rser.2020.110637
4. Petriglieri F, Nierychlo M, Nielsen PH *et al.* In situ visualisation of the abundant Chloroflexi populations in full-scale anaerobic digesters and the fate of immigrating species. *PLoS One* 2018;**13**:1–14. https://doi.org/10.1371/journal.pone.0206255
5. Bovio-Winkler P, Cabezas A, Etchebehere C *et al.* Database mining to unravel the ecology of the phylum Chloroflexi in methanogenic full scale bioreactors. *Phylum Chloroflexi in Methanogenic Full Scale Bioreactors Front Microbiol* 2021;**11**:603234. https://doi.org/10.3389/fmicb.2020.603234
6. Speirs LBM, Dyson ZA, Tucci J *et al.* Eikelboom filamentous morphotypes 0675 and 0041 embrace members of the Chloroflexi: resolving their phylogeny, and design of fluorescence in situ hybridisation probes for their identification. *FEMS Microbiol Ecol* 2017;**93**:1–13. https://doi.org/10.1093/femsec/fix115
7. Andersen MH, McIlroy SJ, Nierychlo M *et al.* Genomic insights into Candidatus Amarolinea aalborgensis gen. Nov., sp. nov., associated with settleability problems in wastewater treatment plants. *Syst Appl Microbiol* 2019;**42**:77–84. https://doi.org/10.1016/j.syapm.2018.08.001
8. Kragelund C, Levantesi C, Borger A *et al.* Identity, abundance and ecophysiology of filamentous Chloroflexi species present in activated sludge treatment plants. *FEMS Microbiol Ecol* 2007;**59**:671–82. https://doi.org/10.1111/j.1574-6941.2006.00251.x
9. Nierychlo M, Miłobędzka A, Petriglieri F *et al.* The morphology and metabolic potential of the Chloroflexi in full-scale activated sludge wastewater treatment plants. *FEMS Microbiol Ecol* 2019;**95**:1–11. https://doi.org/10.1093/femsec/fiy228
10. Sekiguchi Y, Takahashi H, Kamagata Y *et al.* In situ detection, isolation, and physiological properties of a thin filamentous microorganism abundant in methanogenic granular Sludges: a novel isolate affiliated with a clone cluster, the green non-Sulfur bacteria, subdivision I. *Appl Environ Microbiol* 2001;**67**:5740–9. https://doi.org/10.1128/AEM.67.12.5740-5749.2001
11. Björnsson L, Hugenholtz P, Tyson GW *et al.* Filamentous Chloroflexi (green non-sulfur bacteria) are abundant in wastewater treatment processes with biological nutrient removal. *Microbiology (N Y)* 2002;**148**:2309–18
12. Li J, Hu B, Zheng P *et al.* Filamentous granular sludge bulking in a laboratory scale UASB reactor. *Bioresour Technol* 2008;**99**:3431–8. https://doi.org/10.1016/j.biortech.2007.08.005
13. Song YX, Liao Q, Yu C *et al.* Physicochemical and microbial properties of settled and floating anammox granules in upflow reactor. *Biochem Eng J* 2017;**123**:75–85. https://doi.org/10.1016/j.bej.2017.04.002
14. Nierychlo M, McIlroy SJ, Kucheryavskiy S *et al.* Candidatus Amarolinea and Candidatus Microthrix are mainly responsible for filamentous bulking in Danish municipal wastewater treatment plants. *Front Microbiol* 2020;**11**:1–17. https://doi.org/10.3389/fmicb.2020.01214
15. Van Loosdrecht MCM, Brdjanovic D. Anticipating the next century of wastewater treatment. *Science (1979)* 2014;**344**:1452–3. Available from. https://www.science.org/doi/10.1126/science.1255183
16. Mills S, Trego AC, Prevedello M *et al.* Unifying concepts in methanogenic, aerobic, and anammox sludge granulation. *Environmental Science and Ecotechnology* 2024;**17**:100310. https://doi.org/10.1016/j.ese.2023.100310
17. Trego AC, Mills S, Collins G. Granular biofilms: function, application, and new trends as model microbial communities. *Crit Rev Environ Sci Technol* 2021;**51**:1702–25. Available from. https://www.tandfonline.com/doi/abs/10.1080/10643389.2020.1769433
18. Klocke M, Mähnert P, Mundt K *et al.* Microbial community analysis of a biogas-producing completely stirred tank reactor fed continuously with fodder beet silage as mono-substrate. *Syst Appl Microbiol* 2007;**30**:139–51. https://doi.org/10.1016/j.syapm.2006.03.007 Available from. https://pubmed.ncbi.nlm.nih.gov/16697135/
19. Hofman-Bang J, Zheng D, Westermann P *et al.* Molecular ecology of anaerobic reactor systems. *Adv Biochem Eng Biotechnol* 2003;**81**:151–203. Available from. https://pubmed.ncbi.nlm.nih.gov/12747563/
20. Petriglieri F, Kondrotaite Z, Singleton C *et al.* A comprehensive overview of the Chloroflexota community in wastewater treatment plants worldwide. *mSystems* 2023;**8**:1–25. https://doi.org/10.1128/msystems.00667-23
21. Bovio-Winkler P, Guerrero LD, Erijman L *et al.* Genome-centric metagenomic insights into the role of Chloroflexi in anammox, activated sludge and methanogenic reactors. *BMC Microbiol* 2023;**23**:1–19. https://doi.org/10.1186/s12866-023-02765-5
22. Yamada T, Imachi H, Ohashi A *et al.* Bellilinea caldifistulae gen. Nov., sp. nov and Longilinea arvoryzae gen. Nov., sp. nov., strictly anaerobic, filamentous bacteria of the phylum Chloroflexi isolated from methanogenic propionate-degrading consortia. *Int J Syst Evol Microbiol* 2007;**57**:2299–306. https://doi.org/10.1099/ijs.0.65098-0
23. Yamada T, Sekiguchi Y. Cultivation of uncultured chloroflexi subphyla: significance and ecophysiology of formerly uncultured chloroflexi "subphylum i" with natural and biotechnological relevance. *Microbes and environments/JSME* 2009;**24**:205–16. https://doi.org/10.1264/jsme2.ME09151S
24. Nunoura T, Hirai M, Miyazaki M *et al.* Isolation and characterization of a thermophilic, obligately anaerobic and heterotrophic marine Chloroflexi bacterium from a Chloroflexi-dominated microbial community associated with a Japanese shallow hydrothermal system, and proposal for Thermomarinilin. *Microbes and environments/JSME [Internet]* 2013;**28**:228–35. https://doi.org/10.1264/jsme2.ME12193 Available from. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4070665&tool=pmcentrez&rendertype=abstract
25. Albertsen M, Hugenholtz P, Skarshewski A *et al.* Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* 2013;**31**:533–8. https://doi.org/10.1038/nbt.2579
26. McIlroy SJ, Karst SM, Nierychlo M *et al.* Genomic and in situ investigations of the novel uncultured Chloroflexi associated with 0092 morphotype filamentous bulking in activated sludge. *ISME J* 2016;**10**:2223–34. https://doi.org/10.1038/ismej.2016.14
27. Nielsen PH, Kragelund C, Seviour RJ *et al.* Identity and ecophysiology of filamentous bacteria in activated sludge. *FEMS Microbiol Rev* 2009;**33**:969–98. https://doi.org/10.1111/j.1574-6976.2009.00186.x
28. McIlroy SJ, Kirkegaard RH, Dueholm MS *et al.* Culture-independent analyses reveal novel anaerolineaceae as abundant primary fermenters in anaerobic digesters treating waste activated sludge. *Front Microbiol* 2017;8
29. Nierychlo, M., Andersen KS, Xu Y, *et al.* Species-level microbiome composition of activated sludge - introducing the MiDAS 3 ecosystem-specific reference database and taxonomy. *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*. 2019;**1**:2019. https://doi.org/10.1016/j.watres.2020.115955

30. Sekiguchi Y, Yamada T, Hanada S *et al.* Anaerolinea thermophila gen. Nov., sp. nov. and Caldilinea aerophila gen. Nov., sp. nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain bacteria at the subphylum level. *Int J Syst Evol Microbiol* 2003;**53**:1843–51. https://doi.org/10.1099/ijs.0.02699-0
31. Yamada T, Sekiguchi Y, Hanada S *et al.* Anaerolinea thermolimosa sp. nov., Levilinea saccharolytica gen. Nov., sp. nov. and Leptolinea tardivitalis gen. Nov., sp. nov., novel filamentous anaerobes, and description of the new classes Anaerolineae classis nov. and Caldilineae classis nov. in the bacterial phylum Chloroflexi. *Int J Syst Evol Microbiol* 2006;**56**:1331–40. https://doi.org/10.1099/ijs.0.64169-0
32. Sun L, Toyonaga M, Ohashi A *et al.* Isolation and characterization of Flexilinea flocculi gen. Nov., sp. nov., a filamentous, anaerobic bacterium belonging to the class Anaerolineae in the phylum Chloroflexi. *Int J Syst Evol Microbiol* 2016;**66**:988–96. https://doi.org/10.1099/ijsem.0.000822
33. Lawson CE, Wu S, Bhattacharjee AS *et al.* Metabolic network analysis reveals microbial community interactions in anammox granules. *Nat Commun* 2017;**8**:1–12. https://doi.org/10.1038/ncomms15416
34. Zhao Y, Liu SS, Jiang B *et al.* Genome-Centered metagenomics analysis reveals the symbiotic organisms possessing ability to cross-feed with Anammox bacteria in Anammox consortia. *Environ Sci Technol* 2018;**52**:11285–96. https://doi.org/10.1021/acs.est.8b02599
35. Bovio P, Cabezas A, Etchebehere C. Preliminary analysis of Chloroflexi populations in full-scale UASB methanogenic reactors. *J Appl Microbiol* 2019;**126**:667–83. https://doi.org/10.1111/jam.14115
36. Hao L, Michaelsen TY, Singleton CM *et al.* Novel syntrophic bacteria in full-scale anaerobic digesters revealed by genome-centric metatranscriptomics. *ISME J* 2020;**14**:906–18. https://doi.org/10.1038/s41396-019-0571-0
37. Li C, He P, Hao L *et al.* Diverse acetate-oxidizing syntrophs contributing to biogas production from food waste in full-scale anaerobic digesters in China. *Renew Energy* 2022;**193**:240–50. https://doi.org/10.1016/j.renene.2022.04.143
38. Liang J, Mai W, Wang J *et al.* Performance and microbial communities of a novel integrated industrial-scale pulp and paper wastewater treatment plant. *J Clean Prod* 2021;**278**:123896. https://doi.org/10.1016/j.jclepro.2020.123896
39. Puig-Castellví F, Midoux C, Guenne A *et al.* Metataxonomics, metagenomics and metabolomics analysis of the influence of temperature modification in full-scale anaerobic digesters. *Bioresour Technol* 2022;**346**:126612. https://doi.org/10.1016/j.biortech.2021.126612
40. Soo RM, Skennerton CT, Sekiguchi Y *et al.* An expanded genomic representation of the phylum cyanobacteria. *Genome Biol Evol* 2014;**6**:1031–45. https://doi.org/10.1093/gbe/evu073
41. Yang Y, Daims H, Liu Y, *et al.* Activity and metabolic versatility of complete ammonia oxidizers in full-scale wastewater treatment systems downloaded from [internet]. 2020. Available from: http://mbio.asm.org/
42. Fan L, Peng W, Duan H *et al. Presence and Role of Viruses in Anaerobic Digestion of Food Waste under Environmental Variability*, 2022.
43. Andrews. FastQC: a quality control tool for high throughput sequence data. *Babraham Institute [Internet]* 2010 [cited 2020 Aug 25]; Available from:. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
44. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114–20. Available from. https://doi.org/10.1093/bioinformatics/btu170. http://www.ncbi.nlm.nih.gov/pubmed/24695404
45. Li D, Luo R, Liu CM *et al.* MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 2016;**102**:3–11. https://doi.org/10.1016/j.ymeth.2016.02.020
46. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 2018;**6**:158Available from. https://doi.org/https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-018-0541-1
47. Chaumeil PA, Mussig AJ, Hugenholtz P *et al.* GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 2020;**36**:1925–7. Available from. https://doi.org/10.1093/bioinformatics/btz848. https://academic.oup.com/bioinformatics/article/36/6/1925/5626182
48. Parks DH, Chuvochina M, Chaumeil PA *et al.* A complete domain-to-species taxonomy for bacteria and archaea. *Nat Biotechnol* 2020;**38**:1079–86. https://doi.org/10.1038/s41587-020-0501-8
49. Gurevich A, Saveliev V, Vyahhi N *et al.* QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;**29**:1072–5. Available from. https://doi.org/10.1093/bioinformatics/btt086. https://pubmed.ncbi.nlm.nih.gov/23422339/
50. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. bioRxiv. 2022;2022.07.11.499243. Available from: https://www.biorxiv.org/content/10.1101/2022.07.11.499243v1
51. Orakov A, Fullam A, Coelho LP *et al.* GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome Biol* 2021;**22**:178–19. Available from. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02393-0
52. Haryono MA, Law YY, Arumugam K *et al.* Recovery of high quality metagenome-assembled genomes from full-scale activated sludge microbial communities in a tropical climate using longitudinal metagenome sampling. *Front Microbiol* 2022;**13**:869135.
53. Pérez MV, Guerrero LD, Orellana E *et al.* Time series genome-centric analysis unveils bacterial response to operational disturbance in activated sludge. *MSystems* 2019;**4**:10–1128.
54. Schneider D, Zühlke D, Poehlein A *et al.* Metagenome-assembled genome sequences from different wastewater treatment stages in Germany. *Microbiology resource announcements* 2021;**10**:10–1128.
55. Singleton CM, Petriglieri F, Kristensen JM *et al.* Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. *Nat Commun* 2021;**12**:2009.
56. Wang Y, Qin W, Jiang X *et al.* Seasonal prevalence of ammonia-oxidizing archaea in a full-scale municipal wastewater treatment plant treating saline wastewater revealed by a 6-year time-series analysis. *Environ Sci Technol* 2021;**55**:2662–73.
57. Ye L, Mei R, Liu WT *et al.* Machine learning-aided analyses of thousands of draft genomes reveal specific features of activated sludge processes. *Microbiome* 2020;**8**:16.
58. Brandt C, Bongcam-Rudloff E, Müller B. Abundance tracking by long-read nanopore sequencing of complex microbial communities in samples from 20 different biogas/wastewater plants. *Appl Sci* 2020;**10**:7518.
59. Campanaro S, Treu L, Rodriguez-R LM *et al.* New insights from the biogas microbiome by comprehensive genome-resolved

metagenomics of nearly 1600 species originating from multiple anaerobic digesters. *Biotechnology for biofuels* 2020;**13**:1–18.
60. Hao L, Michaelsen TY, Singleton CM *et al.* Novel syntrophic bacteria in full-scale anaerobic digesters revealed by genome-centric metatranscriptomics. *The ISME journal* 2020;**14**:906–18.
61. Ma S, Jiang F, Huang Y *et al.* A microbial gene catalog of anaerobic digestion from full-scale biogas plants. *Gigascience* 2021;**10**:giaa164.
62. Raguideau S, Trego A, Farrell F *et al.* Novel microbial syntrophies identified by longitudinal metagenomics. *BioRxiv* 2021-07.
63. Park SJ, Andrei AŞ, Bulzu PA *et al.* Expanded diversity and metabolic versatility of marine nitrite-oxidizing bacteria revealed by cultivation- and genomics-based approaches. *Appl Environ Microbiol* 2020;**86**:e01667–20.
64. Eisenhofer R, Odriozola I, Alberdi A. Impact of microbial genome completeness on metagenomic functional inference. *ISME Communications* 2023;**3**:1–5. Available from. https://doi.org/10.1038/s43705-023-00221-z. https://www.nature.com/articles/s43705-023-00221-z
65. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal* 2017;**11**:2864–8. Available from: https://www.nature.com/articles/ismej2017126, https://doi.org/10.1038/ismej.2017.126
66. Letunic I, Bork P. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;**47**:W256–9. https://doi.org/10.1093/nar/gkz239
67. Kanehisa M, Goto S. *KEGG: Kyoto Encyclopedia of Genes and Genomes [Internet]*. Vol. **28**, *Nucleic Acids Res*. Oxford University Press; 2000 [cited 2021 Jan 17]. p. 27–30. Available from: /pmc/articles/PMC102409/?report=abstract
68. Lombard V, Golaconda Ramulu H, Drula E *et al.* The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 2014;**42**. Available from: D490–5. https://doi.org/10.1093/nar/gkt1178. https://pubmed.ncbi.nlm.nih.gov/24270786/
69. R Core Team. R Core Team [Internet]. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 2020. Available from: https://www.eea.europa.eu/data-and-maps/indicators/oxygen-consuming-substances-in-rivers/r-development-core-team-2006
70. Oksanen J, Blanchet FG, Kindt R, *et al.* Vegan: community ecology package. R Package Version. 2.0-10 [Internet]. 2013 [cited 2020 Aug 6]. Available from: https://www.researchgate.net/publication/258996451_Vegan_Community_Ecology_Package_R_Package_Version_20-10
71. Andersen KS, Kirkegaard RH, Karst SM *et al.* ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. *bioRxiv* 2018;299537
72. Wickham H. *ggplot2 - Elegant Graphics for Data Analysis | Hadley Wickham* | Springer [Internet]. 2017 [cited 2020 Jun 15]. 213 p. Available from: https://www.springer.com/gp/book/9780387981413
73. Kolde R. Pheatmap: pretty heatmaps version 1.0.12. 2019 [cited 2021 Jan 17]; Available from: https://rdrr.io/cran/pheatmap/
74. Hammer Ø. *PAST PAleontological STatistics Reference Manual*, 1999
75. Rodríguez-Gijón A, Nuy JK, Mehrshad M *et al.* A genomic perspective across Earth's microbiomes reveals that genome size in archaea and bacteria is linked to ecosystem type and trophic strategyRunning title: Archaea and Bacteria genome size distribution. https://doi.org/10.1101/2021.01.18.427069
76. Nielsen DA, Fierer N, Geoghegan JL *et al.* Aerobic bacteria and archaea tend to have larger and more versatile genomes. *Oikos* 2021;**130**:501–11. https://doi.org/10.1111/oik.07912
77. Blesa A, Averhoff B, Berenguer J. Horizontal gene transfer in Thermus spp. *Curr Issues Mol Biol* 2018;**29**:23–36. Available from: https://pubmed.ncbi.nlm.nih.gov/29648539/, https://doi.org/10.21775/cimb.029.023
78. Borges KM, Bergquist PL. Genomic restriction map of the extremely thermophilic bacterium *Thermus thermophilus* HB8. *J Bacteriol* 1993;**175**:103–10. Available from: https://pubmed.ncbi.nlm.nih.gov/8416889/, https://doi.org/10.1128/jb.175.1.103-110.1993
79. Chen MY, Teng WK, Zhao L *et al.* Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. *The ISME Journal* 2020;**15**:211–27
80. Cobo-Simón M, Tamames J. Relating genomic characteristics to environmental preferences and ubiquity in different microbial taxa. *BMC Genomics* 2017;**18**:1–11. https://doi.org/10.1186/s12864-017-3888-y
81. Konstantinidis KT, Tiedje JM. Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc Natl Acad Sci* 2004;**101**:3160–5. Available from: https://www.pnas.org/content/101/9/3160, https://doi.org/10.1073/pnas.0308653100
82. Bentkowski P, Van Oosterhout C, Mock T. A model of genome size evolution for prokaryotes in stable and fluctuating environments. *Genome Biol Evol* 2015;**7**. Available from: /pmc/articles/PMC4558865/:2344–51. https://doi.org/10.1093/gbe/evv148
83. Daramola MO, Aransiola EF, Adeogun AG. Comparative study of thermophilic and mesophilic anaerobic treatment of purified terephthalic acid (PTA) wastewater. *Nat Sci (Irvine)* 2011;**03**:371–8. https://doi.org/10.4236/ns.2011.35050
84. Speirs LBM, Rice DTF, Petrovski S *et al.* The phylogeny, biodiversity, and ecology of the Chloroflexi in activated sludge. *Front Microbiol* 2019;**10**. https://doi.org/10.3389/fmicb.2019.02015
85. Bowers RM, Kyrpides NC, Stepanauskas R, *et al. Minimum Information about a Single Amplified Genome (MISAG) and a Metagenome-Assembled Genome (MIMAG) of Bacteria and Archaea [Internet]*. Vol. **35**, Nature Biotechnology. Nature Publishing Group; 2017. p. 725–31. Available from: http://gensc.org
86. Wilkens C, Busk PK, Pilgaard B *et al.* Diversity of microbial carbohydrate-active enzymes in Danish anaerobic digesters fed with wastewater treatment sludge. *Biotechnol Biofuels* 2017;**10**:1–14. https://doi.org/10.1186/s13068-017-0840-y
87. Ruiken CJ, Breuer G, Klaversma E *et al.* Sieving wastewater – cellulose recovery, economic and energy evaluation. *Water Res* 2013;**47**:43–8. https://doi.org/10.1016/j.watres.2012.08.023
88. Zheng Y, Maruoka M, Nanatani K *et al.* High cellulolytic potential of the Ktedonobacteria lineage revealed by genome-wide analysis of CAZymes. *J Biosci Bioeng* 2021;**131**:622–30. https://doi.org/10.1016/j.jbiosc.2021.01.008
89. Xia Y, Wang YY, Wang YY *et al.* Cellular adhesiveness and cellulolytic capacity in Anaerolineae revealed by omics-based genome interpretation. *Biotechnol Biofuels.* 2016;**9**:1–13. https://doi.org/10.1186/s13068-016-0524-z
90. Podosokorskaya OA, Bonch-Osmolovskaya EA, Novikov AA *et al.* Ornatilinea apprima gen. Nov., sp. nov., a cellulolytic representative of the class Anaerolineae. *Int J Syst Evol Microbiol* 2013;**63**:86–92. https://doi.org/10.1099/ijs.0.041012-0
91. Xia Y, Kong Y, Nielsen PH. In situ detection of protein-hydrolysing microorganisms in activated sludge. *FEMS*

Microbiol Ecol 2007;**60**:156–65. Available from. https://doi.org/10.1111/j.1574-6941.2007.00279.x. https://pubmed.ncbi.nlm.nih.gov/17313663/

92. Anantharaman K, Brown CT, Burstein D *et al.* Analysis of five complete genome sequences for members of the class Peribacteria in the recently recognized Peregrinibacteria bacterial phylum. *PeerJ* 2016;**4**. Available from::e1607. https://doi.org/10.7717/peerj.1607. https://pubmed.ncbi.nlm.nih.gov/26844018/
93. Onetto CA, Grbin PR, McIlroy SJ *et al.* Genomic insights into the metabolism of "Candidatus Defluviicoccus seviourii", a member of Defluviicoccus cluster III abundant in industrial activated sludge. *FEMS Microbiol Ecol* 2019;**95**:1–12. https://doi.org/10.1093/femsec/fiy231
94. Nakahara N, Nobu MK, Takaki Y *et al.* Aggregatilinea lenta gen. Nov., sp. nov., a slow-growing, facultatively anaerobic bacterium isolated from subseafloor sediment, and proposal of the new order aggregatilineales Ord. Nov. within the class anaerolineae of the phylum chloroflexi. *Int J Syst Evol Microbiol* 2019;**69**:1185–94. https://doi.org/10.1099/ijsem.0.003291
95. Hemp J, Ward LM, Pace LA *et al.* Draft genome sequence of Ornatilinea apprima P3M-1, an anaerobic member of the Chloroflexi class Anaerolineae. *Genome Announc* 2015;**3**:1353–68. Available from. https://journals.asm.org/doi/10.1128/genomeA.01353-15
96. Ward LM, Hemp J, Pace LA *et al.* Draft genome sequence of Leptolinea tardivitalis YMTK-2, a mesophilic anaerobe from the Chloroflexi class Anaerolineae. *Genome Announc* 2015;**3**. Available from:. https://doi.org/10.1128/genomeA.01356-15. https://pubmed.ncbi.nlm.nih.gov/26586893/
97. Ward LM, Hemp J, Shih PM *et al.* Evolution of phototrophy in the Chloroflexi phylum driven by horizontal gene transfer. *Front Microbiol* 2018;**9**:1–16. https://doi.org/10.3389/fmicb.2018.00260
98. Schramm A, Santegoeds CM, Nielsen HK, *et al.* On the occurrence of anoxic microniches, denitrification, and sulfate reduction in aerated activated sludge. *Appl Environ Microbiol* 1999;**65**:4189–96. Available from: https://journals.asm.org/journal/aem, https://doi.org/10.1128/AEM.65.9.4189-4196.1999
99. Campanaro S, Treu L, Rodriguez-R LM *et al.* New insights from the biogas microbiome by comprehensive genome-resolved metagenomics of nearly 1600 species originating from multiple anaerobic digesters. *Biotechnol Biofuels* 2020;**13**:1–18. https://doi.org/10.1186/s13068-020-01679-y
100. Yoon DN, Park SJ, Kim SJ *et al.* Isolation, characterization, and abundance of filamentous members of Caldilineae in activated sludge. *J Microbiol* 2010;**48**:275–83. https://doi.org/10.1007/s12275-010-9366-8
101. Grégoire P, Bohli M, Cayol JL *et al.* Caldilinea tarbellica sp. nov., a filamentous, thermophilic, anaerobic bacterium isolated from a deep hot aquifer in the Aquitaine Basin. *Int J Syst Evol Microbiol* 2011;**61**:1436–41. https://doi.org/10.1099/ijs.0.025676-0
102. Kale V, Björnsdóttir SH, Fridjónsson ÓH *et al.* Litorilinea aerophila gen. Nov., sp. nov., an aerobic member of the class Caldilineae, phylum Chloroflexi, isolated from an intertidal hot spring. *Int J Syst Evol Microbiol* 2013;**63**:1149–54. https://doi.org/10.1099/ijs.0.044115-0