Bioresource Technology 441 (2026) 133657

Contents lists available at ScienceDirect

Bioresource Technology

journal homepage: www.elsevier.com/locate/biortech

Who calls granules "home"? Domain-spanning *meta*-analysis charting microbial ecosystems underlying aerobic granular sludge reactors

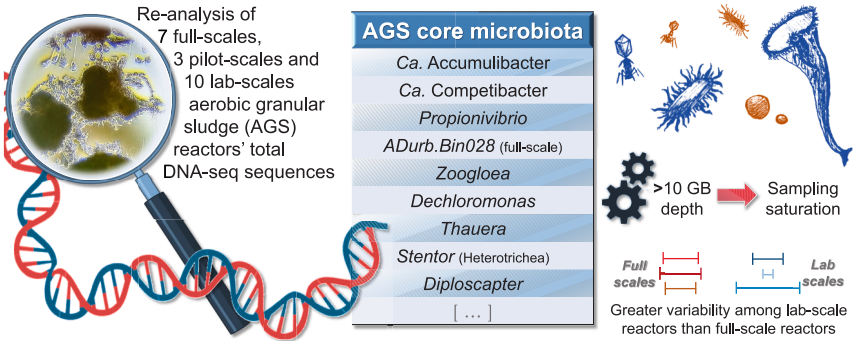
Leandro Di Gloria [a], Tommaso Lotti [b], Mark CM van Loosdrecht [c], Matteo Ramazzotti [a,*]

[a] Department of Experimental and Clinical Biomedical Sciences, University of Florence, Florence, Italy
[b] Department of Civil and Environmental Engineering, University of Florence, Florence, Italy
[c] Department of Biotechnology, Delft University of Technology, Delft, the Netherlands

HIGHLIGHTS

- A sequencing depth of 10 GB allows profiling the majority of the AGS community;
- Revealed the AGS core microbiota irrespective of reactors' configuration and scale;
- ADurb.Bin028 is a recurrent genus in reactors fed with real wastewater;
- *Rotaria*, *Diploscapter* and *Stentor* rank among the most abundant eukaryotes in AGS;
- The reactors fed with real wastewater exhibited similar taxonomic profiles.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Microbiota
Ecology
Microbial ecology
Aerobic granular sludge
Meta-analysis
Wastewater treatment

ABSTRACT

The theatre of activity of complex microbial communities underpins the Aerobic Granular Sludge (AGS) systems, resulting in efficient wastewater treatment. Here, we present the first meta-analysis of DNA sequencing data from both published and newly generated AGS samples, aiming to define the "core microbiota" of AGS reactors, consisting of bacteria, archaea, eukaryotes and DNA viruses consistently featured and shared across different scales and operational settings. Briefly, the results indicated that a sequencing depth of at least 10 GB is required to profile the majority of the AGS community, revealed the core taxa, detected the recurrent presence of the uncultured genus ADurb.Bin028 in full-scale reactors and identified *Rotaria* and *Diploscapter*, as well as the sessile ciliates *Stentor* and *Thuricola*, as the most abundant eukaryotes in AGS. In conclusion, this work provided a taxonomic overview of AGS' common microbes and addressed potential technical caveats, aiming to establish a reference for future studies.

1. Introduction

The biological treatments that enables the mitigation of increasing environmental pollution can be viewed as the result of challenging established beliefs and explore new directions. In 1914, Ardern and Lockett questioned the prevailing practice of discarding the flocculent

* Corresponding author at: Giovanni Battista Morgagni 50, 50134 FLORENCE (FI), Italy.
E-mail address: matteo.ramazzotti@unifi.it (M. Ramazzotti).

https://doi.org/10.1016/j.biortech.2025.133657
Received 2 September 2025; Received in revised form 9 November 2025; Accepted 13 November 2025
Available online 15 November 2025

solids in wastewater treatment plants (WWTPs) bioreactors and identified them as the potential means for purifying the incoming wastewater. As a result, they referred to the resulting floccular system as "activated sludge" (AS) in the absence of a more precise term (Orhon, 2015). Nowadays, we recognized the so called AS flocs as active microbial aggregates composed primarily of bacteria, however their involvement was initially misunderstood or framed as ancillary (Orhon, 2015). Mishima and Nakamura further refined the concept underlying the flocs in 1990 by allowing denser microbial aggregates, namely the microbial granules, to be maintained while discarding the biomass with higher settling time (Wilén et al., 2018). This system was initially operated under continuous aerobic conditions and then defined as "Aerobic Granular Sludge" (AGS) (Wilén et al., 2018). In 1998, Heijnen and van Loosdrecht introduced a cycle of redox conditions (aerobic, anoxic and anaerobic phase), along with a feast-famine regime (with the feast phase occurring under anaerobic conditions), favouring organisms responsible for granule formation (Wilén et al., 2018). Briefly, most of the organic carbon in the mixed liquor is consumed during the anaerobic phase by phosphate-accumulating organisms (PAOs) and glycogen-accumulating organisms (GAOs). These organisms utilize glycogen and, in the case of PAOs, also polyphosphate compounds synthesized during the aerobic phase as reserves, thereby gaining an ecological advantage over faster-growing microbes when oxygen is absent (Pronk et al., 2023). Moreover, due to the distinct ecological niches thriving in the various redox conditions throughout the granule, the resulting biomass is capable of simultaneous nitrification, denitrification, carbon oxidation and phosphorus removal (Wilén et al., 2018). Accordingly, microbial taxa exchange metabolic products within the reactor mixed liquor. For instance, nitrifiers generate nitrate and nitrite that can serve as substrate for denitrifiers. Furthermore, many granule-associated microorganisms produce polymeric substances and poly-hydroxy-alkanoates (PHA) which can be recovered and valorised for industrial applications, promoting the circular economy. Finally, the transition from AGS lab-scale reactors to actual full-scale applications was first achieved in the Netherlands in 2010 (Pronk et al., 2023).

Full-scale AGS reactors typically exhibit longer start-up times and greater operational instability compared to the lab-scales (Wilén et al., 2018). These limitations can be mitigated through close monitoring of operational parameters and by introducing selected microbial consortia to support bio-augmentation or restore the microbial ecosystem that underpins the process (Han et al., 2022). In light of this, identifying the microbial actors ("microbiota") of AGS and their "theatre of activity" ("microbiome") (Berg et al., 2020) is essential for effective monitoring, operability and prediction of reactor treatment efficiency in response to certain stimuli (e.g. the nutrient available in the influent (Zhang et al., 2024)). Despite the substantial progress made in characterizing key microbial players within AGS systems since their ideation, it remains widely acknowledged that our understanding of the AGS microbiome is still incomplete and far from being fully harnessed (Wilén et al., 2018). For example, the GAOs and filamentous bacterial heterotrophs have long been considered undesirable due to their competitive interactions with PAOs, which, conversely, are crucial in AGS systems as they uptake the orthophosphate during the aerobic phase to replenish their intracellular polyphosphates, thereby acting as the main drivers of the phosphorous removal (Wang et al., 2021). Herein lies a recurring misconception regarding the nature of micro-organism communities: the term "microbiota" encompasses the ecological suffix "-biota" which indicates all of the living organisms in a certain area as a single unit (European Environment Agency (EEA)). This implies that the members of the community exert at least a minimal direct or indirect influence on one another. In fact, although simplifying complex topics is a human tendency for better understanding, ecological community are intricate networks wherein multispecies interaction occurs. For example, the quorum sensing is one of the most famous communication methods between bacteria, even across distinct clades, and it has been proven to be critical to granules formation (Berg et al., 2020; Wilén et al., 2018). Nowadays, GAOs and other minor microbial populations are recognized as essential actors that must be maintained in a balanced state within AGS reactors, due to their cooperative interactions with PAOs and their role in granule structural stability (Cardona et al., 2025; Wang et al., 2021). Accordingly, numerous studies have adopted molecular ecology approaches to investigate AGS communities as whole pictures, without overlooking the potential contributions of archaea, eukaryotes and viruses. Notably, predation is another important interaction in any biota, with protozoa and bacteriophages being large contributors in the bioreactors. The most prevalent protozoa in WWTPs belong to sessile ciliates which attach to the biomass and fed on suspended particulate matter and bacteria (Barrios-Hernández et al., 2021; Weber et al., 2007). As for the bacteriophages, they account for up to about 70 % of the bacterial mortality in aquatic ecosystems and they are by far the most abundant entities in both natural environments and WWTPs (Nieuwenhuijse et al., 2020; Wilén et al., 2018; Zhao et al., 2024). Ideally, the bacteriophages regulate the amount of filamentous bacteria and remove dangerous bacterial pathogens (Yang et al., 2017), acting according to the mechanism referred to as "kill the winner". Briefly, this mechanism involves a positive feedback loop in which increased prey abundance leads to intensified predation, and thus contribute to limiting bacteria abundances in the ecosystems (Barr et al., 2010). Conversely, certain phages can infect and reduce populations of important bacterial strains, potentially destabilizing the reactor (Barr et al., 2010). Nevertheless, the contribute of protozoa and bacteriophages should not be oversimplified to mere predation: protozoa can excrete compounds stimulating the bacterial growth or serve as physical substrate on which begin the granule development (Weber et al., 2007) while bacteriophages are the vector of gene transduction between bacteria and also provide auxiliary metabolic genes related to antibiotic resistance or carbon, nitrogen and phosphate metabolisms (Zhao et al., 2024).

To support research on the AGS microbiome, this article aims at better defining the core members of the AGS microbiota. To achieve this, we surveyed 18 publicly available AGS whole genome sequencing data from sequencing batch reactors (SBRs) and rectors operating under continuous flow system (CFRs). We further included data from two novel reactors to strengthen our findings, totalling 20 AGS reactors. A bioinformatic approach was used to identify and collect bacteria, archaea, eukaryotes and DNA viruses, which was previously validated through an AGS-focused benchmark (Di Gloria et al., 2025). Finally, we addressed key technical challenges in AGS microbiota analysis, providing essential insights for future research design.

2. Materials and methods

2.1. Sample collection

Short-read DNA sequencing data from AGS reactors publicly available and associated with peer-reviewed scientific publications, were downloaded from the NCBI SRA archive (Table 1).

All datasets providing raw data of adequate quality and pertaining to AGS reactors alternating anaerobic, aerobic and anoxic phases were included, with the aim of representing the variety of the AGS communities as far as permitted by the low number of currently available related datasets. Accordingly, a total of 14 datasets from SBRs, each previously described in the literature and characterized by distinct operational conditions, were retrieved. To further strengthen the analysis, data from two new reactors was included. In fact, total DNA was extracted from four samples of mixed liquor collected from a novel pilot-scale AGS SBR treating real municipal wastewater operated in Tuscany (Italy) and from one sample of a laboratory-scale reactor under similar operational settings but treating synthetic wastewater. DNA was purified using the QIAGEN PowerSoil DNA Extraction Kit and sequenced was sequenced in paired-end mode (2 x 150 cycles) on the Illumina NovaSeq X platform for a target output of 20 GB for sample. These newly sequenced samples were analysed along with the downloaded data and

Table 1
List of the data sets included in the *meta*-analysis. The column "Type" indicates whether the reactors in each data set are sequencing batch reactors (SBRs) or continuous flow reactors (CFR). The "Scales" column refers to the operational scale of the reactors (full, pilot or laboratory-scale). The "Seq depth (GB)" column reports the total sequencing depth, expressed in giga-bases (GB), while the "Read pairs (millions)" column contains the approximate number of million paired reads for sample in each dataset. The "Main characteristics" column summarizes key features of each dataset relevant to this *meta*-analysis. This summary is intended to provide an overview and is not a substitute for the detailed methodological descriptions available in the original publications corresponding to each dataset.

Data set identification number	Type	Paper DOI	SRA ID	Scale	Seq depth (GB)	Read pairs (millions)	Main characteristics
1	SBR	https://doi.org/10.1016/j.jhazmat.2022.129528	PRJEB50804	Full	<5	10	Domestic wastewater with micropollutants (antibiotics and other pharmaceuticals)
2	SBR	https://doi.org/10.1016/j.ese.2025.100560	PRJNA1054024	Full	10	40	Plant treating a load of about 1.7 million population equivalents
3	SBR	https://doi.org/10.1016/j.watres.2022.118571	PRJNA783874	Full	5 ~ 10	30	Samples sieved in 3 different fractions of biomass according to granules dimensions
4	SBR	https://doi.org/10.1016/j.watres.2023.120700	PRJNA792132	Full	15	50	Analysing three different plants in the Netherlands
5	SBR	*this paper*	PRJNA1311054	Pilot, Lab	20	80	See the related paragraph of this paper
6	SBR	https://doi.org/10.1128/mra.00102–24	PRJEB38840	Lab	<10	30	Testing diverse DNA extraction methods and carbon substrates mixtures as supplements (volatile fatty acids or complex mixtures of aminoacids and glucose)
7	SBR	https://doi.org/10.1016/j.cej.2023.145638	PRJNA985882	Lab	<10	20	Low P, supported the NO3– concentration during the anoxic phase
8	SBR	https://doi.org/10.1016/j.resenv.2024.100149	PRJNA890448	Lab	15	60	Samples sieved in 3 different fractions of biomass according to granules dimensions. The influent featured micropollutants, such as pharmaceuticals. Moreover, one of the two reactor was supplied with sodium acetate and acetic acid
9	SBR	https://doi.org/10.1016/j.jwpe.2023.103577	PRJNA902909	Lab	10	40	Synthetic influent with ratio COD:N:P equal to 100:5:1
10	SBR	https://doi.org/10.1101/2020.01.10.901413	PRJNA576469	Lab	15	50	Primary carbon source switched between acetate and propionate every two SBR cycles
11	SBR	https://doi.org/10.1111/1462–2920.13019	PRJNA231881	Lab	10	80	Synthetic wastewater containing volatile fatty acids and PO_4^{3-} for the enrichment of PAOs
12	CFR	https://doi.org/10.1016/j.jwpe.2023.104584	PRJNA1015110	Lab	10	30	Semi-continuous process optimized for phosphorus removal and denitrification, the inoculum was the biomass of project 7's reactor
13	CFR	https://doi.org/10.1016/j.bej.2020.107711	PRJNA1082061	Pilot, Lab	<10	20	CFR treating mainly containing organic pollutants in real industrial wastewater
14	CFR	https://doi.org/10.2139/ssrn.4823902	PRJNA606844	Pilot	<10	20	Continuous flow partial nitrification AGS treating landfill leachate

collectively labelled as "dataset 5″.

Cumulatively, 16 AGS SBRs, alternating anaerobic, aerobic and anoxic phases, were collected from the datasets: 7 full-scales, 1 pilot-scale and 8 laboratory-scale. Furthermore, 2 pilot-scale and 2 laboratory-scale CFRs, retrieved from datasets 12, 13 and 14, were also included. Among these, the pilot-scale samples from datasets 14 represented a partial nitrification-AGS community. Due to the higher number of samples from SBR systems, this operational condition served as the primary focus of the taxonomic analysis, while CFR were included as a secondary comparison. As more than one sample was available for most of the reactors, this analysis comprised 57 sample in total. All full-scale and pilot-scale reactors were fed with real wastewater influent, whose nutrient composition likely varies depending on the country of origin, while laboratory-scale reactors used synthetic wastewater. Variable sequencing depth, measured as "GB" (number of giga-bases sequenced) in Table 1, was observed across datasets. However, datasets 6 and 11 were characterized by shorter read lengths (100 bp) compared to the other datasets (150 bp), and thus featured a slightly higher reads number at the same GB.

Detailed settings for each reactor included in this meta-analysis can be found in the respective original publications (see Table 1 column 2: "Paper DOI"). Specific information for the two newly sequenced reactors is briefly reported in Sections 2.2. The full list of sample identifiers and associated metadata, including the information necessary to link the sample labels used in this meta-analysis to the sample identity in the original publications, is provided in Table S1.

2.2. Set-up of the reactors belonging to the "dataset 5"

The pilot-scale reactor had an operative volume of 2 m^3 and was localized in Florence (Tuscany). Briefly, this reactor was inoculated with 1.5 m^3 AS taken from a full-scale SBR treating municipal wastewater. From that moment on, the pilot plant was fed with real municipal wastewater from a combined sewer system of a peripheral area of the municipality of Florence. Due to the presence of low levels of carbon to nitrogen ratios in the raw influent (4.5 g COD/g N), 150 mg COD/L of acetic acid was added to the row influent to speed up the granulation process. The reactor cycle structure consists of six phases: anaerobic phases of i) feeding and ii) mixing; iii) an aerobic phase with external air supply; iv) a mixed anoxic phase; v) settling and consequent vi) withdrawal of the effluent. Granulation from floccular biomass was achieved through the application of metabolic and hydraulic selective pressures.

The laboratory-scale reactor had an operative volume of 5 L. Its cycle was structured in two macro-phases: i) an up-flow anaerobic feeding and an anaerobic mixing phase with recirculation of oxygen-free process-gas to favour the metabolic selection of slow-growing microorganisms such as PAOs and GAOs; ii) an aerobic phase with external air supply for dissolved oxygen control, aimed at promoting aerobic/anoxic phosphates uptake by PAOs. The reactor was fed with synthetic wastewater which composition in terms of organic matter, N and P was 400 mg COD/L (as acetic acid), 50 mg NH_4^+–N/L and 7 mg PO_4^{3-}–P/L, respectively.

2.3. Bioinformatic workflow

The reads were processed and analysed according to a recently published benchmark tailored on an AGS-representative community (Di Gloria et al., 2025).

Briefly, demultiplexed sequence reads were preprocessed using BBDuk (module of BBTools suite, version 39.06) to remove reads sourced from Illumina adapters or phiX, very-low complexity sequences and 3′ ends regions with Q-score lower than 20, while taking into account the in paired-end nature of the sequencing. Afterwards, potential human reads were removed using Kraken v 2.1.3 (Wood et al., 2019) with the human GRCh38 reference genome as database and a confidence threshold set to 0.99. Finally, Kaiju v 1.10 (Menzel et al., 2016) was used with the parameter −m 30 (minimal match length) to classify reads into taxonomic clades using a custom database. This database was based on Kaiju's standard complete database, with the addition of selected eukaryotic organisms potentially relevant to AGS ecosystems. Taxa observed only as singletons or doubletons across samples were discarded to mitigate the risk of false negatives (Amos et al., 2020; Di Gloria et al., 2025).

The ecological analyses on microbial communities were performed in R 4.3 with the help of the packages phyloseq v 1.44.0 (McMurdie and Holmes, 2013), vegan v 2.6–4, ecodist v 2.1.3 and other packages satisfying their dependencies. The packages ggplot2 3.3.6 and ggh4x 0.2.2 were used to plot data and results. A rarefaction analysis on genera was performed on every sample using the function rarecurve, further processed to highlight highly unsaturated samples (arbitrarily defined as samples with a final slope in the rarefaction curve with an increment in taxon number per reads $< 1e\text{-}3$). For the sake of simplicity, taxa lacking a defined genus name in the NCBI taxonomy (e.g., *Flavobacteriales bacterium*) were reported in genus-level tables and plots using the name of the lowest available clade. The most abundant taxa were identified according to their highest average relative abundance among the samples, excluding the unclassified reads from the computation. The average abundances of archaea, eukaryotes or viruses in the dedicated bar plots and tables where computed including only the taxon of the corresponding domain, due to their low relative abundance when the eubacteria were included in the dataset. PCoA (Principal Coordinate Analysis) was performed at species level using the Hellinger distance given the sparse and compositional properties of the data (Legendre and Legendre, 1998). The alpha diversity was computed after excluding rare species, defined as observation with average lower than 0.0001 % across samples (threshold established according to the rarefaction curve results), to mitigate the influence of the varying sequencing depths. The Pielou's evenness index was calculated using the formula $E = S/\log(R)$, where S is the Shannon diversity index and R is the observed species richness in the sample. The potential core microbiota was here defined as the group of taxa with maximal relative abundance above 0.0025 % across samples and identified in more than 75 % of the full-scale, pilot-scale and laboratory-scale SBR samples. Due to the low ecological saturation observed for viruses, core microbiota analysis was not performed for this clade. Further details are available in the Bash and R scripts provided in the GitHub repository indicated in the Data Availability section.

2.4. Light microscopy

Samples were observed using light microscopy (BW Optics, model MXD-400) to identify eukaryotic taxa. Aliquots of mixed liquor were placed on glass slides and the images were acquired at 10 × magnification (granules image in the graphical abstract) or 40 × magnification (figure 4B).

3. Results and discussion

3.1. Data processing considerations

Venturing into the ecology of AGS-forming communities is akin to exploring a dense forest where microbial taxa, many still hidden from view or exhibiting behaviours yet to be understood, plays the role of plants and animals interacting with each other while maintaining environmental equilibrium. As every exploration begins by testing the equipment, the first steps of this analysis were taken on the technical terrain that precedes any experiment design phase. Approximately 5–20 % of the raw reads in each sample were filtered out by BBDuk, primarily due to low quality scores. Subsequently, Kraken2 removed fewer than 100 reads per sample identified as of human origin, with the exception of dataset 3, where approximately 500 out of 30,000,000 raw reads pairs per sample (less than 0.0001 %) were classified as human-derived. Accordingly, the percentage of human contaminant reads was nearly zero, even in reactors fed with real wastewater. In light of this, the decontamination step may not be advisable in most scenarios, given the risk of misclassifying key AS and AGS bacteria as *Homo sapiens* (Di Gloria et al., 2025). Despite most reads passing these processing steps, approximately 75 % of the reads in each sample remained unclassified by Kaiju with the current settings. Of these, up to approximately 30 % may be attributed to the intrinsic limitations of the classifier and related settings (Di Gloria et al., 2025). Thus, we speculate that currently only about half of the characteristic microbes of AGS (including viruses) are represented in the NCBI database. A total of 7584 genera were identified across samples, of which 5663 (74.7 %) belong to bacteria or archaea clades.

Finally, the rarefaction curves of each sample were computed to assess their ecological saturation (Fig. 1, Supplementary Fig. 1).

All samples showed saturation of prokaryotes at the genus level (Fig. 1A), although the samples from dataset 1, characterized by the lowest sequencing depth in this meta-analysis, appeared poorly saturated. When rarefaction curves were computed at the species level (Fig. 1B), the dataset 1 samples were highly unsaturated, and none of the curves reached a plateau, including datasets 5 and 8, which had the highest read counts. Moreover, even in the datasets with the highest sequencing depth, species-level rarefaction curves did not reach a complete plateau. Filtering out rare prokaryotes with an average relative abundance below 0.0001 % was sufficient to achieve saturation at the species level for all samples. On the other hand, most datasets appeared ecologically unsaturated when analysing eukaryotic and viral diversity, particularly in the case of viruses (Fig. S1). This pattern persisted even after filtering out eukaryotes and viruses with an average relative abundance below 0.0001 %, with only samples featuring a sequencing depth equal to or greater than 10 GB achieving ecological saturation at the eukaryotic genus level. Taken together, these results indicate that a sequencing depth of 5 GB (considering a paired-end sequencing with read length of 150 bp) is not sufficient to fully capture the ecological complexity of an AGS reactor. However, even a sequencing depth of 20 GB was insufficient to reach a complete plateau in the prokaryotic species-level and eukaryotic genus-level saturation curves in a pilot-scale reactor, at least without applying average abundance filters (which are recommendable in meta-genomic analyses (Amos et al., 2020; Di Gloria et al., 2025)). Moreover, the viral saturation curves showed the most pronounced limitation. A high sampling effort is crucial to estimate the environmental richness, as rare observations may otherwise be missed in the dataset (Sgarbi et al., 2020). Nonetheless, even unsaturated samples frequently usually capture the relative abundances of the most abundant observations within an acceptable margin of error (Sgarbi et al., 2020). In light of this, and considering the observed shapes of the rarefaction curves (i.e. their trend toward a

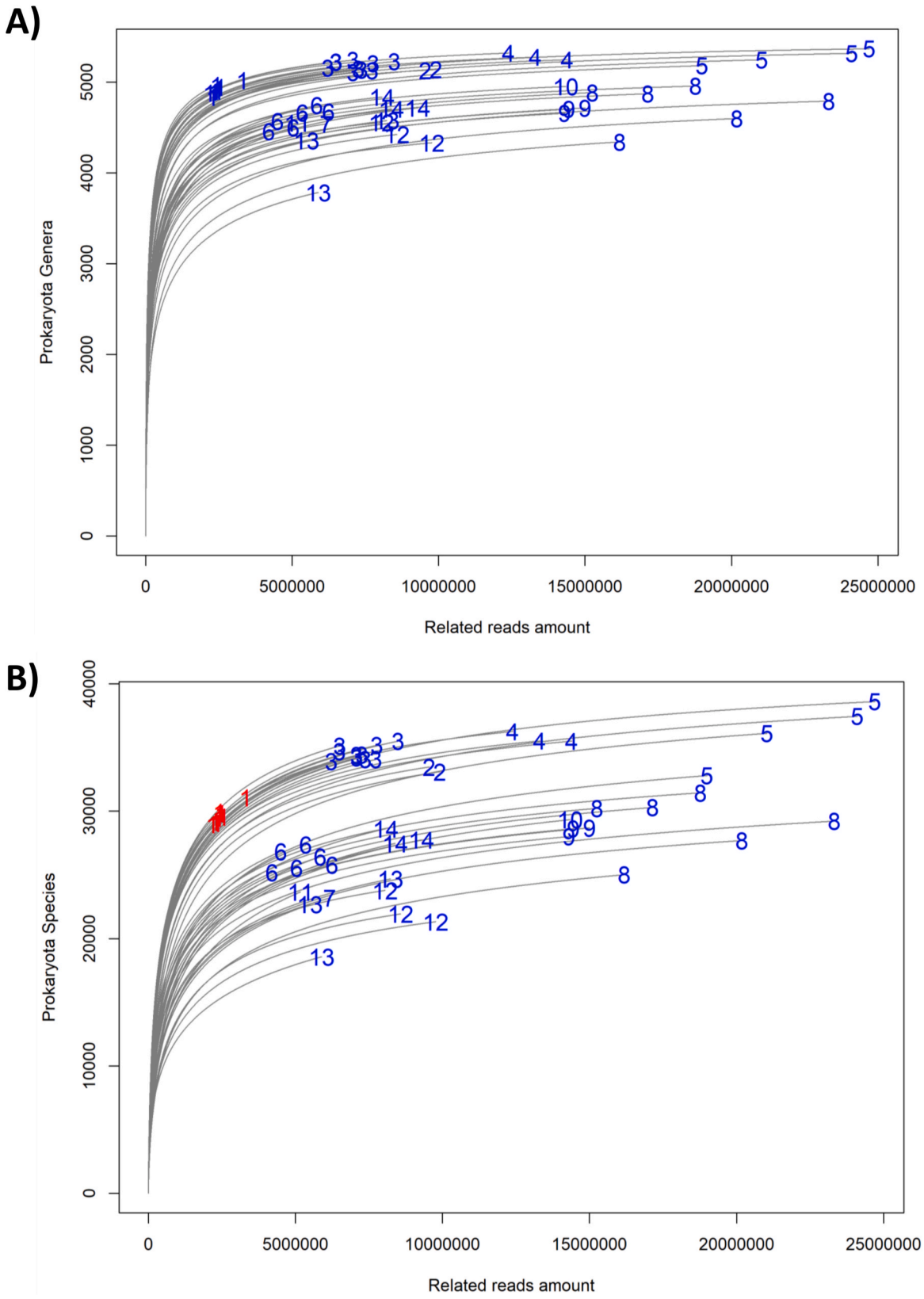


Fig. 1. Saturation curve of prokaryota genera (A) and prokaryota species (B) of each sample, displaying the number of distinct taxa observed with increasing read counts available for analysis, thus the samples' saturation degree. The dataset identification numbers are reported on the end of each related curve to distinguish samples of different datasets. The labels in red indicate the samples which were evaluated as highly unsaturated, as they remained far from the plateau.

plateau as read numbers increased), we conclude that at least 10 GB of sequencing depth is sufficient to achieve a reasonable ecological saturation at least in case of prokaryotic species and eukaryotic genera in AGS with the current short-read DNA sequencing technology, even in full-scale reactors, whereas 5 GB may still provide a cheaper superficial overview.

3.2. Overview of SBR community across highest taxonomic ranks

Overall, bacteria, archaea, virus and eukaryotes, including traces of fungi and algae, were detected in every reactor (Fig. 2).

In most samples, approximately 99 % to nearly 100 % of the reads were classified as bacterial, with a few exceptions showing distinct features. A more meticulous profiling of eukaryotes and viruses may require targeted approaches beyond bulk DNA sequencing, given that the AGS communities were composed almost entirely of prokaryotes. Samples 1 and 3 from dataset 6, which were obtained via DNA extraction using the CTAB protocol from reactors supplemented with complex monomeric carbon mixtures, exhibited slightly higher proportions of viral reads compared to other samples. Reactor 12 from dataset 8 showed an increased fungal proportion. The fungi overgrowth in dataset 8 was also discussed in the related publication (Burzio et al., 2024). These peculiarities may be explained by the DNA extraction method using the CTAB salt (common in virus purification protocols (Huang et al., 2000)) in the case of dataset 6, and by supplementation with sodium acetate and acetic acid in reactor 12, respectively. In particular, our results highlight that developing a method to concentrate viral DNA prior to sequencing is necessary to achieve ecological saturation of phage communities, at least in reactors fed with real wastewater where viral richness was markedly higher than in laboratory-scale reactors (Fig. S1, Fig. S6). The dataset 1 exhibited a higher predominance of Rotifers compared to other datasets, which may reflect its lower sequencing depth or a richness in planktonic bacteria susceptible to predation.

The observed proportions among domain abundances must consider that DNA from dead microbes, which is an integral structural component of AGS microbial matrices (Cardona et al., 2025) is sequenced alongside that from live cells. On the other hand, these proportions are also influenced by differences in genome sizes due to the metagenomics short-DNA sequencing protocol, with eukaryotes (especially multicellular ones, which harbour a genome copy in each cell) being correspondingly overestimated. In fact, in a community consisting of only 0.1 % eukaryotes, up to approximately 90 % of the extracted DNA may derive from this clade (Gabor et al., 2003). Conversely, viruses have small genomes and, in addition, are currently poorly represented in the genome databases (Nieuwenhuijse et al., 2020). Accordingly, accounting for these potential biases, we still interpret the previously observed predominance of bacteria in all AGS communities as a robust result, while viruses warrant further investigation.

3.3. Overview of SBR community at genus level

The differences across the datasets became more evident when the most abundant genera (Fig. 3, Table 2) and species (Fig. S2, Table S2) were assessed.

All the most abundant taxa were bacteria, with differences in abundance variability between full-scale and laboratory-scale reactors groups. In fact, the full-scale reactors displayed similar community profiles despite differing in settings and influent composition, whereas each laboratory-scale showed distinctive patterns. Notably, the genera ADurb.BinA028 (NCBI taxon ID 1866933), belonging to an identified bacteria clade, and *Candidatus* Brachybacter were consistently present in full-scales reactors with similar percent abundance but were detected only in trace amounts in laboratory-scale reactors. Conversely, *Micropruina* was abundant in laboratory-scale reactors of dataset 6 and 9 but occurred only in trace amounts in full-scales and pilot-scale reactors. The reactors 11 and 12 of the dataset 8, characterized by the absence and the addition of sodium acetate, respectively, showed markedly different profiles from each other. For example, the reactor 12 exhibited a strong predominance of *Zoogloea*, similarly to the dataset 9 reactor, which was characterized by a high COD:N:P ratio. *Candidatus* Accumulibacter was particularly abundant in the laboratory-scale reactors of datasets 5, 6 and 11, while its presence in other reactors was not as prominent. Similar results were observed also at species level. Notably, the most abundant identified species of *Candidatus* Accumulibacter genus were *Accumulibacter* sp. and *Candidatus* Accumulibacter sp. ACC003, while *Candidatus* Competibacter denitrificans was the predominant species of its clade, especially in datasets 5, 6 and 11 (Fig. S2A). Results remained largely consistent when restricting the computation of the most abundant taxa across samples of full and pilot-

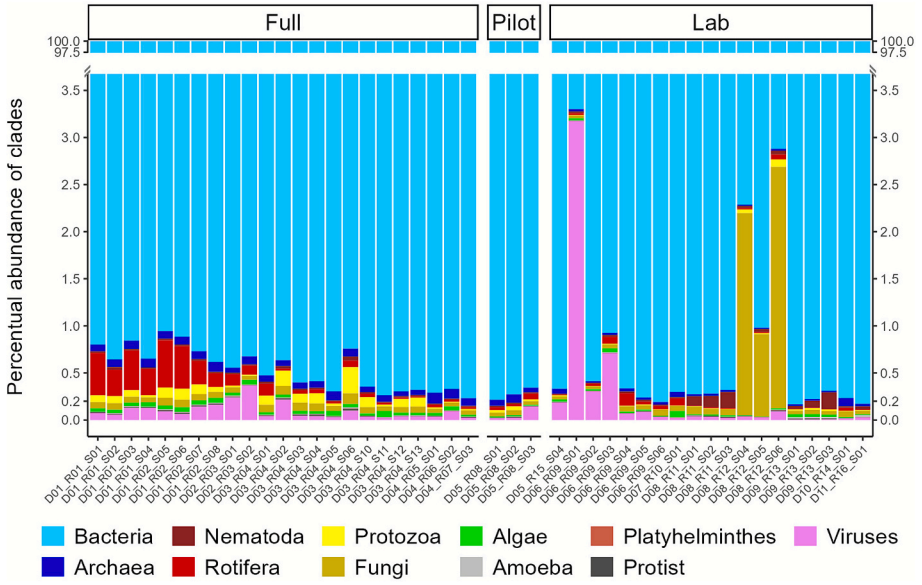


Fig. 2. Bar plot showing the proportions of the domains across SBRs samples. The reactor scales are indicated above each sample group, while sample names are shown on the x-axis ("D" = datasets, "R" = reactor, "S" = sample). Each domain is represented by a distinct colour. The y-axis values ranging from 3.5 % to 97.5 %, which represents only bacteria related abundances, were hidden from the plot to improve the visualization of less abundant domains.

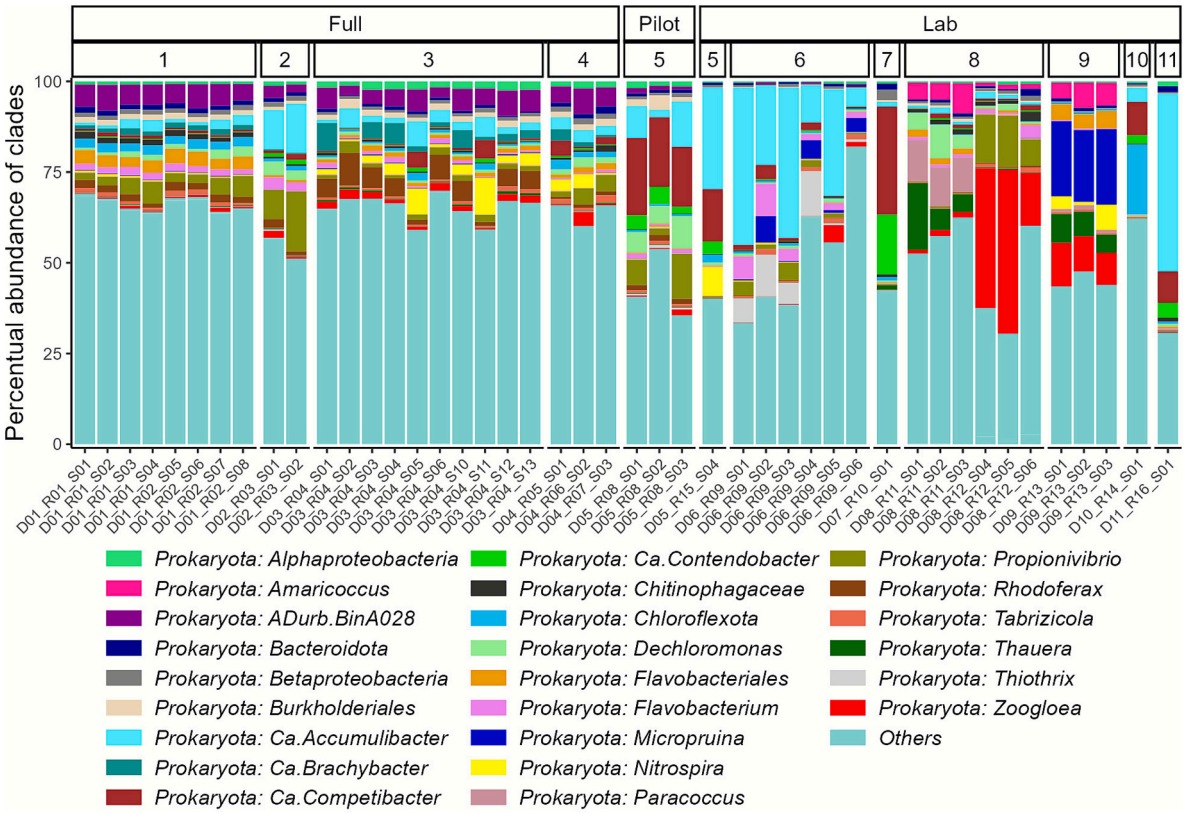


Fig. 3. Bar plot showing the proportions of the most abundant genera across SBRs samples. The reactor scales and the dataset identification numbers are indicated above each sample group, while sample names are shown on the x-axis ("D" = datasets, "R" = reactor, "S" = sample). Each genus is represented by a distinct colour, whereas the group "Others" include every other less abundant genera.

Table 2
Average percent abundances of the most abundant genera across SBR samples. The averages are computed on the total microbial community. The "Main role" column reports the most known role or characteristic of that clade in AS or AGS environments and it is not intended to be a comprehensive summary of the clade's potential functions or characteristics. "PAO = polyphosphate-accumulating bacteria, "GAO" = glycogen-accumulating bacteria, "AOO" = ammonia-oxidizing organisms, "NOO" = nitrite-oxidizing organisms.

Overall average	Averages Full-scales	Averages Pilot-scales	Averages Lab-scales	Taxon	Domain	Main role
7.61	3.44	7.78	12.63	*Candidatus* Accumulibacter	Bacteria	PAO
3.79	1.21	0.79	7.39	*Zoogloea*	Bacteria	
3.54	1.18	18.96	3.95	*Candidatus* Competibacter	Bacteria	GAO
3.34	3.47	7.03	2.59	*Propionivibrio*	Bacteria	GAO
2.72	5.12	1.11	0.07	*ADurb.BinA028*	Bacteria	
1.85	0.15	0.06	4.18	*Micropruina*	Bacteria	GAO
1.77	3.13	1.53	0.15	*Rhodoferax*	Bacteria	
1.70	1.30	6.44	1.43	*Dechloromonas*	Bacteria	PAO
1.48	1.31	1.00	1.76	*Flavobacterium*	Bacteria	
1.46	1.92	0.18	1.10	*Nitrospira*	Bacteria	NOO
1.45	1.92	0.22	1.09	*Flavobacteriales*	Bacteria	
1.42	1.54	0.56	1.42	*Chloroflexota*	Bacteria	
1.30	0.28	0.20	2.71	*Thauera*	Bacteria	
1.07	2.08	0.01	0.02	*Candidatus* Brachybacter	Bacteria	
1.05	0.35	3.52	1.52	*Candidatus* Contendobacter	Bacteria	GAO
1.04	1.44	1.48	0.50	*Alphaproteobacteria*	Bacteria	
1.03	1.48	2.33	0.27	*Burkholderiales*	Bacteria	
0.95	1.09	0.93	0.78	*Bacteroidota*	Bacteria	
0.92	0.09	0.56	1.98	*Thiothrix*	Bacteria	
0.91	1.12	0.86	0.66	*Betaproteobacteria*	Bacteria	
0.90	0.06	0.03	2.05	*Amaricoccus*	Bacteria	
0.88	0.14	0.14	1.88	*Paracoccus*	Bacteria	
0.75	0.88	0.14	0.69	*Chitinophagaceae*	Bacteria	
0.74	0.85	0.83	0.60	*Tabrizicola*	Bacteria	

scale reactors, even at the species level (Fig. S2B). Notably, even under this restriction, the variability across full-scale datasets was driven mainly by differences in clade proportions rather than by species presence, at least for most of the dominant species.

Among the most abundant archaeal genera (Fig. S3, Table S3), *Methanothrix* and *Methanosaeta* were predominantly observed in full-

scale and pilot-scale reactors, whereas *Candidatus* Nitrosopumilus was more prevalent in most laboratory-scale reactors. The most abundant viral taxa (Fig. S4, Table S4) largely belonged to the Caudoviricetes class, including representatives of the "EBPR podoviruses". Notably, *Klosneuvirus KNV1* and the clade associated with taxon ID 340016 (a broad group of uncultured environmental viruses targeting protozoa (Schoch, 2020) were detected in most of the full-scale reactors. In addition, reads assigned to Human Papillomavirus (HPV) were also found in considerable abundance, although this result is likely attributable to misclassification (as discussed below).

As for the most abundant eukaryotes (Fig. S5, Table S5), the rotifer genera *Rotaria* and *Adineta*, along with *Symbiodinium*, were common and abundant across samples. In particular, *Rotaria* was more predominant in full and pilot-scale reactors than in laboratory-scale ones. Conversely, the relative abundance of the nematode *Diploscapter* was higher in laboratory-scale samples. Fungi such as *Rhizopus*, *Lipomyces*, *Aspergillus* and *Russula*, as well as the nematodes *Brugia* and *Anisakis*, showed similar average abundances across reactor types.

It is noteworthy that the observation of *Brugia*, *Anisaki* and *Symbiodinium* may be a misclassification. The chromist Symbiodinium, although belonging to the paraphyletic group "Alveolata" according to NCBI taxonomy (Cavalier-Smith, 2018), is a microscopic algae-like organism living as an endosymbiotic and photosynthetic autotroph (Roth, 2014). Considering the limited light penetration in AGS reactors and within their mixed liquor, we assume that its recurrent detection across samples is more likely the result of a misclassification. To further support this, *Symbiodinium* was frequently misassigned from *Candidatus* Competibacter phosphatis reads in the AS and AGS microbial

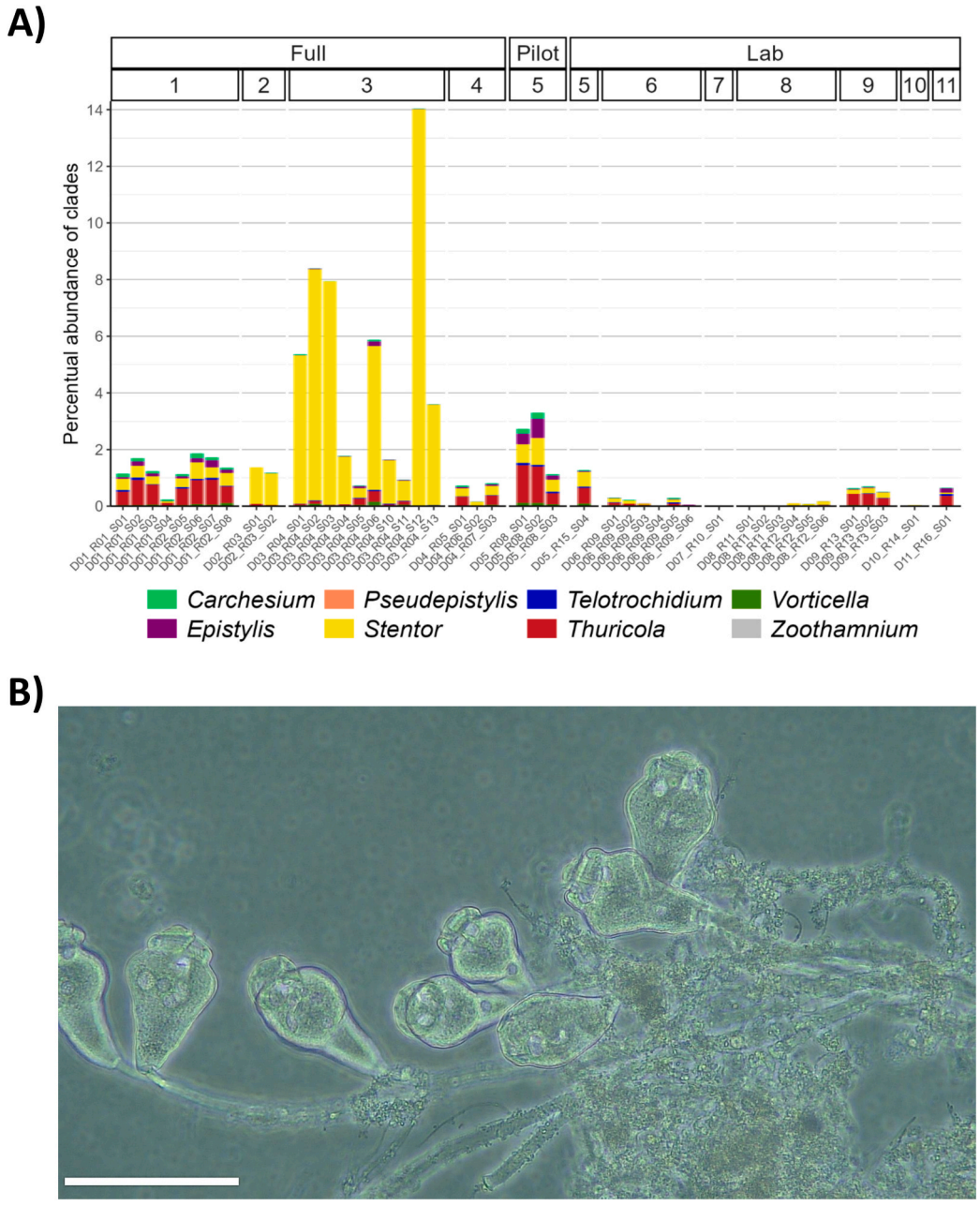


Fig. 4. (A) Bar plot showing the proportions of the sessile ciliates across SBRs samples. The reactor scales and the dataset identification numbers are indicated above each sample group, while sample names are shown on the x-axis ("D" = dataset, "R" = reactor, "S" = sample). The percent abundances displayed along the y-axis are computed using only reads assigned to the Eukaryota domain. (B) Image of the sessile ciliates observed on the top of the pilot-scale reactors' granules by light microscopy at 40 × magnification. Their taxonomic identification is detailed in the result section of the paper. The white scale bar corresponds to 100 μm.

community tailored benchmark (Di Gloria et al., 2025). Likewise, the nematodes *Brugia* and *Anisakis*, detected here at extremely low abundances across samples, were also erroneously classified from *Homo* or *Diploscapter* reads in the same benchmark (Di Gloria et al., 2025), indicating that these sequences likely originated from *Diploscapter*.

Notably, the fungal genus *Magnusiomyces* dominated the eukaryotic community in reactor 12 of dataset 8.

The Heterotrichida *Stentor* was also featured among the most abundant eukaryotes. All of the sessile ciliates' reads were taxonomically assigned to the genera *Carchesium, Epistylis, Pseudepistylis, Stentor, Telotrochidium, Thuricola, Vorticella* or *Zoothamnium* (Fig. 4A).

In particular, *Stentor* and *Thuricola* were the most prevalent sessile ciliates across samples, whereas *Vorticella* and *Zoothamnium* were the least abundant. Moreover, the sessile ciliates were almost undetected across laboratory-scale reactors, potentially as a result of the absence of particulate matter in the synthetic influent (Barrios-Hernández et al., 2021).

3.4. Further insights into ciliates population

The presence of stalked ciliates on the granules of the pilot-scale reactor of dataset 5 was confirmed by light microscopy (Fig. 4B). Based on zooid form and size, these ciliates could be assigned to *Carchesium*, *Vorticella* or *Epistylis* species (Rivas-Castillo et al., 2022). The identification was narrowed to *Carchesium* or *Epistylis* due to the stalk ramification, since *Vorticella* spp. form colonies with zooids on single stalks (Amos, 1972; Rivas-Castillo et al., 2022). Finally, the independent contraction of stalks led to identify them as *Carchesium* spp. (Rivas-Castillo et al., 2022), corroborating the identification of this genus through DNA-sequencing. The low abundance of Vorticella-assigned reads across reactors, combined with the lack of Vorticella detection by microscopy, challenge the use of the term "*Vorticella*-like organisms" used in AGS related literature (Wilén et al., 2018; Winkler et al., 2012), at least assuming that the low detection is not attributable to under-representation in reference databases. It should be noted that the absence of *Vorticella* in our microscopy observations does not exclude its occurrence in samples from other reactors. In fact, different ciliates exhibit distinct ecological preferences and tolerances to wastewater compounds. For instance, *Vorticella* is reported to thrive in environments with low organic loads and to be capable of degrading petroleum hydrocarbons while tolerating heavy metals (Rivas-Castillo et al., 2022). Nevertheless, the presence of *Carchesium*, *Epistylis* or *Vorticella* is considered a potential indicator of good reactor performance in AS and AGS systems (Rivas-Castillo et al., 2022; Weber et al., 2007).

3.5. Functional group analysis of SBR

Peculiar patterns were observed also when the bacteria and the archaea were grouped by their putative functional groups, namely AOOs (ammonia oxidizing organisms), NOOs (nitrite oxidizing organisms), PAOs, GAOs and other PHA accumulating organisms with different metabolisms from PAOs and GAOs, such as *Thauera* and *Zoogloea* (Fig. 5, Table S6).

In general, PAOs, GAOs, AOOs and NOOs were well represented across reactors, displaying patterns already well described in the literature, such as the higher PAO:GAO ratio in settings with high carbon and phosphorous concentration or in case of alternation of molecule used as readily degradable carbon supply (Wang et al., 2021) (Fig. 5).

The full-scale reactors exhibited lower relative abundance of PAOs and GAOs than laboratory-scale reactors, with the predominance of *Candidatus* Accumulibacter on others PAOs abundances being less pronounced in laboratory-scale ones. *Candidatus* Neomicrothrix, *Nostocoides* (also referred to as "Tetrasphaera"), *Tessaracoccus, Gemmatimonas* and *Azonexus* were among the most abundant PAOs alongside with *Candidatus* Accumulibacter. Regarding GAOs, *Candidatus* Competibacter was not consistently the dominant GAO, except in reactors from datasets 5, 7 and 10. In the laboratory-scale reactor group, the most abundant GAO was *Micropruina*, whereas in full-scale reactors it was *Propionivibrio*. However, overall average abundance of *Micropruina* was disproportionately influenced by dataset 9. Accordingly, *Propionivibrio* is confirmed as the most abundant and recurrent GAO across reactor scales, corroborating recent findings in the literature (Ekholm et al., 2024; Yuan et al., 2024). Moreover, Cardona et al. linked *Propionivibrio* activity to granule stability and inferred its metabolic interactions with

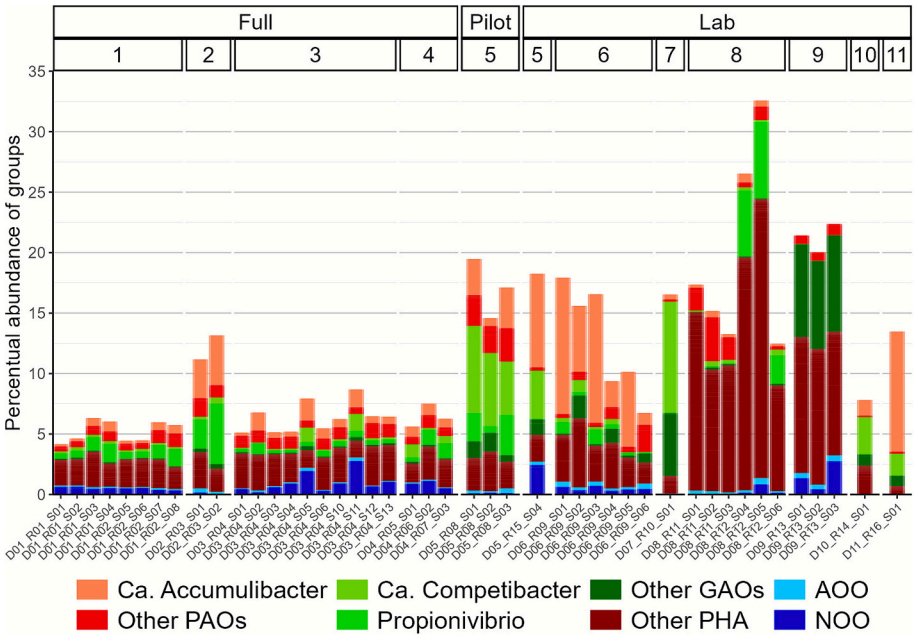


Fig. 5. Bar plot showing the proportions of the organisms functional groups across SBRs samples. The reactor scales and the dataset identification numbers are indicated above each sample group, while sample names are shown on the x-axis ("D" = dataset, "R" = reactor, "S" = sample). Each group is represented by a distinct colour, whereas the group "Others PHA" refers to the PHA accumulating organisms with different metabolisms from PAOs and GAOs. The percent abundances are computed including also the taxa not categorized into these functional groups, which were omitted in this plot.

Candidatus Accumulibacter (Cardona et al., 2025). The consistent detection of *Propionivibrio* also highlights the importance of a methodological bias already reported in literature: the 16S probes PAO462 and PAO846, used alongside PAO651 to identify *Candidatus* Accumulibacter, can also hybridize with *Propionivibrio* 16S sequences (Yuan et al., 2024), inflating PAO abundance in studies using this probe mix. Nevertheless, both recent literature and our results suggest that *Propionivibrio* warrants as much attention as *Candidatus* Competibacter, which has traditionally been considered the model GAO.

Among the AOOs, *Nitrosomonas* was the most abundant genus across samples, followed by *Nitrosospira* and *Nitrosococcus*. As for the NOOs, *Nitrospira* was the most abundant across samples, followed by *Candidatus* Nitrotoga and *Nitrobacter*.

A total of 72 PHA producing bacteria not classifiable as PAOs and GAOs were identified, in particular in the reactors of datasets 8 and 9. Among these bacteria, *Zoogloea*, *Dechloromonas*, *Thauera*, *Flavobacterium*, *Amaricoccus, Tsukamurella* and *Paracoccus* had the high relative abundances across reactors. It is noteworthy that several additional bacteria belonging to these functional groups were identified, with the complete list provided in Table S6.

3.6. Alpha and beta diversity of SBR

Alpha diversity indices of prokaryotes, eukaryotes and viruses (Fig. S6) revealed substantially higher observed richness in full and pilot-scale reactors compared to laboratory-scale reactors, especially for prokaryotes and viruses. Moreover, prokaryotes communities exhibited greater abundances evenness in full-scale systems, whereas pilot and most laboratory-scale reactors were more dominated by few prokaryotic taxa. Beta diversity analysis (Fig. S7) confirmed the mutual similarity among full-scale reactors, with the pilot-scale reactor showing closer affinity to the full-scale systems. However, the percentage of variance explained by the two axes and the overall sample distribution did not indicate markedly distinct prokaryote compositions across reactor scales. In the case of eukaryotes, inter-sample differences were less pronounced, also due to the inclusion of reactor 12 from dataset 8, the sample of which appeared as outliers in the PCoA plot.

3.7. Core microbiota of SBR with comparison to CFR

The observed "AGS core microbiota", defined in this paper as the recurrent taxa across the various samples, comprised 1622 genera (Table S7), the most abundant of which were all bacteria (Table 3).

These included PAOs, GAOs, AOOs, NOOs, as well as other PHA-accumulating bacteria distinct from PAOs and GAOs. Most of these abundant and recurrent bacterial taxa were also observed also in CFR samples, including *Zoogloea*, *Rhodoferax*, *Thauera*, the GAOs *Micropruina* and *Propionivibrio*, the PAO *Dechloromonas,* the AOO Nitrosomonas as well as the NOOs *Nitrospira* and *Candidatus* Nitrotoga. Moreover, *Candidatus* Accumulibacter, *Candidatus* Competibacter, *Candidatus* Contendobacter and *Paracoccus* resulted among the most abundant genera across the samples of both the reactor types (Fig. 3 and Fig. S8 A). Yet, *Candidatus* Accumulibacter and *Candidatus* Competibacter were not included in CFRs core microbiota, as their occurrence did not reach the thresholds applied in this study (Table 3). Additional CFR samples are required to corroborate this result.

As expected, *Nitrosomonas* was significantly more abundant in samples from dataset 14 (Fig. S8A), which was based on a partial nitrification system. While the denitrification potential is not addressed in this paper to avoid over-complication, it is noteworthy that many of the most abundant bacteria and core taxa are also known denitrifiers.

Although not among the most abundant taxa, it is noteworthy that the bacterial genera *Bdellovibrio* and *Micavibrio* were also part of the SBR and CFR core communities.

Table 3
Average percent abundances of the most abundant genera frequently observed across full, pilot and laboratory-scales SBRs. The "Overall averages across common" column displays the percentages averages reported calculated by including only the common genera identified as recurrent within each scale, in order to better reflect the proportions among these genera. The "Main role" column reports the most known role or characteristic of that clade in AS or AGS environments and it is not intended to be a comprehensive summary of the clade's potential functions or characteristics. The "Also in CFR" column indicates if the taxon was found also as recurrent observation across CFR samples.

Overall averages across common	Averages Full-scales	Averages Pilot-scales	Averages Lab-scales	Taxon	Domain	Main role	Also in CFR
1.94	0.95	2.32	3.08	*Candidatus* Accumulibacter	Bacteria	PAO	
1.42	0.33	0.23	2.91	*Zoogloea*	Bacteria		*
1.03	0.32	5.88	1.13	*Candidatus* Competibacter	Bacteria	GAO	
1.02	0.95	2.08	0.94	*Propionivibrio*	Bacteria	GAO	*
0.73	1.37	0.35	0.02	*ADurb.BinA028*	Bacteria		
0.62	0.04	0.02	1.42	*Micropruina*	Bacteria	GAO	*
0.50	0.35	1.93	0.45	*Dechloromonas*	Bacteria	PAO	*
0.49	0.85	0.47	0.05	*Rhodoferax*	Bacteria		*
0.44	0.08	0.06	0.93	*Thauera*	Bacteria		*
0.42	0.52	0.06	0.36	*Nitrospira*	Bacteria	NOO	*
0.42	0.50	0.07	0.39	*Flavobacteriales*	Bacteria		*
0.41	0.40	0.17	0.45	*Chloroflexota*	Bacteria		*
0.40	0.35	0.31	0.48	*Flavobacterium*	Bacteria		*
0.31	0.02	0.01	0.72	*Amaricoccus*	Bacteria		*
0.31	0.10	1.11	0.44	*Candidatus* Contendobacter	Bacteria	GAO	
0.29	0.39	0.46	0.15	*Alphaproteobacteria*	Bacteria		*
0.29	0.56	0.00	0.01	*Candidatus* Brachybacter	Bacteria		
0.29	0.40	0.73	0.08	*Burkholderiales*	Bacteria		*
0.28	0.04	0.04	0.61	*Paracoccus*	Bacteria		*
0.27	0.29	0.29	0.24	*Bacteroidota*	Bacteria		*
0.26	0.30	0.26	0.21	*Betaproteobacteria*	Bacteria		*
0.22	0.02	0.17	0.47	*Thiothrix*	Bacteria		*
0.21	0.23	0.04	0.22	*Chitinophagaceae*	Bacteria		*
0.20	0.22	0.26	0.17	*Tabrizicola*	Bacteria		*
0.18	0.23	0.19	0.12	*Rhodocyclaceae*	Bacteria		*
0.18	0.19	0.05	0.19	*Candidatus* Nitrotoga	Bacteria	NOO	
0.18	0.11	0.23	0.25	*Nitrosomonas*	Bacteria	AOO	*
0.17	0.27	0.12	0.07	*Sulfuritalea*	Bacteria		*
0.17	0.16	0.27	0.17	*Bacteroidetes*	Bacteria		*

The archaea *Methanothrix*, *Methanosarcina* and *Candidatus* Nitrosopumilus were also identified as core members of the SBR and CFR communities (Supplemetary Table 8) with average percent abundance over 0.001 %. The presence of various methanogenic archaea even among the core taxa is an important observation when discussing the functions inferred from the taxonomy, as the detection these clades may suggest the potential role of AGS in the context of the green economy. At this regard, many PHA-producing organisms with metabolism different from PAOs and GAOs, such as *Thauera, Paracoccus* and *Zoogloea*, were also identified as core AGS taxa. These bacteria likely contribute to consuming the carbon remaining available during the aerobic phase. Ji et al. likewise reported *Thauera* and *Paracoccus* among the most abundant bacteria in their laboratory-scale AGS reactor operated under salinity conditions, alongside other taxa such as *Azoarcus*, *Zoogloea* and *Candidatus* Competibacter, and emphasized their role in secreting polymeric substances essential for granule stability (Ji et al., 2025).

As for the eukaryotic taxa (Supplemetary Table 8), the fungi *Rhizopus*, *Vanrija, Metschnikowia, Batrachochytrium* and *Cutaneotrichosporon*, together with *Tetrahymena*, *Blepharisma*, the rotifers *Rotaria* and *Adineta* (both belonging to the detritivores Bdelloides class), *Symbiodinium* as well as the nematode *Diploscapter*, were consistently observed across both SBRs and CFRs.

Rotifers and Nematodes, along with protozoa, are eukaryotes commonly observed in WWTPs (Bergtold et al., 2007; Li et al., 2013). They feed on microorganisms and graze complex structures, thereby contributing to the mineralization of substances that, in turn, benefit their prey (Barrios-Hernández et al., 2021; Li et al., 2013; Rueda-Ramírez et al., 2023). On the other hand, fungi are seldom discussed in AGS-related publications, despite they may fulfil activities such as being additional backbones for the granulation or participating in the element cycles (Sharaf et al., 2019; Weber et al., 2007; Wilén et al., 2018). Nevertheless, fungi relative abundance is expected to be extremely low in a properly functioning AGS, as observed in our results, otherwise their overgrowth can lead to filamentous granules with lower settling velocity (Burzio et al., 2024; Sharaf et al., 2019).

Regarding sessile ciliates, the genus *Stentor* (class Heterotrichea) was identified as part of the AGS core microbiota in SBRs, however it was not always found in CFRs samples. Additionally, *Epistylis*, *Carchesium* and *Thuricola* (class Sessilida) were recurrent across reactors, though not ubiquitous.

Concerning exclusive occurrences, species of the uncultured bacterium A1Q1 as well as the fungi *Magnusiomyces* and *Apiotrichum* were identified as taxa unique to full and pilot-scale reactors with average percent abundance over 0.001 %. However, these species were also detected in trace amounts in a few laboratory-scale reactors, although their recurrence did not meet the threshold required to be classified as "core members" within the laboratory-scale group. Moreover, ADurb. BinA028 ranked among the most abundant genera in SBRs (Table 2) and occurred mainly in reactors fed with real wastewater. This genus remains uncultured to date and corresponds to a bacterial draft genome reconstructed from full-scale anaerobic digesters associated with AS reactors (NCBI BioSample: SAMN05004214, 2016). Its genome predicts the presence of proteins such as Polyphosphate Kinase 2 (PPK2) and PHA depolymerase (NCBI Protein database, 2025), supporting the plausibility of its presence in this environment, though it may alternatively represent a composite genome generated by chimeric assembly.

In conclusion, it is noteworthy that although CFR samples formed a distinct cluster in the PCoA (Fig. S8B), the community profile divergence between SBR and CFR was not particularly marked. Indeed, CFR samples from dataset 12 clustered closely with the SBR sample from dataset 7, from which their inoculum originated (Chen et al., 2024).

3.8. Ecological perspectives

The microbial richness is higher in full-scale and pilot-scale reactors (Fig. S6), which are feed on real wastewater, in line with the idea that multitude of microbes are introduced by real influents. However, these reactors also display higher abundances' evenness compared to the laboratory-scale reactors, despite the simpler ecological landscapes of the latter. This outcome aligns with the Intermediate Disturbance Hypothesis (IDH), which affirms that the biodiversity and the abundance evenness are higher when moderate disturbances occur intermittently, allowing diverse taxa to thrive at different times (Harris, 2020; Osman, 2015). In fact, in real scenarios, disturbances as the arrival of diverse competitors, fluctuations in nutrients, antibiotics and temperature may temporarily affect the reactors, whereas more stable and defined conditions are set in laboratory-scale reactors. On note, the nutrient's concentration or type are not usually listed as potential disturbances in IDH, but we still include them here on account of the different metabolisms and subsequent grow rates, which underpin AGS ecology. In fact, in real influents also a broader spectrum of organic substrates is expected compared to synthetic influents. In addition to the environment influences, the members of the ecosystem regulate themselves, as microbes produce molecules to either support or inhibit the thriving of certain others (Berg et al., 2020; Mihaylova-Garnizova et al., 2024; Wilén et al., 2018). As a result, it can be hypothesized that when a few organisms become dominant, they further shape the community by subsequently reducing abundance evenness. In light of these results and considerations, we assume that reactors fed with real influent are far less prone to "ecological drifts" (i.e., the development of different communities from the same inoculum due to stochastic factors) typically occurring in laboratory-scale reactors (Zhou et al., 2013), although experimental validation is required to empirically validate these assumptions. Notably, also the inter-sample variability among full-scale reactors was markedly lower than that of laboratory-scale reactors, despite the lower microbial richness of the latter (Fig. 3, Fig. S6, Fig. S7). The observed similarity in the most abundant taxa across full-scale reactors was unexpected, given the substantial differences in influents and operational settings. Nevertheless, differences are expected when focusing on individual reactors or on less abundant bacteria, which are better described in the reactor-specific studies (Table 1). Moreover, taxonomic similarity does not prove an identical microbial activity profile, as different stimuli (e.g. influent characteristics) lead to diverse gene transcriptional regulation even within the same strain.

Furthermore, the presence of phages limits the number and diversity of bacteria, acting like an uncontrolled immune system within the reactor. Most viruses detected across the AGS reactors belonged to the Caudovirales order, consistent with previous reports of a high abundance of viral tail-like proteins in these system (Barr et al., 2010) (Fig. S4, Table S4). For example, the EBPR podoviruses were among the most abundant identified viruses and they are known to infect *Candidatus* Accumulibacter strains (Mihara et al., 2016), which are themselves highly prevalent in the community. For the same reasons, we assume that the human Papillomavirus, inferred among the dominant viral taxa, likely resulted from the misclassification of a similar and yet unknown phage. Regarding strain-specific predation, it is also significant that the bacterial genera *Micavibrio* and *Bdellovibrio*, which harbour strains identified as obligate predators of *Nitrospira* lineage I and as specialized parasites of PHA-producing bacteria (Dolinšek et al., 2013; Wang et al., 2011), respectively, resulted to be part of the AGS core microbiota.

3.9. Paper limitations

Given the nature of this study, inherent limitations must be disclosed and emphasized. For instance, the methods employed do not encompass RNA viruses. In addition, it is noteworthy that the use of different DNA extraction kits may bias the representation of certain species across datasets, potentially resulting in underestimating the abundance for some taxa (Gabor et al., 2003; Szóstak et al., 2022). However, this does not undermine the conclusions regarding taxa that were consistently observed across reactors. Moreover, the discussion in based primarily on the genera rather than species, with particular focus on the observations

exceeding very low abundance thresholds (e.g. 0.0001 %). Although this approach was preferred also to mitigate the risk of false positives, as low abundant counts are often attributable to misclassifications (Di Gloria et al., 2025), it may have led to the omission of rare taxa. Nevertheless, species level results and comprehensive abundance tables were still reported in the results section and supplementary data, enabling further exploration. Either way, the detection of certain taxa does not demonstrate their functional involvement. However, in light of Baas Becking's renowned phrase "*Everything is everywhere, but the environment selects*" (De Wit & Bouvier, 2006), we regard observed abundances as potential indications of the microbial suitability to the environment, consequently providing a qualitative inference of their characteristic activity, albeit not quantitative. The number of reactors currently available in the literature is insufficient to adequately explore similarities and differences between AGS SBRs and CFRs, which we could only examine superficially (Fig. S8). Likewise, we deemed that it was not possible to perform robust statistical comparisons between different operational settings due to the insufficient number of datasets per each setting group, particularly given the intrinsically high variability of laboratory-scale communities.

4. Conclusions

Notwithstanding its limitations, this work stands as the first meta-analysis highlighting the most characteristic microbes and the technical limitations that can be expected when studying an AGS reactor. Despite substantial variations in influent composition and reactors operative settings, we observed a notable similarity in the most abundant bacteria across reactors fed with real wastewater, whereas the previously reported high variability in laboratory-scale systems was confirmed. Additionally, we provide novel insights into the potential abundance of key AGS eukaryotes, including sessile ciliates. While further details will emerge as more AGS total DNA sequences become publicly available, our results revealed a conserved AGS core microbiota while offering a springboard for future microbial ecology studies in engineered systems.

CRediT authorship contribution statement

Leandro Di Gloria: Writing – review & editing, Writing – original draft, Software, Investigation, Formal analysis, Data curation, Conceptualization. **Tommaso Lotti:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Mark CM van Loosdrecht:** Writing – review & editing, Supervision. **Matteo Ramazzotti:** Writing – review & editing, Validation, Supervision, Resources, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by Piano Nazionale Di Ripresa E Resilienza (PNRR), Missione 4 "Istruzione e Ricerca" – Componente *C*2, Investimento 1.1, "Fondo per il Programma Nazionale di Ricerca e Progetti di Rilevante Interesse Nazionale (PRIN)". Project code MUR: 20228AF8JN. CUP: B53D23005590006. Title: "Integrating innovative *N*-removing biofilm processes and excess sludge valorization technologies for the development of energy- and material-efficient wastewater treatment plants" – Acronym: "N4En" and partly by POR FESR Toscana 2014–2020 ("IDRO.SMART" project, CUP: 3647.04032020.15700 0040).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biortech.2025.133657.

Data availability

All sequences analysed in this study are publicly available on the NCBI SRA under the accession codes listed in Table 1 column 3: "SRA ID". The scripts, the processed data and further details are accessible via the GitHub repository at the link https://github.com/LeandroD94/Papers/tree/main/2025_AGS_across_Domains.

References

Amos, G.C.A., Logan, A., Anwar, S., Fritzsche, M., Mate, R., Bleazard, T., Rijpkema, S., 2020. Developing standards for the microbiome field. Microbiome 8 (1), 98.
Amos, W.B., 1972. Structure and coiling of the stalk in the peritrich ciliates Vorticella and Carchesium. J. Cell Sci. 10 (1), 95–122.
Barr, J.J., Slater, F.R., Fukushima, T., Bond, P.L., 2010. Evidence for bacteriophage activity causing community and performance changes in a phosphorus-removal activated sludge. FEMS Microbiol. Ecol. 74 (3), 631–642.
Barrios-Hernández, M.L., Bettinelli, C., Mora-Cabrera, K., Vanegas-Camero, M.C., Garcia, H., van de Vossenberg, J., Prats, D., Brdjanovic, D., van Loosdrecht, M.C.M., Hooijmans, C.M., 2021. Unravelling the removal mechanisms of bacterial and viral surrogates in aerobic granular sludge systems. Water Res. 195, 116992.
Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C.-C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G.H., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J.A., Maguin, E., Mauchline, T., McClure, R., Mitter, B., Ryan, M., Sarand, I., Smidt, H., Schelkle, B., Roume, H., Kiran, G.S., Selvin, J., Souza, R.S.C.D., van Overbeek, L., Singh, B.K., Wagner, M., Walsh, A., Sessitsch, A., Schloter, M., 2020. Microbiome definition re-visited: Old concepts and new challenges. Microbiome 8 (1), 103.
Bergtold, M., Mayr, G., Traunspurger, W., 2007. Nematodes in wastewater biofilms—Appearance and density of species in three biofilter reactors. Water Res. 41 (1), 145–151.
Burzio, C., Mohammadi, A.S., Smith, S., Abadikhah, M., Svahn, O., Modin, O., Persson, F., Wilén, B.-M., 2024. Sorption of pharmaceuticals to foam and aerobic granular sludge with different morphologies. Resour. Environ. Sustainability 15, 100149.
Cardona, L., Saini, J.S., Rodilla Ramírez, P.N., Adler, A., Holliger, C., 2025. Multiple extracellular polymeric substance pathways transcribed by Accumulibacter and the flanking community during aerobic granule formation and after influent modification. Appl. Environ. Microbiol. 91 (4), e0176924.
Cavalier-Smith, T., 2018. Kingdom Chromista and its eight phyla: a new synthesis emphasising periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. Protoplasma 255 (1), 297–357.
Chen, B., Li, Y., Luo, Z., Lei, M., Li, J., Zhang, X., 2024. Enhanced phosphorus removal from biological secondary effluent using denitrifying phosphorus-removal granular sludge. J. Water Process Eng. 57, 104584.
De Wit, R., Bouvier, T., 2006. 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say?, 8(4), 755–758.
Di Gloria, L., Casbarra, L., Lotti, T., Ramazzotti, M., 2025. Testing the limits of short-reads metagenomic classifications programs in wastewater treating microbial communities. Sci. Rep. 15 (1), 23997.
Dolinšek, J., Lagkouvardos, I., Wanek, W., Wagner, M., Daims, H., 2013. Interactions of nitrifying bacteria and heterotrophs: Identification of a micavibrio-like putative predator of Nitrospira spp. Appl. Environ. Microbiol. 79 (6), 2027–2037.
Ekholm, J., Persson, F., de Blois, M., Modin, O., Gustavsson, D.J.I., Pronk, M., van Loosdrecht, M.C.M., Wilén, B.-M., 2024. Microbiome structure and function in parallel full-scale aerobic granular sludge and activated sludge processes. Appl. Microbiol. Biotechnol. 108 (1), 334.
European Environment Agency (EEA) Glossary. https://www.eea.europa.eu/help/glossary/eea-glossary/biota.
Gabor, E.M., de Vries, E.J., Janssen, D.B., 2003. Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. FEMS Microbiol. Ecol. 44 (2), 153–163.
Han, X., Jin, Y., Yu, J., 2022. Rapid formation of aerobic granular sludge by bioaugmentation technology: A review. Chem. Eng. J. 437, 134971.
Harris, P.T., 2020. Chapter 4 - Biogeography, benthic ecology and habitat classification schemes. In: Harris, P.T. (Ed.), Seafloor Geomorphology as Benthic Habitat, Second Edition. Elsevier, E. Baker, pp. 63–96.
Huang, P.W., Laborde, D., Land, V.R., Matson, D.O., Smith, A.W., Jiang, X., 2000. Concentration and detection of caliciviruses in water samples by reverse transcription-PCR. Appl. Environ. Microbiol. 66 (10), 4383–4388.
Ji, Y., Wang, C., Fan, X., Liu, J., Wang, K., Xu, X., Zhu, L., 2025. Revealing the structural stability of aerobic granular sludge (AGS) under salinity condition from a thermodynamic perspective. ACS ES&T Water 5 (6), 3403–3414.
Legendre, P., Legendre, L. 1998. Chapter 7 - Ecological resemblance. Numerical Ecology: Developments in Environmental Modelling 20, 265–335.
Li, J., Ma, L., Wei, S., Horn, H., 2013. Aerobic granules dwelling vorticella and rotifers in an SBR fed with domestic wastewater. Sep. Purif. Technol. 110, 127–131.

McMurdie, P.J., Holmes, S., 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8 (4), e61217.
Menzel, P., Ng, K.L., Krogh, A., 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat. Commun. 7, 11257.
Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., Hingamp, P., Goto, S., Ogata, H., 2016. Linking virus genomes with host taxonomy. Viruses 8 (3), 66.
Mihaylova-Garnizova, R., Davidova, S., Hodzhev, Y., Satchanska, G., 2024. Antimicrobial peptides derived from bacteria: classification, sources, and mechanism of action against multidrug-resistant bacteria. Int. J. Mol. Sci. 25 (19).
NCBI BioSample: SAMN05004214. 2016. https://www.ncbi.nlm.nih.gov/biosample?term=txid1866933[orgn]%20AND%20%22isolate%20ADurb+BinA028%22[All%20Fields].
NCBI proteins database, proteins of the organism with taxonomy ID: 1866933. 2025. https://www.ncbi.nlm.nih.gov/protein/?term=txid1866933[Organism:noexp].
Nieuwenhuijse, D.F., Oude Munnink, B.B., Phan, M.V.T., Hendriksen, R.S., Bego, A., Rees, C., Neilson, E.H., Coventry, K., Collignon, P., Allerberger, F., Rahube, T.O., Oliveira, G., Ivanov, I., Sopheak, T., Vuthy, Y., Yost, C.K., Tabo, D.-A., Cuadros-Orellana, S., Ke, C., Zheng, H., Baisheng, L., Jiao, X., Donado-Godoy, P., Coulibaly, K.J., Hrenovic, J., Jergović, M., Karpíšková, R., Elsborg, B., Legesse, M., Eguale, T., Heikinheimo, A., Villacis, J.E., Sanneh, B., Malania, L., Nitsche, A., Brinkmann, A., Saba, C.K.S., Kocsis, B., Solymosi, N., Thorsteinsdottir, T.R., Hatha, A.M., Alebouyeh, M., Morris, D., O'Connor, L., Cormican, M., Moran-Gilad, J., Battisti, A., Alba, P., Shakenova, Z., Kiiyukia, C., Ngeno, E., Raka, L., Bērziņš, A., Avsejenko, J., Bartkevics, V., Penny, C., Rajandas, H., Parimannan, S., Haber, M.V., Pal, P., Schmitt, H., VanPassel, M., van de Schans, M.G.M., Zuidema, T., Jeunen, G.-J., Gemmell, N., Fashae, K., Wester, A.L., Holmstad, R., Hasan, R., Shakoor, S., Rojas, M.L.Z., Wasyl, D., Bosevska, G., Kochubovski, M., Radu, C., Gassama, A., Radosavljevic, V., Tay, M.Y.F., Zuniga-Montanez, R., Wuertz, S., Gavačová, D., Trkov, M., Keddy, K., Esterhuyse, K., Cerdà-Cuéllar, M., Pathirage, S., Larsson, D.G.J., Norrgren, L., Örn, S., Van der Heijden, T., Kumburu, H.H., de RodaHusman, A.M., Njanpop-Lafourcade, B.-M., Bidjada, P., Nikiema-Pessinaba, S. C., Levent, B., Meschke, J.S., Beck, N.K., Van Dang, C., et al., 2020. Setting a baseline for global urban virome surveillance in sewage. Sci. Rep. 10 (1), 13748.
Orhon, D., 2015. Evolution of the activated sludge process: the first 50 years. 90(4), 608–640.
Osman, R.W. 2015. The Intermediate Disturbance Hypothesis. in: Encyclopedia of Ecology (Second Edition), (Ed.) B. Fath, Elsevier. Oxford, pp. 441–450.
Pronk, M., van Dijk, E.J.H., van Loosdrecht, M.C.M. 2023. Aerobic granular sludge. in: Biological Wastewater Treatment: Principles, Modelling and Design, (Eds.) G. Chen, M.C.M. van Loosdrecht, G.A. Ekama, D. Brdjanovic, IWA Publishing, pp. 0.
Rivas-Castillo, A.M., Garcia-Barrera, A.A., Garrido-Hernandez, A., Martinez-Valdez, F.J., Cruz-Romero, M.S., Quezada-Cruz, M., 2022. Ciliated peritrichous protozoa in a tezontle-packed sequencing batch reactor as potential indicators of water quality. Pol. J. Microbiol. 71 (4), 539–551.
Roth, M.S., 2014. The engine of the reef: Photobiology of the coral-algal symbiosis. Front. Microbiol. 5, 422.
Rueda-Ramírez, D., Palevsky, E., Ruess, L., 2023. Soil Nematodes as a Means of Conservation of Soil Predatory Mites for Biocontrol. 13(1), 32.
Schoch, C.L., 2020. NCBI Taxonomy: A comprehensive update on curation, resources and tools.
Sgarbi, L.F., Bini, L.M., Heino, J., Jyrkänkallio-Mikkola, J., Landeiro, V.L., Santos, E.P., Schneck, F., Siqueira, T., Soininen, J., Tolonen, K.T., Melo, A.S., 2020. Sampling effort and information quality provided by rare and common species in estimating assemblage structure. Ecol. Ind. 110, 105937.
Sharaf, A., Guo, B., Liu, Y., 2019. Impact of the filamentous fungi overgrowth on the aerobic granular sludge process. Bioresour. Technol. Rep. 7, 100272.
Szóstak, N., Szymanek, A., Havránek, J., Tomela, K., Rakoczy, M., Samelak-Czajka, A., Schmidt, M., Figlerowicz, M., Majta, J., Milanowska-Zabel, K., Handschuh, L., Philips, A., 2022. The standardisation of the approach to metagenomic human gut analysis: From sample collection to microbiome profiling. Sci. Rep. 12 (1), 8470.
Wang, Y., Gao, H., 2021. Integrated omics analyses reveal differential gene expression and potential for cooperation between denitrifying polyphosphate and glycogen accumulating organisms. Environ. Microbiol. 23 (6), 3274–3293.
Wang, Z., Kadouri, D.E., Wu, M., 2011. Genomic insights into an obligate epibiotic bacterial predator: Micavibrio aeruginosavorus ARL-13. BMC Genomics 12 (1), 453.
Weber, S.D., Ludwig, W., Schleifer, K.H., Fried, J., 2007. Microbial composition and structure of aerobic granular sewage biofilms. Appl. Environ. Microbiol. 73 (19), 6233–6240.
Wilén, B.M., Liébana, R., Persson, F., Modin, O., Hermansson, M., 2018. The mechanisms of granulation of activated sludge in wastewater treatment, its optimization, and impact on effluent quality. Appl. Microbiol. Biotechnol. 102 (12), 5005–5020.
Winkler, M.K.H., Kleerebezem, R., Khunjar, W.O., de Bruin, B., van Loosdrecht, M.C.M., 2012. Evaluating the solid retention time of bacteria in flocculent and granular sludge. Water Res. 46 (16), 4973–4980.
Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. Genome Biol. 20 (1), 257.
Yang, Q., Zhao, H., Du, B., 2017. Bacteria and bacteriophage communities in bulking and non-bulking activated sludge in full-scale municipal wastewater treatment systems. Biochem. Eng. J. 119, 101–111.
Yuan, J., Deng, X., Xie, X., Chen, L., Wei, C., Feng, C., Qiu, G., 2024. Blind spots of universal primers and specific FISH probes for functional microbe and community characterization in EBPR systems. ISME Commun. 4 (1), ycae011.
Zhang, Y., Qiu, X., Luo, J., Li, H., How, S.W., Wu, D., He, J., Cheng, Z., Gao, Y., Lu, H., 2024. A review of the phosphorus removal of polyphosphate-accumulating organisms in natural and engineered systems. Sci. Total Environ. 912, 169103.
Zhao, H., Yang, M., Fan, X., Gui, Q., Yi, H., Tong, Y., Xiao, W., 2024. A Metagenomic Investigation of Potential Health Risks and Element Cycling Functions of Bacteria and Viruses in Wastewater Treatment Plants. 16(4), 535.
Zhou, J., Liu, W., Deng, Y., Jiang, Y.-H., Xue, K., He, Z., Van Nostrand Joy, D., Wu, L., Yang, Y., Wang, A., 2013. Stochastic assembly leads to alternative communities with distinct functions in a bioreactor microbial community. MBio 4 (2). https://doi.org/10.1128/mbio.00584-12.

Glossary

AS: activated sludge
AGS: aerobic granular sludge
WWTP: wastewater treatment plant
PAO: phosphate-accumulating organisms
GAO: glycogen-accumulating organisms
PHA: poly-hydroxy-alkanoates
PCoA: principal coordinate analysis
SBR: sequencing batch reactor
CFR: continuous flow reactor