


Transcriptional profiling elucidates biofilm functionality in the dynamic environment of an enhanced biological phosphorus removal reactor

Didrik Villard^a, Lars Snipen^b, Knut Rudi^a, Sverre Branders^a, Torgeir Saltnes^{c,e}, Sondre Eikås^d and Wenche Johansen ^{a,*}

^a Faculty for Applied Ecology, Agriculture and Biotechnology, Inland Norway University of Applied Sciences, Hamar, Norway

^b Faculty of Chemistry, Biotechnology and Food Science, University of Life Sciences, Ås, Norway

^c Hias How2O, Sandvikavegen 136, Ottestad, Hamar 2312, Norway

^d Hias IKS, Sandvikavegen 136, Ottestad, Hamar 2312, Norway

^e Present address: Envidan AS, Storgata 10B, Oslo 0155, Norway

*Corresponding author. E-mail: wenche.johansen@inn.no

 WJ, 0000-0002-9240-586X

ABSTRACT

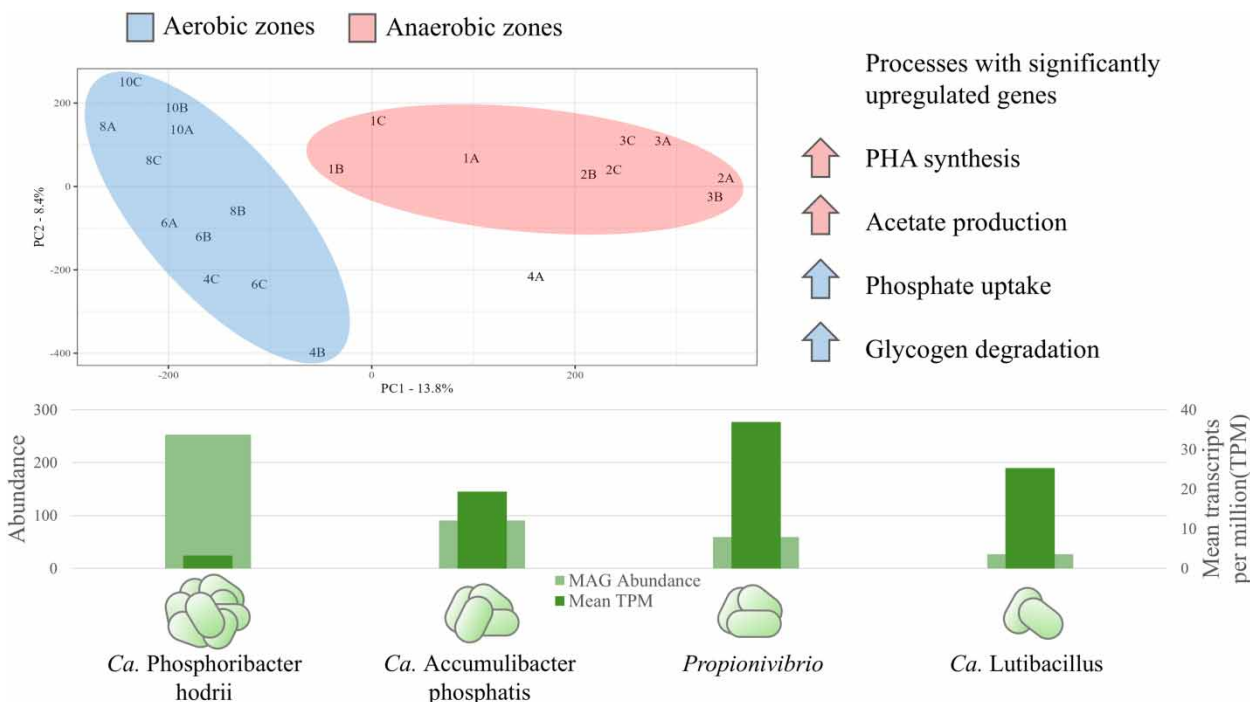
Recently, biofilms, complex and dynamic structures of microorganisms, have been applied to enhanced biological phosphorus removal (EBPR), a wastewater treatment configuration dependent on cyclic shifts between anaerobic and aerobic conditions. In this study, comparative metagenomics and metatranscriptomics were performed on biofilms collected from seven sites of a moving-bed-biofilm-reactor-based EBPR process. The aim was to examine the functional ecology of phosphorus-accumulating biofilms throughout a single EBPR cycle. Taxonomic profiling revealed high microbial diversity, stable throughout the EBPR cycle. The dominant phosphorus-accumulating organisms (PAOs) were identified as *Candidatus* accumulibacter, *Candidatus* phosphoribacter, and *Candidatus* lutibacillus. However, these did not show the highest transcriptional activities. Propionivibrio, a glycogen-accumulating organism, was the most transcriptionally active. Comparative analysis of biofilms from different EBPR stages showed a progressive change in metatranscriptome composition, correlating with nutrient removal. Analysis of differentially expressed genes in abundant PAOs revealed key genes associated with the uptake of phosphorus, degradation of glycogen, biosynthesis of polyhydroxyalkanoates, and acetate production. In conclusion, this study reveals that biofilms possess the capability to adapt to environmental fluctuations primarily through alterations in microbial gene expression activity and subsequent metabolic modulation, and dominant taxa may not necessarily exhibit the highest transcriptional activity in complex microbial communities.

Key words: biofilm, enhanced biological phosphorus removal (EBPR), functional ecology, metatranscriptomics, polyphosphate-accumulating metabolism

HIGHLIGHTS

- The microbial composition of the biofilm is stable throughout a complete EBPR cycle.
- Metatranscriptome dynamics correlate with shifts in the nutrient profile of the wastewater.
- PAOs and GAOs are the most abundant and transcriptionally active taxa in the EBPR biomass.
- The dominating taxon exhibit lower transcriptional activity compared to less abundant taxa.

GRAPHICAL ABSTRACT



INTRODUCTION

Biofilms are complex and dynamic assemblies of microbial cells attached to abiotic and biotic surfaces, encased in a matrix of biopolymers and water, called the extracellular polymeric substances (EPSs). Approximately 80% of the prokaryotic cells on the Earth's surface are estimated to reside in biofilms, establishing biofilms as the most natural microbial lifeform (Flemming & Wuertz 2019). Biofilms play significant roles in diverse natural environments, such as rivers and streams, where they are essential drivers to biogeochemical cycling of nutrients within the ecosystem (Battin *et al.* 2016). Biofilms are also extensively exploited for diverse biotechnological applications, for example, for wastewater treatment, where environmental pollutants are removed or transformed into non-toxic chemicals (Edel *et al.* 2019). Structurally, biofilms are highly complex due to their diverse microbial composition, intricate metabolic interactions, and spatial organization (Flemming *et al.* 2023). Biofilms are also highly dynamic, as they must respond to changes in the environment such as temperature, the availability of nutrients and oxygen. By adjusting their composition and metabolic activity, biofilm can adapt to environmental changes. Understanding these responses is important to improve their efficiency and utilization for biotechnological applications. For example, understanding biofilm physiology is crucial for effectively implementing biofilms in wastewater treatment, as it enables optimized nutrient removal and reduces the risk of eutrophication in aquatic environments. To study the functional dynamics of biofilms and other multispecies communities, multi-omics approaches, which integrate data from genomics, transcriptomics, proteomics, and metabolomics have emerged as valuable tools (McDaniel *et al.* 2021b; Chen *et al.* 2022). These approaches allow for the comprehensive analyses of the genetic potential, gene expression patterns, protein abundances, and metabolic activities of individual species within the community, providing insights into their interactions and ecological roles.

The Hias Process is a novel biotechnological invention utilizing moving-bed-biofilm-reactor (MBBR) technology in combination with enhanced biological phosphorus removal (EBPR) to clean phosphorus (P) from municipal wastewater (Saltnes *et al.* 2017). EBPR leverages varying environmental conditions to promote microbial processes for P removal by cycling between anaerobic carbon-feasting and aerobic carbon-famine conditions. This cycling promotes the proliferation of microorganisms capable of storing large amounts of P as polyphosphate (poly-P) within their cells, known as the polyphosphate-accumulating organisms (PAOs) (Mino *et al.* 1998). The Hias Process consists of a single continuous reactor divided into ten stages (zones), the first three being anaerobic and the following seven oxic (Figure 1).

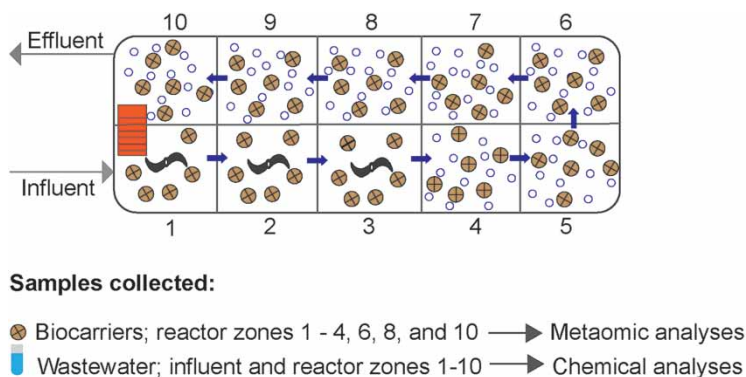


Figure 1 | Schematic diagram of the Hias Process, an enhanced biological phosphorus removal (EBPR) wastewater treatment system. The Hias EBPR process is designed as a moving-bed-biofilm-reactor (MBBR), comprising a single reactor divided into ten zones. The initial three zones, following the influent introduction, operate under anoxic conditions, while the subsequent seven zones are aerobic with aeration (shown as blue circles). A conveyor belt (shown in orange) facilitates the movement of carriers between the tenth and first zones. Biocarriers were sampled from each anaerobic zone and every alternate aerobic zone, while wastewater samples were collected from all zones, including the influent water.

After an initial primary filtration of particulates, the wastewater enters the first reactor zone in a mechanically agitated anaerobic part where it is mixed with biocarriers harboring biofilm. Wastewater and biofilm carriers flow by gravity through the remaining two anaerobic and seven oxic zones in a plug-flow pattern in a cycle of between 4 and 12 h, depending on influent volume. In the last reactor zone, treated wastewater and sloughed-off biofilm leave the reactor, while the biofilms are conveyed dry back to the start of the reactor in a continuous process. The Hias Process is currently operating on full scale at the 200,000 PE (population-equivalent) wastewater treatment plant (WWTP) located in Ottestad, Norway. The process has proven to be highly resilient for fluctuation in external carbon-load, performing stable P removal over time (Saltnes *et al.* 2017). The Hias Process eliminates the use of chemicals for phosphorous removal with a considerably lower reactor volume compared to similar processes. These attributes make it a potentially important solution for emerging countries facing eutrophication challenges.

Several PAO genera relevant for WWTPs has been described and are recently thoroughly reviewed (Zhao *et al.* 2022; Ruiz-Haddad *et al.* 2024). Of these, the most widely characterized is *Candidatus Accumulibacter*, for which a consensus metabolic model has been established (reviewed by He & McMahon 2011). In short, during anaerobic carbon-feasting conditions, *Ca. Accumulibacter* assimilates volatile fatty acids (VFAs) such as acetate, propionate, and butyrate into polyhydroxyalkanoate (PHA) which are stored intracellularly. Energy (ATP) for PHA biosynthesis is provided by glycogen degradation and poly-P hydrolysis, while reducing equivalents (NAD(P)H) is provided by the glycolytic consumption of glucose and/or the anaerobic operation of the tricarboxylic acid cycle (TCA) (reviewed by Zhou *et al.* 2010). Under aerobic carbon-famine conditions, stored PHA is oxidized providing ATP for biomass growth and to replenish the poly-P and glycogen pools. Some species are capable of denitrification, using nitrite (NO_2^-) and nitrate (NO_3^-) as terminal electron acceptors for anoxic P uptake (Flowers *et al.* 2009). Although common to full-scale WWTPs globally, *Ca. Accumulibacter* is often found to be less abundant when compared to members belonging to the actinobacterial *Tetrasphaera* group (Nielsen *et al.* 2019). Previous phylogenetic analysis based on 16S rDNA sequence information divided the *Tetrasphaera* lineage into three clades, 1, 2 and 3 (Nguyen *et al.* 2011). However, more recent phylogenetic analyses based on metagenome-assembled genomes (MAGs) and the full-length 16S rRNA gene have resulted in the reclassification of the *Tetrasphaera* group into at least five different genera (reviewed by Ruiz-Haddad *et al.* 2024). The *Tetrasphaera* clade 3 group, previously represented by environmental (uncultured) sequences only, is now divided into two new proposed genera, namely *Candidatus Phosphoribacter* and *Candidatus Lutibacillus*, of which the former is reported as the most dominant PAO in EBPR-based WWTPs worldwide (Singleton *et al.* 2022). Both genera are shown experimentally to exhibit dynamic cycling of intracellular poly-P during anaerobic-aerobic alternations (Singleton *et al.* 2022). Recently, a new *Tetrasphaera*-related clade, clade 4, represented by the genus *Austwickia*, was proposed (Yu *et al.* 2023). However, experimental data showing poly-P-accumulating metabolism (PAM) or cycling of poly-P in this genus is lacking. The metabolism of the genera formally assigned to the *Tetrasphaera* lineage is distinctly different from that of *Ca. Accumulibacter* (Zhao *et al.* 2022). During anaerobic conditions

glucose (and other sugars) are fermented to acetate, lactate, succinate, and alanine with the simultaneous biosynthesis of glycogen as a storage polymer using the energy derived from poly-P hydrolysis and substrate fermentation (Kristiansen *et al.* 2013). Under aerobic conditions, with low access to external nutrients, the stored glycogen is degraded, providing energy for growth and to replenish the poly-P pool. In addition to glycogen, free amino acids have been identified as both a carbon source anaerobically and as a storage product to be consumed in the aerobic phase for uptake of P (Nguyen *et al.* 2015). It has been suggested that *Ca. Accumulibacter* and the former *Tetrasphaera* lineage play cooperative roles in EBPR (Kristiansen *et al.* 2013; Fernando *et al.* 2019; Nielsen *et al.* 2019), specifically that *Tetrasphaera* provide fermentations products, such as acetate, lactate, succinate and amino acids, as substrates for *Ca. Accumulibacter*. However, in a recent study it was shown that a simple synergistic relationship between *Ca. Accumulibacter* and *Tetrasphaera* may not exist as their interactions are predominantly influenced by the specific fermentation products released by *Tetrasphaera* (Chen *et al.* 2023).

Here, the authors have used metagenomic and metatranscriptomic approaches to analyze the microbial composition and the transcriptional activities of biofilms in an EBPR-based wastewater treatment system throughout one reactor cycle. The aim was to investigate how the biofilm responds to anaerobic/aerobic cycling conditions to facilitate P removal, and to identify prevalent and transcriptionally active microorganisms involved in the Hias Process. This information is important to gain insight into the functional ecology of P-accumulating biofilms.

METHODS

A chart diagram showing the experimental design of the study is provided in Figure 2.

Sample collection and processing

Samples were collected on January 27, 2022, from a 10,000-population-equivalent (PE) demonstration plant designed for the Hias Process (EBPR and MBBR) at the Hias WWTP in Ottestad, Norway. Prior to sample collection, the wastewater inlet volume was fixed at 15 L/s for 24 consecutive hours. On the day of sampling the temperature of the influent wastewater was 9 °C. Three to five biofilm carriers were collected at random from the center of each of reactor zones 1, 2, 3, 4, 6, 8, 10 and immediately frozen in liquid nitrogen (N₂) to preserve nucleic acids (NAs). Simultaneously, wastewater samples were collected from the inlet and from each of the ten reactor zones and further processed for chemical analysis. A scheme of the Hias Process and sampling sites are provided in Figure 1.

Chemical analysis

All samples were filtered through 1 µm fiberglass filter and analyzed for dissolved phosphorus (PO₄³⁻), dissolved chemical oxygen demand (SCOD), ammonium (NH₄⁺), nitrate (NO₃⁻) and nitrite NO₂⁻ with a NOVA Spectroquant 60 spectrophotometer according to standard procedures.

Extraction of DNA and RNA from biofilm

Three biofilm carriers from each of the reactor zones 1, 2, 3, 4, 6, 8 and 10 (a total of 21 samples) were processed individually and grinded to a fine powder using liquid N₂. DNA and total RNA was isolated separately from each of approximately 200 mg of grinded powder. DNA was isolated using the DNeasy PowerSoil Pro kit (Qiagen, Germany) according to the manufacturer's instructions and eluted in 50 µL Buffer C6. For RNA isolation, grinded biofilm powder was transferred to PowerSoil Pro bead tubes (Qiagen, Germany) and immediately added 1 mL of chilled (4 °C) TRIzol reagent (Invitrogen, US). The samples were processed on Beadbeater 96 (BioSpec Products, US) for 3 × 40 s with 1 min incubation on ice between each run. After incubation on ice for 5 min, the suspension was carefully transferred to an Eppendorf tube and added 300 µL of chloroform and 30 µL of RNase-free H₂O, thoroughly mixed by vortexing and incubated for 3 min on ice. The samples were centrifuged at 12,000 × g for 15 min at 4 °C and the upper aqueous phase (500 µL) was transferred to a new Eppendorf tube. An equal amount of freshly diluted 70% ethanol was added to each sample, and the sample transferred to the GeneJET RNA Purification Kit (Thermo Fisher Scientific, USA) spin column, which was further used to purify and collect total RNA according to the manufacturer's instructions. Total RNA was eluted with 50 µL of RNase-free H₂O. NAs were quantified using NanoDrop or Qubit assay, and the integrity was analyzed using 1% agarose in 0.5X TBE (Tris-borate-EDTA) gels. Samples were stored at -80 °C until further processing.

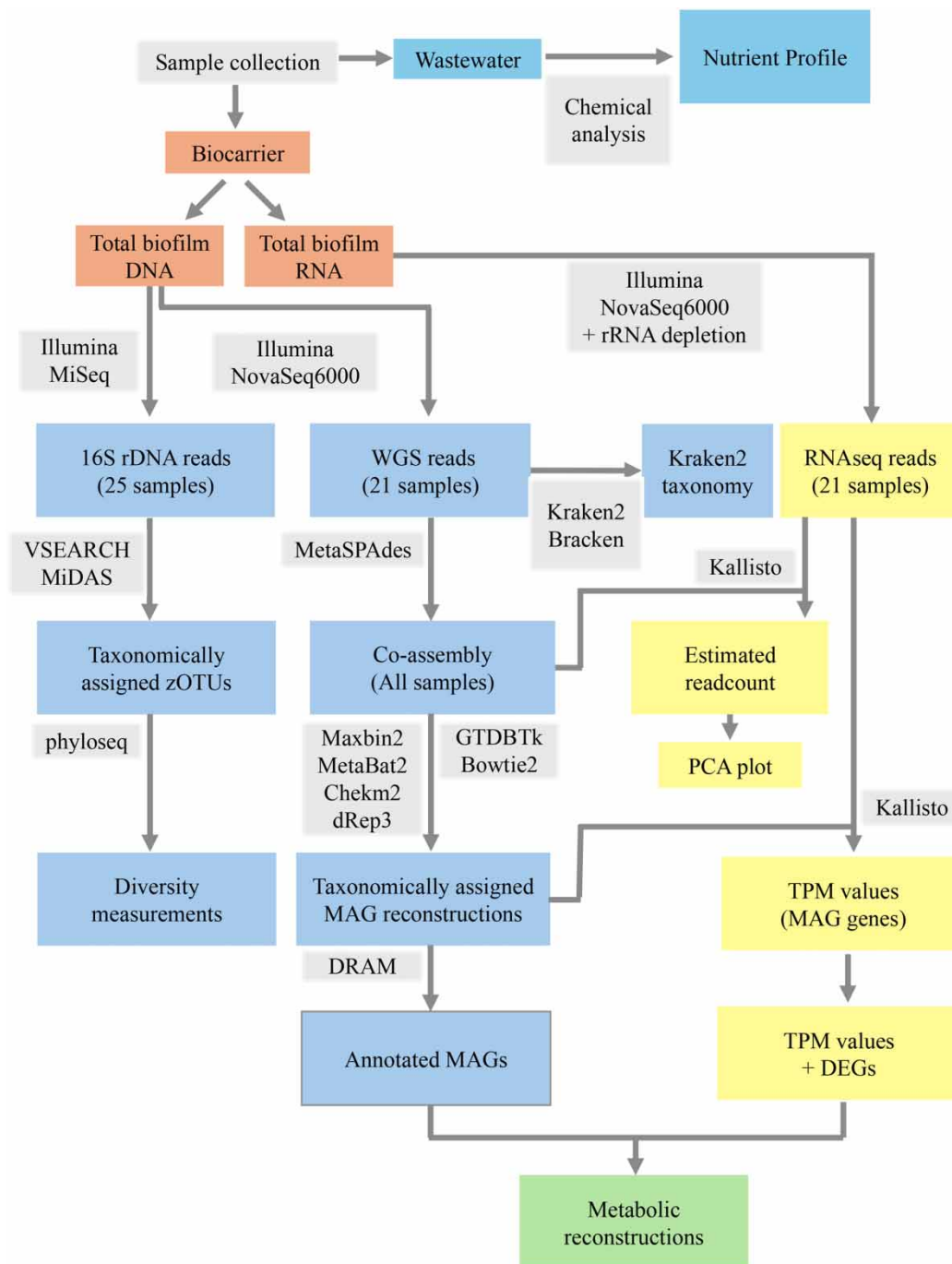


Figure 2 | Flowchart providing an overview of the analysis performed. Wastewater samples (light blue) and biofilm material (orange) were both analyzed. From biofilm, total RNA (yellow) and total DNA (blue) was isolated and sequenced, and analyzed using various bioinformatics tools (grey). Metabolic reconstructions (green) were performed using both DNA and RNA sequencing data. Abbreviations: DEG, differentially expressed gene; MAG, Metagenome-assembled genome; PCA, principal component analysis; TPM, transcripts per million; WGS, whole-genome sequencing; and zOTU, zero-distance operational taxonomic units.

16S rRNA metagenomics

For 16S rRNA gene sequencing, the variable V3-V4 region was amplified using ~10 ng DNA as template in a 1x HOT FIRE-Pol® Blend Master Mix Ready to Load (Solis Biodyne, Estonia) reaction mixture containing 0.2 μM of each of the universal

primer pair 341F and 806R (Takahashi *et al.* 2014). Amplification was performed using a 2720 Thermal Cycler (Applied Biosystem, USA) and the following cycling parameters: 95 °C for 15 min, 25 cycles at 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 45 s, and a final extension step at 72 °C for 7 min. Polymerase chain reaction (PCR) products were purified using Sera-Mag Speedbeads (Cytiva, USA) according to the manufacturer's instructions. Index PCR was performed using the same components and reaction conditions as described for V3-V4 16S rDNA amplification except that Illumina primers designed for dual indexing were used. Amplification was conducted on a 2720 Thermal Cycler with the following cycling parameters: 95 °C for 5 min, then 10 cycles of 95 °C for 30 s, 55 °C for 1 min, 72 °C for 45 s, followed by a final extension step at 72 °C for 7 min. The resulting PCR products were quantified using the hsDNA Qubit kit (Thermo Fisher Scientific, USA) according to manufacturer's protocol. An equal amount of amplified DNA from each sample was transferred to a new tube to create a pooled library, which was subsequently purified using the Sera-Mag Speedbeads protocol. Sequencing was performed at the Norwegian Sequencing Centre (NSC) (Lyle *et al.* 2012) using the Illumina MiSeq platform and 300 base pair (bp) paired end reads.

zOTU generation and diversity metrics for 16S sequencing data

The 16S rRNA gene sequencing data were clustered into zero-radius Operational Taxonomic Units (zOTU) using the UNOISE algorithm (Edgar 2016b) implemented in the VSEARCH v.2.21.1 software (Rognes *et al.* 2016) and these were taxonomically assigned using the SINTAX algorithm (Edgar 2016a) and MiDAS 4.8.1 database (Dueholm *et al.* 2022). Diversity metrics were calculated using the R package phyloseq v. 1.46.0.

Library preparation and sequencing for whole-genome sequencing (WGS) metagenomics and metatranscriptomics

Library preparation and sequencing were performed at Novogene (Cambridge, UK). The Novogene NGS DNA library prep set (Cat. #PT004) and the Novogene NGS RNA library prep set (Cat. # PT042) were used to construct the metagenome and metatranscriptome libraries, respectively. For generating the metatranscriptomes, eukaryotic and prokaryotic ribosomal RNA (rRNA) were first depleted from the total RNA using the Illumina Ribo-Zero Plus rRNA Depletion Kit (Illumina, USA). Quantification of libraries were analyzed with Qubit and real-time PCR, and bioanalyzer was used for size distribution detection. Libraries were pooled and sequenced using the Illumina NovaSeq6000 platform with 150 bp paired end reads.

Co-assembly and reconstruction of metagenome-assembled genomes (MAGs)

Illumina NovaSeq metagenome reads were filtered using BBduk and overlapping read-pairs were merged using BBmerge from the Bmap software v.39.01 (Bushnell *et al.* 2017). The filtered reads were then taxonomically assigned using Kraken2 (Wood *et al.* 2019) and Bracken (Lu *et al.* 2017) with the GTDB release 207 v. 2 (Parks *et al.* 2018). Metagenome-assembled genomes (MAGs) were reconstructed using a co-assembly approach, in which a subset of randomly selected reads (15 million pairs and 3 million merged reads) from each of the 21 biofilm samples across all reactor zones were pooled into a super sample. The rationale for using a co-assembly approach was the uncovered DNA sequence similarity between the samples from the different reactor zones and that a co-assembly approach could potentially recover MAGs that displayed too low abundance to be assembled from individual samples. Co-assembly was performed using metaSPADes v.3.15.5 (Nurk *et al.* 2017) and the resulting contigs were binned using Maxbin2 v2.2.7 (Wu *et al.* 2016) and MetaBat2 v.2.15 (Kang *et al.* 2019). Bin quality was assessed by checkm2 v.1.0.1 (Chklovski *et al.* 2023) and de-replicated by dRep3 v3.4.0 (Olm *et al.* 2017). The resulting MAGs were assigned to the GTDB taxonomy using GTDB-Tk v.2.2.0 (Chaumeil *et al.* 2019). To quantify the abundance of each reconstructed MAGs in each reactor zone, a subset of 10 million read-pairs from each sample was mapped against the MAGs using bowtie2 (Langmead & Salzberg 2012). The coverage (abundance) of each MAG and for each sample was computed using samtools v.1.9 (Danecek *et al.* 2021) and the `jgi_summarize_bam_contig_depths` from MetaBat2. MAG abundancies were reported as the average of the three biological replicates for each reactor zone.

Gene prediction and annotation of metagenomes, and metatranscriptome mapping

Putative coding genes in the co-assembled metagenome was predicted using prodigal (Hyatt *et al.* 2010). For the reconstructed MAGs, putative genes were predicted and annotated using the Distilled and Refined Annotation of Metabolism tool (DRAM) (Shaffer *et al.* 2020). Kallisto v. 0.48.0 (Bray *et al.* 2016) was employed to map the Illumina NovaSeq RNAseq reads from each sample to the predicted genes of the MAGs and co-assembly, resulting in estimated counts and TPM (transcript per million reads) values for each gene in each sample.

Principal component analysis of metatranscriptome mapped to co-assembly

Estimated read counts from Kallisto were variance stabilized using DESeq2, and a Principal Component Analysis (PCA) was conducted using the stats package. The 100 most extreme loadings from the resulting PC1 were extracted and annotated searching the tax4fun2 database containing KEGG orthologs using the blastp settings hosted by Diamond v. 2.1.6 (Buchfink *et al.* 2021).

Differential expressed genes and expression threshold

To identify MAG-encoding genes exhibiting differential expression between the anaerobic and aerobic zones a Wilcoxon test was conducted. Mean TPM values per gene and per MAG, for both the anaerobic and aerobic zones were calculated and log₂fold-change between them for each mean value was determined. For a gene to be collected as a differentially expressed gene (DEG), the adjusted *p*-value (controlling the false discovery rate) had to be <0.05 and the log fold-change at least 1 or -1 between anaerobic/aerobic conditions. Genes with very low expression values (<0.01) in either aerobic or anaerobic zones were also discarded to avoid artefacts from comparing very weak signals. To establish a TPM threshold for gene expression the lower 0.5% quantile of all mean TPM values from upregulated zones for each DEG was identified and applied to all non-DEG transcripts.

Metabolic reconstructions

Metabolic reconstructions of selected MAGs were performed manually based on the DRAM annotations. In cases where there were discrepancies between DRAM hits from KEGG orthologs and Pfam, BLAST searches were employed to assist in the annotation process. Initially, genes and their corresponding transcripts encoding enzymes known to be pertinent for biochemical pathways relevant to EBPR physiology were compiled. Subsequently, DEGs encoding enzymes involved in these pathways were gathered, along with the genes and transcripts necessary for the complete metabolic pathways associated with the DEGs in question. The models were then constructed using all collected genes and transcripts for the MAGs.

RESULTS AND DISCUSSION

Wastewater nutrient profile throughout one reactor cycle of the Hias process

To be able to correlate biofilm functionality to microbial composition and function, EBPR process performance was investigated by collecting the nutrient profile of the wastewater at the time of biofilm sampling (Figure 3).

After an initial increase in P from an inlet concentration of 6.6–42.6 mg L⁻¹ in reactor zone 3 (corresponding to the last anaerobic zone), the concentration of P was gradually decreased by 99.1% to 0.06 mg L⁻¹ in reactor zone 10 (corresponding to the last aerobic zone). Simultaneously, the total soluble chemical oxygen demand (SCOD) content in the wastewater was reduced by 77% from 495 mg L⁻¹ in the influent to 116 mg L⁻¹ in reactor zone 10. The reduction in SCOD concentration was rapid in the three anaerobic zones where <65% of the total SCOD lost, was removed. These observations are consistent with a highly effective EBPR process under cyclic anaerobic and aerobic conditions as described elsewhere (Nielsen *et al.* 2019). In the aerobic phase of the process, ammonium (NH₄⁺) was reduced from 76.5 to 57.5 mg L⁻¹ without significant increase in nitrite (NO₂⁻) and nitrate (NO₃⁻) accumulation, indicating simultaneous nitrification and denitrification (SND) in the reactor at the time of sampling (Figure 3). Also, a small reduction in the concentration of NO₂⁻ and NO₃⁻ was observed in the anaerobic zones, indicating heterotrophic growth using these nitrogenous compounds as terminal electron acceptors.

The microbial composition of the Hias biofilm is stable throughout one reactor cycle

The dynamics of the Hias biofilm community structure throughout one reactor cycle was investigated by analyzing the microbial composition of biofilms extracted from each of the different reactor zones, using both 16S rRNA gene sequencing and whole-genome shotgun (WGS) metagenomics. In total, 2236 zero-radius operational taxonomic units (zOTU) were identified, 80% of which were taxonomically assigned using the MIDAS 4 database (Dueholm *et al.* 2022). The biofilms demonstrated a high degree of biodiversity, with zOTU richness in individual samples ranging from 1822 to 2071 (Supplementary material 1, Figure S1(a)) and with Simpson's diversity index scores between 0.98 and 0.99 (Supplementary material 1, Figure S1(b)), which could be explained by the continuous supply of microorganisms originating from the incoming wastewater. A previous study investigating microbial communities with approximately 200, 250 and 300 OTUs found that higher biodiversity was associated with faster recovery from environmental perturbations, suggesting that heterogenous biofilms might be more resilient to environmental disturbance (Feng *et al.* 2017). Thus, the comparatively high biodiversity of the

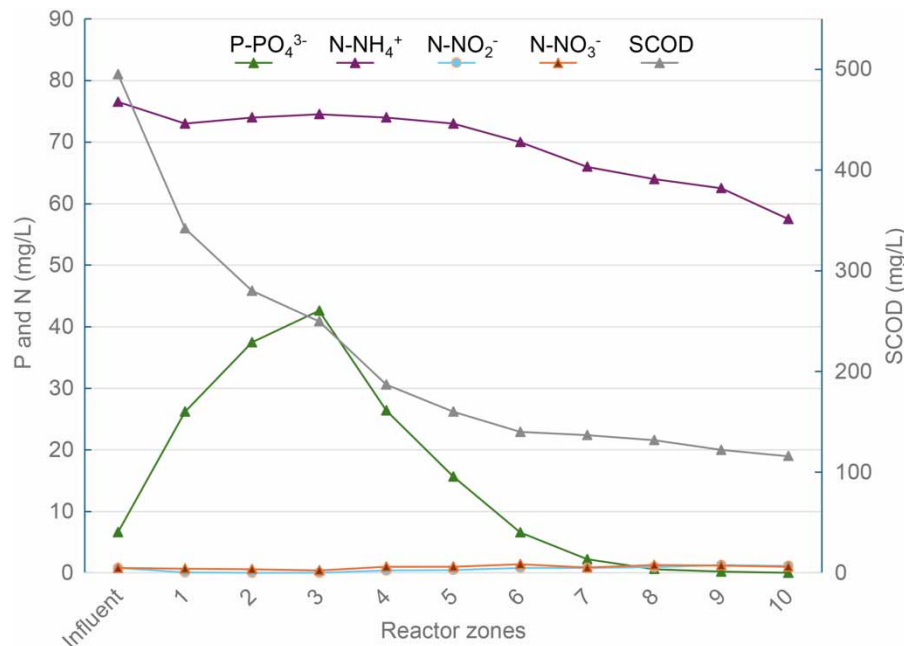


Figure 3 | Nutrient profile of wastewater through one reactor cycle of the Hias Process. The graph shows concentrations of the following nutrients: phosphate (PO_4^{3-}), ammonium (NH_4^+), nitrite (NO_2^-), nitrate (NO_3^-) and soluble carbon oxygen demand (SCOD). Water samples for chemical analysis were collected from each of the reactor zone, as well as the influent wastewater.

Hias biofilm might provide an explanation to the robust stability of the Hias Process in response to fluctuations in inlet wastewater carbon loads over time, as shown in a previous study (Saltnes *et al.* 2017). The microbial community structure of the biofilms was compared using pairwise Wilcoxon tests, revealing no significant differences (p -value > 0.05) in species richness neither between biofilms harvested from the different reactor zones, nor between the anaerobic (reactor zones 1–3) and aerobic (reactor zones 4, 6, 8 and 10) phases of the process. A permutational multivariate analysis of variance analysis (PERMANOVA) based on Bray-Curtis dissimilarities of the relative abundance of zOTUs revealed no significant differences in the microbial composition between biofilms extracted from the different reactor zones (p -value = 0.59). The community structure dynamics were also investigated using the taxonomical profiles of each biofilm sample generated from the analysis of the WGS metagenomic data (Supplementary material 1, Figure S2(a)). Similarly to the 16S rRNA gene sequencing approach, a PERMANOVA analysis using Bray-Curtis distance matrix (based on the pairwise taxonomic profiles) showed no significant differences in microbial community composition between samples extracted from the different reactor zones (p -value = 0.48). However, the most abundant phyla identified using the two different sequencing approaches (WGS and 16S-based metagenomics) differed (compare Figures S2(a) and S2(b), Supplementary material 1). This might be attributed to the fact that only 20% of the WGS reads were taxonomically assigned using Kraken2 and Bracken, as compared to 80% of the 16S-based reads using the SINTAX and the MiDAS 4 database. Thus, to compare the biofilm community composition independently from taxonomical assignments, the mash-distances of the WGS reads between each pair of biofilm samples were calculated (Supplementary material 1, Figure S3). The distance between all biofilm samples was small, only 2–3%, and similar both within biological replicates and between samples extracted from the different reactor zones, with no apparent grouping. Again, a PERMANOVA analysis based on Bray-Curtis dissimilarity using the pairwise mash-distances between samples showed no significant differences in the sequence composition of the biofilm extracted from each of the reactor zones (p -value = 0.20). Finally, metagenome-assembled genomes (MAGs) were also reconstructed, calculating their relative abundance (coverage) in each reactor zones (Supplementary material 2, Table S1), revealing no significant differences in MAG composition and abundance (p -value = 0.06). Taken together, both the 16S-based and WGS metagenomic approaches showed that the microbial composition of the biofilm does not change significantly from one reactor zone to another, suggesting that the Hias biofilm microbiota is stable within one reactor cycle. Thus, the changes in biofilm functionality during one reactor cycle (e.g. P release and uptake, and SCOD reduction) cannot be explained by changes in the microbial composition of the biofilm.

The most abundant biofilm MAGs belong to prominent and putative PAOs and GAOs

MAGs were reconstructed using a co-assembly approach, in which a subset of the WGS reads from each of the 21 biofilm samples across all reactor zones were pooled into a super sample. 180 MAGs of minimum quality (>75% completeness and <25% contamination) were reconstructed (Supplementary material 2, Table S1) and when mapping the WGS reads to the contigs of these MAGs, 37–38% of the reads were aligned, indicating that the reconstructed MAGs accounted for only ~40% of the total biofilm DNA. The taxonomy of the 180 reconstructed MAGs was inferred by GTDB-Tk using GTDB release 207 (Supplementary material 2, Table S1). Among the top six most abundant MAGs, four putative PAOs and two putative glycogen-accumulating organisms (GAOs) were identified. The most abundant MAG (accounting for 13.2% of the total MAG abundancy; Supplementary material 1, Table S2) was metabat.36 (hereafter referred to as MAG1). MAG1 was annotated with the GTDB 207 accession GCA-2748155 sp016707175, which corresponds to *Ca. Phosphoribacter hodrii* in the GTDB 214 release. *Ca. Phosphoribacter hodrii* is a novel species previously assigned to the clade 3 group of *Tetrasphaera* (Singleton *et al.* 2022). Among the top six abundant MAGs, maxbin.015 (MAG4; 4.7%) and maxbin.017 (MAG5; 4.2%) were also identified, both taxonomically assigned to the well-known PAO *Ca. Accumulibacter phosphatis*. The third most abundant MAG, metabat.106 (MAG3; 5.9%) belonged to a currently uncharacterized genus in GTDB, namely JADJTI01 sp016711265. The corresponding NCBI annotation in GTDB 207 classify JADJTI01 sp016711265 to *Austwickia*, a genus recently proposed to represent a new clade (clade 4) of *Tetrasphaera*-PAOs (Yu *et al.* 2023). The remaining two of the top six abundant MAGs, metabat.288 and metabat.396, belonged to the proposed glycogen-accumulating organism (GAO) *Propionivibrio aalborgensis* (MAG2; 10.0%) (Albertsen *et al.* 2016) and the genus *Propionivibrio* (MAG6; 3.1%), respectively. It is generally accepted that GAOs and PAOs compete for the same carbon resources within EBPR, and that the appearance of GAOs, which do not contribute to biological P removal, is associated with decreased overall P removal efficiency and EBPR process performance (Cech & Hartman 1993; Saunders *et al.* 2003). These findings contrast the results of the current study, which shows that a MAG taxonomically assigned as the GAO *P. aalborgensis* is abundantly present in the biofilm, yet the biomass sustains a highly efficient EBPR performance in the Hias Process (Figure 3). Consistent with our study, Albertsen *et al.* (2016) have shown the presence and often abundance of *Propionivibrio* in full-scale EBPR WWTPs.

A discrepancy was identified when comparing the most abundant MAGs and the most abundant zOTUs (Supplementary material 1, Table S2). In the zOTU analysis, members of the genus *Ca. Accumulibacter* were identified as the overwhelming most abundant zOTUs, represented by 15.2% of all the sequences, whereas organisms related to *Tetrasphaera* accounted for only 1.5% of the zOTUs. In contrast, the most abundant MAG identified was related to *Tetrasphaera*, while *Ca. Accumulibacter* was found to be the fourth and fifth most abundant MAG only. As relative zOTUs relies on the quantification of a single fragment of the 16S rRNA gene, the method is inherently susceptible to difference in the gene copy numbers between genera. According to the Ribosomal RNA Database, *rmDB*, (Stoddard *et al.* 2015), *Ca. Accumulibacter* harbors three copies of the gene, in contrast to one copy in *Tetrasphaera*. Thus, relative zOTUs could therefore overestimate the abundancy of *Accumulibacter*, partially explaining the disparity between our analyses. Singleton *et al.* (2022) utilized the V1-V3 variable region of the 16S rRNA gene to assess abundances, revealing *Ca. Phosphoribacter* as the most abundant functional PAO across 847 EBPR-based plants worldwide. Our data underscores the important of considering methodology factors, such as gene copy number in interpreting microbial community data accurately.

Metatranscriptomic analysis reveal a progressive change in the expression of biofilm genes within each reactor zone

Next, the dynamics of the metatranscriptome throughout one EBPR cycle were analyzed. RNAseq reads, generated from biofilms extracted from the different reactor zones, were mapped to a catalogue of approximately 1.6 million genes, predicted from the co-assembled metagenome. A principal component analysis (PCA) of the estimated counts from Kallisto mapping revealed a progressive change in the expression of the 1.6 million genes throughout the process, with distinctly different transcript profiles in the anaerobic (samples 1–3) and aerobic (samples 4, 6, 8 and 10) phases (Figure 4). Notably, in the anaerobic and aerobic clusters, samples from neighboring zones display a gradient along separate principal components, PC1 and PC2, respectively, positioning reactor zone 1 as the midpoint among of all samples. This dynamic trend indicates that distinctly different gene activities are necessary during the anaerobic and aerobic phases of the process, and that there is a gradual change in gene activity from one reactor zone to the next within each of these two phases. However, when biofilms are transferred between anaerobic and aerobic conditions (reactor zones 3 to 4 and 10 to 1), distinctly different clustering of biofilm samples are evident. In fact, the most distant metatranscriptome profiles was observed between samples extracted from

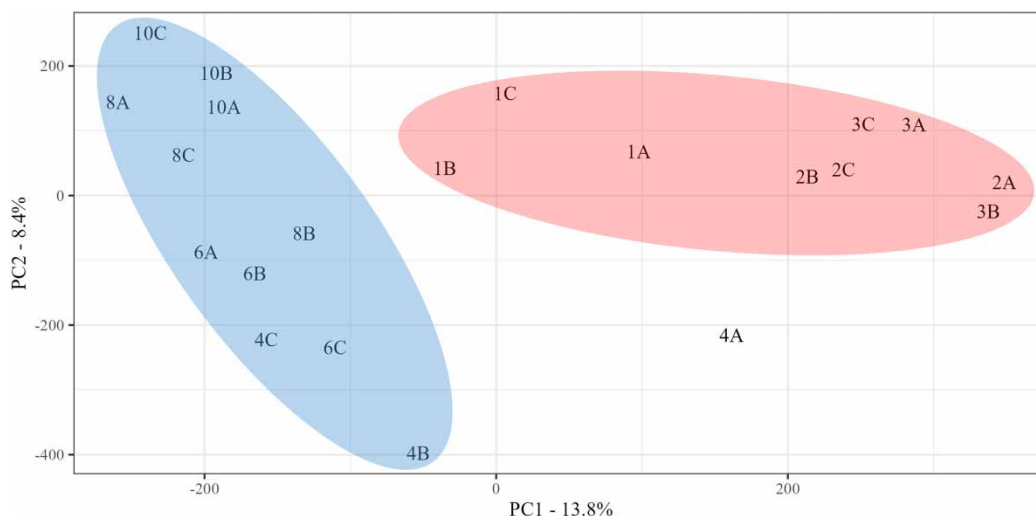


Figure 4 | A principal component analysis (PCA) plot showing the metatranscriptome dynamics across a single EBPR cycle in the Hias Process. This analysis utilized estimated read counts obtained by aligning RNAseq reads to genes predicted from a co-assembly of whole-genome sequencing data, employing Kallisto. The plot illustrates two distinct clusters representing samples collected from anaerobic zones (highlighted in red) and aerobic zones (highlighted in blue). Numerical labels denote the specific reactor zone of sample origin, while alphabetical characters denote biological replicates.

reactor zones 3 and 10, representing the final anaerobic and aerobic reactor zones, respectively. The observed dynamic changes in the metatranscriptome follows the dynamics of the nutrient profiles. The large change in the metatranscriptome from 10 to 1 corresponds to the transfer of biofilms from aerobic conditions with low SCOD in zone 10, to anaerobic conditions with high SCOD in zone 1. Then progressive smaller changes in the metatranscriptome occur within the anaerobic zones, which corresponds to a gradually lower concentration of available carbon (SCOD) from high to low concentrations in zone 1 and 3, respectively. The biofilms then circulate to reactor zone 4 where they become exposed to highly oxygenated conditions, and again a large shift in the metatranscriptome is observed. This large and rapid shift is finally followed by progressive smaller changes in the metatranscriptome within the aerobic zones, correlated with the gradual reduction in the concentration of wastewater P. The metatranscriptomic profiling reveals that the biofilm responds quickly to changes in oxygen and nutrient content by adjusting gene expression activities. The metatranscriptome thus holds a significant potential for comprehending the functionality of a biofilm characterized by a stable microbial composition.

To identify the transcripts responsible for the observed PCA clustering, the 100 most extreme PC1 loadings were extracted (data not shown). The absolute value of these ranged from only 0.03 to 0.07, indicating that the variable had only a weak influence on the component. A manual inspection of the corresponding KEGG-annotations to these genes did not reveal specific transcripts encoding proteins known to be involved in metabolic pathways relevant to the nutrient physiology of the Hias Process. Metabolic reconstruction of the biofilm community was therefore conducted using metagenome-assembled genomes (MAGs) and the analysis of their corresponding metatranscriptome (see below).

Relative MAG abundance is not correlated to its contribution to the metatranscriptome

Putative MAG-encoding genes were predicted using the Distilled and Refined Annotation of Metabolism tool (DRAM) (Shaffer *et al.* 2020), resulting in the identification of a total of 613,814 ORFs encoded by the 180 MAGs. Gene-level TPM values were calculated by mapping the RNAseq reads to each gene using Kallisto, representing the relative gene activity level of each gene. Calculating the ratio of the mean MAG activity (mean TPM) divided by its relative abundance, revealed that the relative abundance of the reconstructed MAG was not correlated with relative transcriptional activity (Figure 5). For example, the mean TPM for the most prevalent MAG (MAG1) was 3.3, only 8.9% of the corresponding value for MAG6, which showed the highest overall mean TPM of 37. MAG6, taxonomically assigned to the genus *Propionivibrio*, demonstrated the overall highest transcriptional activity, whereas MAG7, assigned to the *Dermatophilaceae* family (GTDB accession JADKGA01) and corresponding to *Ca. Lutibacillus* in GTDB release 214 (Supplementary material 2, Table S1),

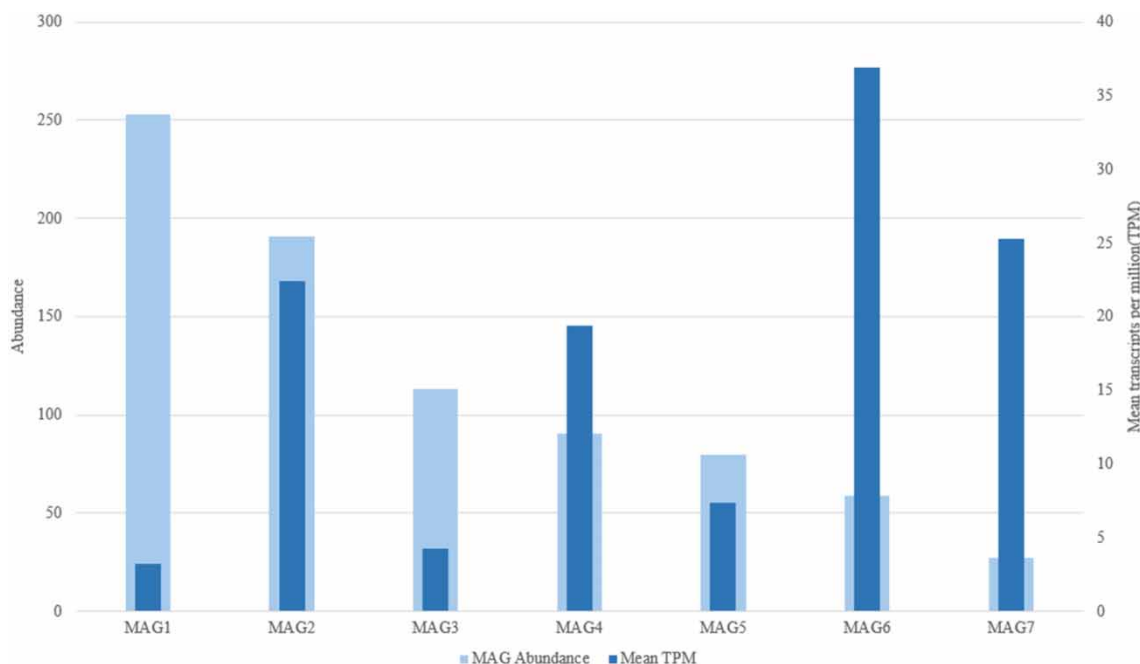


Figure 5 | Abundance and mean transcription of the metagenome-assembled genomes (MAG) investigated. Relative MAG abundance was calculated by mapping 10 million reads from each sample to the MAGs using bowtie2 and abundance was computed with samtools. Mean transcripts per million (TPM)/MAG represents the mean of all TPM values of RNAseq reads mapped to a given MAG. The MAGs shown in the graph was found to have the following taxonomy: MAG1, *Ca. Phosphoribacter*; MAG2, *Propionivibrio aalborgensis*; MAG3, *Austwickia*; MAG4, *Ca. Accumulibacter phosphatis*; MAG5, *Ca. Accumulibacter phosphatis*; MAG6, *Propionivibrio*; and MAG7, *Ca. Lutibacillus*.

ranked second highest in transcriptional activity. These results show that relative MAG abundance is not necessarily correlated to its contribution to the metatranscriptome, indicating that the most abundant biofilm MAGs is not necessarily the most transcriptionally active as a significant lower number of transcripts is derived from it. This suggests that less prevalent MAGs could potentially exhibit larger influence on biofilm functionality compared to more abundant MAGs. In the context of the Hias process, this observation suggests that despite *Ca. Phosphoribacter* (MAG1) having a significantly higher abundance compared to *Ca. Accumulibacter* (MAG4 and MAG5) and *Ca. Lutibacillus* (MAG7), it exhibits substantially lower transcriptional activity. This implies that *Ca. Accumulibacter* (MAG4 and MAG5) and *Ca. Lutibacillus* may play more pivotal functional roles in the Hias Process than *Ca. Phosphoribacter*.

Metabolic reconstruction of putative functional PAOs

Next aim was to reconstruct metabolic models of a subset of the MAGs, specifically those with putative PAM. Potential active metabolic pathways were investigated with differential RNA abundances across the anaerobic and aerobic phases of the reactor paired with metagenomics data as described. For this analysis, the focus was on the six most abundant MAGs, in addition to MAG7. Among these seven MAGs, only five harbor the *pit* gene (Supplementary material 1, Figure S4(a)), encoding the low-affinity P transporter PIT suggested to be common to all PAOs (McIlroy *et al.* 2014). Metabolic models of these five was thus reconstructed. As MAG4 and MAG5, both identified as *Ca. Accumulibacter phosphatis*, largely overlapped in functional profile, only the most transcriptionally active, MAG4 was included.

MAG4 - *Accumulibacter phosphatis*

Transcripts of all genes necessary for the biosynthesis and degradation of the biopolymers PHA, glycogen and poly-P were detected (Figure 6(a) and Supplementary material 1, Figure S4). The PHA synthase genes identified, namely *phaC* and *phaR*, encodes the catalytic and regulatory subunits of the class IV PHA synthase complex, respectively (McCool & Cannon 2001). These genes, and the *phaP* gene, encoding the granule-assembling protein phasin, were all identified as significantly upregulated in the anaerobic zones (Supplementary material 1, Figure S4(b)), suggesting their functional importance in the anaerobic phase of the process for the biosynthesis of PHA. This is in accordance with a recent

metatranscriptomic study on *Ca. Accumulibacter*, where *phaC* was also identified as anaerobically upregulated (McDaniel *et al.* 2021a). Transcripts corresponding to both the high-affinity *pst* and the low-affinity *pit* transporters were identified, the expression of the latter significantly upregulated in the aerobic zones (Supplementary material 1, Figure S4(a)), signifying the importance of the gene product (PIT transporter) during the aerobic phase of the process for the uptake of P. In contrast, the *pst* gene, not *pit*, was found to be aerobically upregulated in the McDaniel *et al.* study (2021a). Expression of the *ppk2* gene, encoding polyphosphate kinase 2 was found significantly upregulated aerobically, suggesting that PPK2 plays an important role in aerobic poly-P cycling in *Ca. Accumulibacter*. While PPK2 has been found to be involved in the synthesis of poly-P (Neville *et al.* 2022), it has mostly been associated with degradation of poly-P to catalyze the generation of GTP from GDP (Zhang *et al.* 2002). To our knowledge, no previous study has reported expression dynamics for *ppk2* in *Ca. Accumulibacter*. Transcripts of all genes necessary for the complete reduction of nitrate (NO_3^-) to N_2 (*nap*, *nir*, *nor* and *nos*) were identified (Supplementary material 1, Figure S4(d)), four of which (*napADG* and *nosD*) were found to be significantly upregulated anaerobically, suggesting that the *Ca. Accumulibacter* conducts complete denitrification in the anaerobic zone, which was also proposed by He & McMahon (2011), who found anaerobic upregulation of *nosZ* in *Ca. Accumulibacter*. Anaerobic upregulation of key genes associated with denitrification is, however, antithetical to the anaerobic biosynthesis of PHA, as the two processes compete for carbon. A possible explanation could be that the two processes (PAM and anaerobic respiration) occur in different spatial biofilm cell populations. A previous study investigating the spatial composition of the Hias biofilm identified high abundances of a *Ca. Accumulibacter* related MAG in the innermost layer of the biofilm (Villard *et al.* 2022). It is possible that cell populations in deeper layers of the biofilm, where oxygen concentrations are low also during aerobic phases of the process, is dependent on anaerobic respiration for growth.

MAG1 - *Ca. Phosphoribacter hodrii*

MAG1, corresponding to *Ca. Phosphoribacter hodrii*, contains all genes and transcripts necessary for the biosynthesis and degradation of glycogen and poly-P (Figure 6(b) and Supplementary material 1, Figure S4(a) and S4(c)). Glycogen is synthesized using the GlgE pathway, using maltose-1-phosphate as the building block. Expression of the glycogen phosphorylase (*glgP*) and pyruvate kinase (*pyk*) genes were significantly upregulated in the aerobic zones. While glycogen storage has yet to be demonstrated experimentally in *Ca. Phosphoribacter* (Singleton *et al.* 2022), our findings suggest that sugars stored as glycogen might function as a carbon and energy source during aerobic conditions when there is low access to external nutrients. Transcripts of all genes encoding the enzymes of the EMP and TCA pathways were detected (data not shown), suggesting that glucose-1-phosphate released in the glycogen phosphorylase reaction could be completely oxidized to CO_2 aerobically, with the committed generation of ATP through electron-transport phosphorylation. In addition to sugars, the significant upregulation of genes involved in amino acids transport and metabolism suggests that amino acids could be important carbon and energy sources for *Ca. Phosphoribacter hodrii* (Figure 6(b)). Aerobically, the *tnaA* and *rocF* genes, encoding tryptophanase and arginase, respectively, are upregulated. Tryptophanase catalyzes the cleavage of L-tryptophan to indole, pyruvate and NH_4^+ , thus potentially contributing to both nitrogen and carbon for growth during the aerobic phase. Interestingly, indole is commonly used by microorganisms as an intra- and interspecies signaling molecule, regulating diverse microbial processes such as motility and biofilm formation (Lee & Lee 2010). Arginase is an enzyme involved in basic nitrogen metabolism and redistribution and hydrolyzes arginine to the key metabolic precursor L-ornithine and urea (Hernández *et al.* 2021). Through the activity of the gene products *rocD* and *rocA*, L-ornithine can be converted into two molecules of glutamate, which can be further funneled and catabolized in the TCA cycle, by conversion to α -ketoglutarate by *rocG*, encoding an NAD-specific glutamate dehydrogenase (Figure 6(b)). Anaerobically, pyruvate seems to be fermented to alanine, lactate, and acetate, by the significant upregulation of the genes encoding alanine dehydrogenase (*ald*), lactate dehydrogenase (*lldE*) or acetyl-CoA hydrolase, respectively (Supplementary material 1, Figure S4(e)). Alanine has been suggested as a potential carbon storage molecule in *Tetrasphaera*-related genera (Kristiansen *et al.* 2013), and it is likely that the anaerobically produced alanine serves as a source for carbon to be catabolized through the TCA cycle in aerobic zones. The production of the VFAs acetate and lactate could potentially be a source of VFAs for *Ca. Accumulibacter*, an interaction that has been described (Chen *et al.* 2023). The enzyme acetyl-CoA hydrolase is suggested to be involved in regulating the intracellular acetyl-CoA or CoASH pools (Buu *et al.* 2003) and does not contribute to the generation of ATP at substrate-level phosphorylation as does acetate kinase (*ackA*), which was not detected in MAG1. Thus, fermentation of sugars through EMP seems to be important for ATP generation during anaerobic conditions. Anaerobic production of acetyl-CoA from pyruvate is likely catalyzed by the enzyme pyruvate ferredoxin oxidoreductase (encoded by the *por* genes) as the negative

regulator (*pdhR*) for the pyruvate dehydrogenase complex (*aceEF* and *lpd* was significantly upregulated in the anaerobic zone. It has been confirmed experimentally that *Ca. Phosphoribacter* exhibit dynamic poly-P cycling during alternating anaerobic and aerobic conditions (Singleton *et al.* 2022). Based on the observed expression profile of relevant genes for MAG1 the authors propose that the energy required for aerobic P accumulation comes from stored amino acids and glycogen, and that the MAG ferments sugars and amino acids in addition to hydrolyzing poly-P to supply energy anaerobically.

MAG7 – *Ca. Lutibacillus*

Ca. Lutibacillus is newly described species, formerly assigned to the clade 3 *Tetrasphaera*, shown experimentally to exhibit dynamic poly-P cycling (Singleton *et al.* 2022). MAG7 (Figure 6(c)) shows potential for poly-P cycling and all relevant transcripts were identified, including *pit* specific transcripts, however, transcripts for two subunits of the high-affinity phosphate transporter, *pst*, (*pstA* and *pstC*) were not identified (Supplementary material 1, Figure S4(a)). Along with *ppk1*, *ppk2* and *ppx*, MAG7 encodes and produced transcripts encoding the enzyme poly-P glucokinase (*ppgk*), which catalyze the phosphorylation of glucose using poly-P as substrate (Pepin & Wood 1986). Genes and transcripts necessary for the biosynthesis of glycogen utilizing maltose-1-phosphate as precursor, and for subsequent glycogen degradation, were present (Supplementary material 1, Figure S4(c)), suggesting the potential of using glycogen as energy storage for poly-P synthesis, however, similar to *Ca. Phosphoribacter*, *in situ* analysis has not detected glycogen storage in *Ca. Lutibacillus* (Singleton *et al.* 2022). Anaerobically, alanine could be produced as a fermentation product of glucose through the activity of alanine dehydrogenase (*ald*), which is significantly upregulated (Supplementary material 1, Figure S4(e)). Fermentation of the amino acid lysin could produce acetate anaerobically through the activity of the genes (*lysP*, *kce* and *atoDA*), which all were significantly upregulated (Figure 6(c)). Additionally, *kamD*, *kamA*, *ackA* and *pta* was only found expressed anaerobically, further supporting the notion that amino acids are important for anaerobic energy generation. It should be mentioned that the PTA-ACKA pathway has not previously been described in *Ca. Lutibacillus*.

MAG6 – *Propionivibrio*

MAG6 was assigned to the genus *Propionivibrio*, a genus closely related to *Ca. Accumulibacter* (Albertsen *et al.* 2016). MAG6 (Figure 6(d) and Supplementary material 1, Figure S4) showed a similar genetic potential to that of *Ca. Accumulibacter* (MAG4). Genes and transcripts for PHA synthesis were expressed, several of which were significantly upregulated anaerobically (*phaR*, *ackA*, *asc* and *ptb*) strongly suggesting that PHA synthesis from butyrate and acetate occur. With respect to poly-P transport, MAG6 was found to encode and express the *pit* gene, but not the *pst* gene (Supplementary material 1, Figure S4(a)). Several other verified PAOs lack the *pst* gene, including one previous MAG assigned as *Ca. Phosphoribacter*, and *Dechloromonas* (Ruiz-Haddad *et al.* 2024), suggesting that poly-P accumulation is not dependent on the high-affinity transporter. MAG6 showed the potential for synthesizing poly-P through expression of *ppk1*, and poly-P degradation through the expression of either *ppk2*, *ppx*, *pap/adk* and/or *ppgk*. Anaerobically, MAG6 was found to encode all three subunits of a multiple sugar transporter (*gguAB* and *chvE*), which both were significantly upregulated (Figure 6(d)). The sugars could be converted to pyruvate through the EMP pathway, providing reducing equivalents for PHA synthesis. Simple sugars could also be used for the biosynthesis of glycogen, as all necessary transcripts for genes involved in the synthesis and degradation of glycogen were identified (Supplementary material 1, Figure S4(c)). Interestingly, MAG6 possesses the genetic potential and produces transcripts similar to those of *Ca. Accumulibacter* (MAG4). Therefore, it remains undetermined whether this MAG function as a GAO or a PAO within the biofilm. However, species belonging to the genus *Propionivibrio* has been reported as GAOs, not being able to contribute to biological P removal, in EBPR systems. If MAG6 is a GAO it is interesting, however, that this MAG exhibit the highest relative transcription of the reconstructed MAGs and thus should be a competitor to PAOs for VFAs for PHA synthesis. Alternatively, PAM in the Hias Process could be less dependent on VFAs and *Accumulibacter*, and be more reliant on the activity of *Ca. Phosphoribacter* and *Ca. Lutibacillus* using sugars and amino acids to drive PAM.

PAOs in the Hias process

The prevalent Hias biofilm MAG1 (*Ca. Phosphoribacter*), MAG4 (*Ca. Accumulibacter*) and MAG7 (*Ca. Lutibacillus*) have all been shown experimentally to perform dynamic P cycling in EBPR systems (Singleton *et al.* 2022). However, without chemical analysis of relevant biofilm biopolymers, such as glycogen, PHA and poly-P, no conclusion regarding which of these are the most important PAOs for EBPR performance in the Hias Process can be drawn. However, *Ca. Lutibacillus* and *Ca. Accumulibacter* display a significant higher relative transcriptional activity compared to *Ca. Phosphoribacter*, which suggests that

the latter is less important for EBPR, despite being the most prevalent. The high abundance of the *Ca. Phosphoribacter* MAG could be explained by a higher transcriptional activity in previous EBPR cycles, and that the low relative expression observed in the current EBPR cycle could result in a lower abundance in successive cycles. The microbial composition may undergo dynamic changes across successive EBPR cycles due to variations in the chemistry of the incoming wastewater or temperature. In response, the biofilm adjusts its microbial composition to adapt to these ongoing and continuous changes. It would be interesting to analyze the biofilm dynamics across successive EBPR cycles to investigate this hypothesis further.

CONCLUSIONS

EBPR in combination with MBBR technology represents a novel approach to municipal wastewater treatment. In the current study, the authors demonstrate that changes in the metatranscriptome are critical in determining the functionality of the biofilm, while the microbial composition and structure of the biofilm remain stable throughout a complete EBPR cycle. The authors further show that the presence of a highly transcriptional active population of GAOs is not detrimental to the efficiency of EBPR. Notably, the abundance of these organisms does not correlate with transcriptional activity. These data are important for advancing our understanding of microbial community dynamics and their role in the functionality of biofilms in wastewater treatment systems. The insights gained from this study could inform the optimization of EBPR processes, contributing to more efficient and reliable wastewater treatment solutions.

ACKNOWLEDGEMENTS

The authors extend their gratitude to Professor Rob Wilson for his assistance and guidance in server-related tasks, Associate professor Abdolrahman Khezri for providing insights into bioinformatic analyses, and Karoline Refsahl for her consistent availability for sample collection at the Hias WWTP. This work was supported by Regional Research Fund Innlandet (RFFI) (Grant number 313463, 2020). RFFI did not have any role in the collection, analysis and interpretation of data, the writing of the report, or in the decision to submit the article for publication.

AUTHOR CONTRIBUTIONS

D.V. contributed to conceptualization, data curation, formal analysis, investigating, visualization, writing – original draft. L.S. contributed to data curation, formal analysis, methodology, software, supervision, visualization, writing – review and editing. K.R. contributed to conceptualization, methodology, supervision, writing – review and editing. S.B. investigated the study. T.S. contributed to conceptualization, resources. S.E. contributed to investigating, resources. W.J. contributed to conceptualization, funding acquisition, methodology, project administration, supervision, visualization, writing – original draft.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- Albertsen, M., Mccilroy, S. J., Stokholm-Bjerregaard, M., Karst, S. M. & Nielsen, P. H. 2016 'Candidatus Propionivibrio aalborgensis': A novel glycogen accumulating organism abundant in full-scale enhanced biological phosphorus removal plants. *Nat Rev Microbiol* **7**, 1033.
- Battin, T. J., Besemer, K., Bengtsson, M. M., Romani, A. M. & Packmann, A. I. 2016 The ecology and biogeochemistry of stream biofilms, *Nature Reviews Microbiology* **14**, 251–263.
- Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. 2016 Near-optimal probabilistic RNA-seq quantification, *Nature Biotechnology* **34**, 525–527.
- Buchfink, B., Reuter, K. & Drost, H.-G. 2021 Sensitive protein alignments at tree-of-life scale using DIAMOND, *Nature Methods* **18**, 366–368.
- Bushnell, B., Rood, J. & Singer, E. 2017 BBMerge – accurate paired shotgun read merging via overlap, *PLoS One* **12**, e0185056.
- Buu, L.-M., Chen, Y.-C. & Lee, F.-J. S. 2003 Functional characterization and localization of acetyl-CoA hydrolase, *ach1p*, in *Saccharomyces cerevisiae*, *Journal of Biological Chemistry* **278**, 17203–17209.
- Cech, J. S. & Hartman, P. 1993 Competition between polyphosphate and polysaccharide accumulating bacteria in enhanced biological phosphate removal systems, *Water Research* **27**, 1219–1225.

- Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. 2019 *GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database*, *Bioinformatics* **36**, 1925–1927.
- Chen, G., Bai, R., Zhang, Y., Zhao, B. & Xiao, Y. 2022 *Application of metagenomics to biological wastewater treatment*, *Science of the Total Environment* **807**, 150737.
- Chen, L., Wei, G., Zhang, Y., Wang, K., Wang, C., Deng, X., Li, Y., Xie, X., Chen, J., Huang, F., Chen, H., Zhang, B., Wei, C. & Qiu, G. 2023 *Candidatus Accumulibacter use fermentation products for enhanced biological phosphorus removal*, *Water Research* **246**, 120713.
- Chklovski, A., Parks, D. H., Woodcroft, B. J. & Tyson, G. W. 2023 *CheckM2: A rapid, scalable and accurate tool for assessing microbial genome quality using machine learning*, *Nature Methods* **20**, 1203–1212.
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A. & Davies, R. M. 2021 *Twelve years of SAMtools and BCFtools*, *Gigascience* **10**, giab008.
- Dueholm, M. K. D., Nierychlo, M., Andersen, K. S., Rudkjøbing, V., Knutsson, S., Albertsen, M. & Nielsen, P. H. 2022 *MiDAS 4: A global catalogue of full-length 16S rRNA gene sequences and taxonomy for studies of bacterial communities in wastewater treatment plants*, *Nature Communications* **13**, 1908.
- Edel, M., Horn, H. & Gescher, J. 2019 *Biofilm systems as tools in biotechnological production*, *Applied Microbiology and Biotechnology* **103**, 5095–5103.
- Edgar, R. C. 2016a *SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences*. *BioRxiv*, 074161.
- Edgar, R. C. 2016b *UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing*. *BioRxiv*, 081257.
- Feng, K., Zhang, Z., Cai, W., Liu, W., Xu, M., Yin, H., Wang, A., He, Z. & Deng, Y. 2017 *Biodiversity and species competition regulate the resilience of microbial biofilm community*, *Molecular Ecology* **26**, 6170–6182.
- Fernando, E. Y., Mcilroy, S. J., Nierychlo, M., Herbst, F. A., Petriglieri, F., Schmid, M. C., Wagner, M., Nielsen, J. L. & Nielsen, P. H. 2019 *Resolving the individual contribution of key microbial populations to enhanced biological phosphorus removal with Raman-FISH*, *ISME Journal* **13**, 1933–1946.
- Flemming, H. C. & Wuertz, S. 2019 *Bacteria and archaea on Earth and their abundance in biofilms*, *Nature Reviews Microbiology* **17**, 247–260.
- Flemming, H.-C., Van Hullebusch, E. D., Neu, T. R., Nielsen, P. H., Seviour, T., Stoodley, P., Wingender, J. & Wuertz, S. 2023 *The biofilm matrix: Multitasking in a shared space*, *Nature Reviews Microbiology* **21**, 70–86.
- Flowers, J. J., He, S., Yilmaz, S., Noguera, D. R. & McMahon, K. D. 2009 *Denitrification capabilities of two biological phosphorus removal sludges dominated by different 'Candidatus Accumulibacter' clades*, *Environmental Microbiology Reports* **1**, 583–588.
- He, S. & McMahon, K. D. 2011 *Microbiology of 'Candidatus Accumulibacter' in activated sludge*, *Microb Biotechnol* **4**, 603–619.
- Hernández, V. M., Arteaga, A. & Dunn, M. F. 2021 *Diversity, properties and functions of bacterial arginases*, *FEMS Microbiology Reviews* **45**, fuab034.
- Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W. & Hauser, L. J. 2010 *Prodigal: Prokaryotic gene recognition and translation initiation site identification*, *BMC Bioinformatics* **11**, 119.
- Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H. & Wang, Z. 2019 *MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies*, *PeerJ* **7**, e7359.
- Kristiansen, R., Nguyen, H. T., Saunders, A. M., Nielsen, J. L., Wimmer, R., Le, V. Q., Mcilroy, S. J., Petrovski, S., Seviour, R. J., Calteau, A., Nielsen, K. L. & Nielsen, P. H. 2013 *A metabolic model for members of the genus Tetrasphaera involved in enhanced biological phosphorus removal*, *ISME Journal* **7**, 543–554.
- Langmead, B. & Salzberg, S. L. 2012 *Fast gapped-read alignment with Bowtie 2*, *Nature Methods* **9**, 357–359.
- Lee, J.-H. & Lee, J. 2010 *Indole as an intercellular signal in microbial communities*, *FEMS Microbiology Reviews* **34**, 426–444.
- Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. 2017 *Bracken: Estimating species abundance in metagenomics data*, *PeerJ Computer Science* **3**, e104.
- Lyle, R., Hughes, T., Undlien, D. & Jakobsen, K. 2012 *The Norwegian Sequencing Centre (NSC)*. **17**.
- McCool, G. J. & Cannon, M. C. 2001 *Phac and PhaR are required for polyhydroxyalkanoic acid synthase activity in Bacillus megaterium*, *Journal of Bacteriology* **183**, 4235–4243.
- Mcdaniel, E. A., Moya-Flores, F., Keene Beach, N., Camejo, P. Y., Oyserman, B. O., Kizacic, M., Khor, E. H., Noguera, D. R. & McMahon, K. D. 2021a *Metabolic differentiation of Co-occurring accumulibacter clades revealed through genome-resolved metatranscriptomics*, *mSystems* **6**, e0047421.
- Mcdaniel, E. A., Wahl, S. A., Ishii, S., Pinto, A., Ziels, R., Nielsen, P. H., McMahon, K. D. & Williams, R. B. H. 2021b *Prospects for multi-omics in the microbial ecology of water engineering*, *Water Research* **205**, 117608.
- Mcilroy, S. J., Albertsen, M., Andresen, E. K., Saunders, A. M., Kristiansen, R., Stokholm-Bjerregaard, M., Nielsen, K. L. & Nielsen, P. H. 2014 *'Candidatus Competibacter'-lineage genomes retrieved from metagenomes reveal functional metabolic diversity*, *ISME Journal* **8**, 613–624.
- Mino, T., Van Loosdrecht, M. C. M. & Heijnen, J. J. 1998 *Microbiology and biochemistry of the enhanced biological phosphate removal process*, *Water Research* **32**, 3193–3207.
- Neville, N., Roberge, N. & Jia, Z. 2022 *Polyphosphate kinase 2 (PPK2) enzymes: Structure, function, and roles in bacterial physiology and virulence*, *International Journal of Molecular sciences* **23**, 670.
- Nguyen, H. T. T., Le, V. Q., Hansen, A. A., Nielsen, J. L. & Nielsen, P. H. 2011 *High diversity and abundance of putative polyphosphate-accumulating Tetrasphaera-related bacteria in activated sludge systems*, *FEMS Microbiology Ecology* **76**, 256–267.

- Nguyen, H. T., Kristiansen, R., Vestergaard, M., Wimmer, R. & Nielsen, P. H. 2015 Intracellular accumulation of glycine in polyphosphate-Accumulating organisms in activated sludge, a novel storage mechanism under dynamic anaerobic-aerobic conditions, *Applied and Environmental Microbiology* **81**, 4809–4818.
- Nielsen, P. H., McIlroy, S. J., Albertsen, M. & Nierychlo, M. 2019 Re-evaluating the microbiology of the enhanced biological phosphorus removal process, *Current Opinion in Biotechnology* **57**, 111–118.
- Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. 2017 metaSPAdes: A new versatile metagenomic assembler, *Genome Research* **27**, 824–834.
- Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. 2017 dRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication, *The ISME Journal* **11**, 2864–2868.
- Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A. & Hugenholtz, P. 2018 A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life, *Nature Biotechnology* **36**, 996–1004.
- Pepin, C. A. & Wood, H. G. 1986 Polyphosphate glucokinase from *Propionibacterium shermanii*. Kinetics and demonstration that the mechanism involves both processive and nonprocessive type reactions, *Journal of Biological Chemistry* **261**, 4476–4480.
- Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. 2016 VSEARCH: A versatile open source tool for metagenomics, *PeerJ* **4**, e2584.
- Ruiz-Haddad, L., Ali, M., Pronk, M., Van Loosdrecht, M. C. M. & Saikaly, P. E. 2024 Demystifying polyphosphate-accumulating organisms relevant to wastewater treatment: A review of their phylogeny, metabolism, and detection, *Environmental Science and Ecotechnology* **21**, 100387.
- Saltnes, T., Sorensen, G. & Eikas, S. 2017 Biological nutrient removal in a continuous biofilm process, *Water Practice and Technology* **12**, 797–805.
- Saunders, A. M., Oehmen, A., Blackall, L. L., Yuan, Z. & Keller, J. 2003 The effect of GAOs (glycogen accumulating organisms) on anaerobic carbon requirements in full-scale Australian EBPR (enhanced biological phosphorus removal) plants, *Water Science and Technology* **47**, 37–43.
- Shaffer, M., Borton, M. A., Mcgovern, B. B., Zayed, A. A., La Rosa, S. L., Solden, L. M., Liu, P., Narrowe, A. B., Rodríguez-Ramos, J., Bolduc, B., Gazitúa, M. C., Daly, R. A., Smith, G. J., Vik, D. R., Pope, P. B., Sullivan, M. B., Roux, S., Wrighton & Kelly, C. 2020 DRAM for distilling microbial metabolism to automate the curation of microbiome function, *Nucleic Acids Research* **48**, 8883–8900.
- Singleton, C. M., Petriglieri, F., Wasmund, K., Nierychlo, M., Kondrotaitė, Z., Petersen, J. F., Peces, M., Dueholm, M. S., Wagner, M. & Nielsen, P. H. 2022 The novel genus, 'Candidatus Phosphoribacter', previously identified as *Tetrasphaera*, is the dominant polyphosphate accumulating lineage in EBPR wastewater treatment plants worldwide, *ISME Journal* **16**, 1605–1616.
- Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. & Schmidt, T. M. 2015 rrn DB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development, *Nucleic Acids Research* **43**, D593–D598.
- Takahashi, S., Tomita, J., Nishioka, K., Hisada, T. & Nishijima, M. 2014 Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing, *PLoS One* **9**, e105592.
- Villard, D., Saltnes, T., Sørensen, G., Angell, I. L., Eikås, S., Johansen, W. & Rudi, K. 2022 Spatial fractionation of phosphorus accumulating biofilm: Stratification of polyphosphate accumulation and dissimilatory nitrogen metabolism, *Biofouling* **38**, 162–172.
- Wood, D. E., Lu, J. & Langmead, B. 2019 Improved metagenomic analysis with Kraken 2, *Genome Biology* **20**, 1–13.
- Wu, Y. W., Simmons, B. A. & Singer, S. W. 2016 Maxbin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets, *Bioinformatics* **32**, 605–607.
- Yu, Z., Xu, S., Wang, P., Liu, D. & Lu, H. 2023 Phosphorus removal and storage polymer synthesis by tetrasphaera-related bacteria with different carbon sources, *ACS ES&T Water* **3**, 1374–1384.
- Zhang, H., Ishige, K. & Kornberg, A. 2002 A polyphosphate kinase (PPK2) widely conserved in bacteria, *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16678–16683.
- Zhao, W., Bi, X., Peng, Y. & Bai, M. 2022 Research advances of the phosphorus-accumulating organisms of *Candidatus Accumulibacter*, *Dechloromonas* and *Tetrasphaera*: Metabolic mechanisms, applications and influencing factors, *Chemosphere* **307**, 135675.
- Zhou, Y., Pijuan, M., Oehmen, A. & Yuan, Z. 2010 The source of reducing power in the anaerobic metabolism of polyphosphate accumulating organisms (PAOs) - a mini-review, *Water Science and Technology* **61**, 1653–1662.

First received 27 May 2024; accepted in revised form 22 August 2024. Available online 17 September 2024