Genetic potential for exopolysaccharide synthesis in activated sludge bacteria uncovered by genome-resolved metagenomics

Dueholm, Morten Kam Dahl; Besteman, Maaike; Zeuner, Emil Juel; Riisgaard-Jensen, Marie; Nielsen, Morten Eneberg; Vestergaard, Sofie Zacho; Heidelbach, Søren; Bekker, Nicolai Sundgaard; Nielsen, Per Halkjær

Published in:
Water Research

DOI (link to publication from Publisher):
10.1016/j.watres.2022.119485

Publication date:
2023

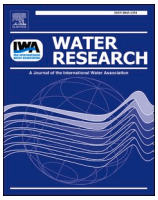
Document Version
Publisher's PDF, also known as Version of record

Water Research 229 (2023) 119485

Contents lists available at ScienceDirect

Water Research

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



Genetic potential for exopolysaccharide synthesis in activated sludge bacteria uncovered by genome-resolved metagenomics

Morten Kam Dahl Dueholm [a,*], Maaike Besteman [b], Emil Juel Zeuner [a], Marie Riisgaard-Jensen [a], Morten Eneberg Nielsen [a], Sofie Zacho Vestergaard [a], Søren Heidelbach [a], Nicolai Sundgaard Bekker [a], Per Halkjær Nielsen [a]

[a] Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark
[b] Department of Agrotechnology and Food Sciences, Wageningen University & Research, Wageningen, Netherlands

ARTICLE INFO

Keywords:
Activated sludge
EPS
Exopolysaccharides
Genome mining

ABSTRACT

A good floc formation of activated sludge (AS) is crucial for solid-liquid separation and production of clean effluent during wastewater treatment. Floc formation is partly controlled by self-produced extracellular polymeric substances (EPS) such as exopolysaccharides, proteins, and nucleic acids. Little is known about the composition, structure, and function of EPS in AS and which bacteria produce them. To address this knowledge gap for the exopolysaccharides, we took advantage of 1083 high-quality metagenome-assembled genomes (MAGs) obtained from 23 Danish wastewater treatment plants. We investigated the genomic potential for exopolysaccharide biosynthesis in bacterial species typical in AS systems based on genome mining and gene synteny analyses. Putative gene clusters associated with the biosynthesis of alginate, cellulose, curdlan, diutan, hyaluronic acids, Pel, poly-β-1,6-N-acetyl-D-glucosamine (PNAG), Psl, S88 capsular polysaccharide, salecan, succinoglycan, and xanthan were identified and linked to individual MAGs, providing a comprehensive overview of the genome-resolved potential for these exopolysaccharides in AS bacteria. The approach and results provide a starting point for a more comprehensive understanding of EPS composition in wastewater treatment systems, which may facilitate a more refined regulation of the activated sludge process for improved stability.

1. Introduction

Wastewater treatment is vital to secure human health, prevent environmental pollution, and recover important resources. Activated sludge (AS), the most widely used technology for biological wastewater treatment, relies on the catabolic activity of suspended microbial aggregates, AS flocs (Seviour and Nielsen, 2010). Good separation between the solid-liquid phase is essential to recover clean effluent from the sludge. The flocculation of the microorganisms determines this process, and poor floc formation can lead to process failures, including deterioration of effluent quality and decreased dewaterability of the sludge (Karr and Keinath, 1978; Wilén et al., 2000).

Extracellular polymeric substances (EPS) constitute a major component of the AS flocs and are essential for linking the cells together in a large three-dimensional matrix (Frølund et al., 1996). The EPS matrix protects against predation and environmental stressors, helps the bacteria sequester nutrients and allows microbial cells to live in close association, thereby facilitating cell-to-cell contact and horizontal gene transfer (Flemming et al., 2016, 2022). The EPS represent a mixture of substances originating from active secretion, cell surface material shedding, cell lysis, and sorption from the environment and include exopolysaccharides, proteins, lipids, and nucleic acids (Liu and Fang, 2003). However, the exact composition of the matrix is highly variable and dynamic. Exopolysaccharides have proved important for shape and structural support of the matrix as well as adhesion and aggregation of cells (Seviour et al., 2019). Additionally, the exopolysaccharides are important for the retention of water and extracellular enzymes (Flemming and Wingender, 2010; Siddharth et al., 2021). Besides their biological roles, EPS have also gained increased attention as they can be extracted and formulated into renewable bioresources, such as flocculants for improved wastewater treatment, eco-friendly flame-retardant materials, and binders for composite materials (Ajao et al., 2018; Feng et al., 2021; Karakas et al., 2020; Kim et al., 2020; Nouha et al., 2018).

Despite their apparent importance, wastewater treatment-related

* Corresponding author.
E-mail address: md@bio.aau.dk (M.K.D. Dueholm).

https://doi.org/10.1016/j.watres.2022.119485
Received 13 October 2022; Received in revised form 8 December 2022; Accepted 10 December 2022
Available online 12 December 2022

exopolysaccharides are poorly described because such investigations traditionally require pure, or at least highly enriched, cultures and subsequent extraction and characterization of the individual EPS (Flemming et al., 2022; Seviour et al., 2019). Among described exopolysaccharides are polymers from a *Zoogloea resiniphila* strain (An et al., 2016; Gao et al., 2018), *Duganella zoogloeoides* ATCC 25,925 (Ikeda et al., 1982), and *Thauera* (Allen et al., 2004). In addition to these pure culture-derived exopolysaccharides, there are granulan (Seviour et al., 2010) and alginate-like exopolymers (ALE) (Lin et al., 2013) which were both extracted from aerobic granular sludge. Granulan and ALE both contain anionic polysaccharides and depend on divalent cations for gel formation (Lin et al., 2010; Seviour et al., 2012). However, as these exopolymers were extracted from complex microbial communities, they represent a mixture of different EPS molecules from several taxa. Accordingly, we have little knowledge about the exact composition, which bacteria have produced them, and the associated biosynthetic machinery.

Most knowledge regarding biofilm-associated exopolysaccharides originates from the clinical fields with pure cultures from genera such as *Pseudomonas, Escherichia,* and *Staphylococcus* (Atkin et al., 2014; Gill et al., 2005; Ma et al., 2006; Monteiro et al., 2009). Some common characteristics are often shared across exopolysaccharides, including the genomic clustering of genes involved in their biogenesis (Schmid et al., 2015). There are three well-described pathways for exopolysaccharide synthesis in bacteria, which each share some conserved key enzymes. These are the synthase-dependent pathway, the Wzx/Wzy-dependent pathway, and the ABC transporter-dependent pathway (Schmid et al., 2015). The conserved nature of the exopolysaccharide synthesis pathways suggests that it is possible to reveal the genetic potential for exopolysaccharides of individual species in complex microbial communities through genome mining of the relevant microbes. However, this has previously been hampered in complex systems by the quality of metagenome-assembled genomes (MAGs) obtained by short-read sequencing, which are often highly fragmented and lack genes that occur in multiple copies or includes repetitive sequences (Koren and Phillippy, 2015). Accordingly, these MAGs are unsuited for gene synteny analyses and may lack important genes, such as those encoding the periplasmic tetratricopeptide repeat (TPR)-containing scaffold proteins involved in the synthase-dependent polysaccharide biogenesis (Whitney and Howell, 2013).

Recently, we recovered 1083 high-quality (HQ) MAGs from 23 Danish WWTPs based on hybrid assembly of Nanopore long-reads and Illumina short-reads (Singleton et al., 2021). These MAGs represent 578 bacterial species and 65 of the top 100 most abundant species in Danish WWTPs (Nierychlo et al., 2020; Singleton et al., 2021). Most MAGs affiliate with the Bacteroidota, Proteobacteria, and Chloroflexota phylum, but MAGs from 29 additional phyla were also recovered. These MAGs have high contiguity and are >90% complete, meeting the MIMAG standards for HQ MAGs (Bowers et al., 2017), which enables the identification and comprehensive characterization of conserved operons and gene clusters, even when these are long and contain repetitive sequences (Goldstein et al., 2019). The HQ MAGs, therefore, represent a valuable foundation for linking microbial identity (taxonomy) with their genetic potential for exopolysaccharide production. Furthermore, most of these MAGs represent common genera across global wastewater treatment systems (Dueholm et al., 2022).

Here, we carried out a literature survey to identify experimentally validated exopolysaccharide biosynthetic pathways encoded in operons or gene clusters within well-characterized bacterial isolates. We identified 16 exopolysaccharide biosynthetic gene clusters (BGCs) and used these to compile a database of the corresponding proteins. This database was then used to identify homologous systems in the AS HQ MAGs, hereby linking the potential for exopolysaccharide production to individual bacterial species. Based on homolog identification and gene proximity analysis, we identified putative gene clusters related to 10 of the examined exopolysaccharide biosynthetic pathways. To our knowledge, this is the first comprehensive description of the genome-resolved potential for these exopolysaccharides in AS bacteria. Generally, the biosynthetic pathways were restricted to one or two phyla. Alginate BGCs were restricted to a few low-abundant MAGs in a single family, questioning the proposed importance of alginate (Zhang et al., 2021) in wastewater treatment systems.

2. Materials and methods

2.1. Source data

Translated gene sequences from experimentally validated polysaccharide biosynthesis pathways (Table 1) were acquired from the National Center for Biotechnology Information (NCBI) database. The HQ MAGs from AS and their taxonomy were obtained from the study by Singleton et al. (2021). Translated coding sequences (CDSs) were identified and annotated using Prokka v. 1.14 (Seemann, 2014) with the following options –meta, –kingdom Bacteria or Archaea (depending on the taxonomy of the MAG).

2.2. Identification of protein homologs

The Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST) (BLAST+ v.2.11.0) (Altschul et al., 1997) was used to identify protein homologs in the Prokka-generated HQ MAG protein database (Singleton et al., 2021). PSI-BLAST parameters were initially optimized using proteins encoded by the PNAG operon as query, and empirically optimal parameters were found to be: -num_iterations 20, -max_hsp 10, -evalue 0.001, -qcov_hsp_perc 20, and -max_target_seq 100, 000. Output format (-out_fmt) 6 was used. These parameters were then used for all searches.

2.3. Prediction and visualization of exopolysaccharide biosynthetic gene clusters

Proteins related to exopolysaccharide biosynthesis gene clusters were evaluated based on the number of homologs found for each exopolysaccharide system and whether their genes were located together in the genomes (Fig. 1). To achieve this, results from PSI-BLAST searches for individual polysaccharide queries were analyzed in R based on the following steps: i) For each query, homologs identified by PSI-BLAST were evaluated based on percent identity thresholds empirically optimized to enhance search specificity while limiting the number of hits outside the identified biosynthetic gene clusters: Homologs with <20% identity were removed for all exopolysaccharide systems, except for the hyaluronic acid system from *Pasteurella multocida* which only contains a single query protein. For this system, a cutoff at <33% identity was applied to reduce the rate of false positive hits. ii) To avoid homologs associated with unrelated pathways, a proximity criterion for genes in putative gene clusters was introduced: Identified homologs whose genes were separated by more than eight annotated genes or 2000 bp from other homologs from the same exopolysaccharide gene cluster were removed. The proximity criteria were set arbitrarily based on an empirical evaluation. iii) Putative gene clusters were filtered based on the number of homologs identified and the presence of essential key enzymes (see Table 1). iv) For the proteins related to the putative gene clusters, amino acid sequences were mapped to the Pfam database (Mistry et al., 2021) using InterProScan search (Jones et al., 2014) to identify conserved protein domains. v) The phylogenetic distribution of each exopolysaccharide system was visualized using the phylogenetic tree created by Singleton et al. (2021). The synteny of gene clusters was examined using gene arrow maps.

The following R-packages were used: seqinr v. 4.2–16 (Gouy et al., 1984), tidyverse v. 1.3.1 (Wickham et al., 2019), gggenes v. 0.4.1 (Wilkins, 2020), ggtree v. 3.2.1 (Yu et al., 2017), and ggtreeExtra v. 1.6.0 (Xu et al., 2021). Final figures were assembled and visually

Table 1
Overview of exopolysaccharide operons or gene clusters used as queries in this study. *Identity threshold for this gene was 33%.

Exopolysaccharide	Organism	Pathway type	Genes	Genes threshold	Essential genes	Accession nr.	Reference
Alginate	*Pseudomonas aeruginosa*	Synthase	12	6		AE004091.2	Franklin et al., 2011 Rehman et al., 2013
Cellulose I	*Komagataeibacter medellinensis*	Synthase	7	2	*bcsAI*	AB015802.1	Matsutani et al., 2015
Cellulose II	*Komagataeibacter medellinensis*	Synthase	4	2	*bcsABII*	AB015803.1	Matsutani et al., 2015
Curdlan	*Agrobacterium* sp.	Synthase	3	2	*crdS*	AECL01000028	Karnezis et al., 2003
Diutan	*Sphingomonas* sp.	Wzx/Wzy	20	10		EU026118.1	Coleman et al., 2008
Hyaluronic acid (*pmHAS*)	*Pasteurella multocida*	Synthase	1	1	*pmHAS**	AF036004.2	DeAngelis et al., 1998
Hyaluronic acid (*hasA*)	*Streptococcus zooepidemicus* *Streptococcus pyogenes*	Synthase	3	3	*hasA*	CP002904.1 AE009949.1	Ma et al., 2011 Smoot et al., 2002
Pel	*Pseudomonas aeruginosa*	Synthase	7	6	*pelF*	AE004091.2	Franklin et al., 2011
PNAG (*eps*)	*Bacillus subtilis*	Synthase	4	3	*epsH* *epsJ*	NC_000964.3	Roux et al., 2015
PNAG (*ica*)	*Staphylococcus epidermidis*	Synthase	4	3	*icaA*	CP000029.1	Atkin et al., 2014 Gill et al., 2005
PNAG (*pga*)	*Escherichia coli*	Synthase	4	3	*pgaC*	NC_000913.3	Wang et al., 2004
Psl	*Pseudomonas aeruginosa*	Wzx/Wzy	12	7		AE004091	Franklin et al., 2011
S88	*Sphingomonas* sp.	Wzx/Wzy	19	9		U51197.1	Yamazaki et al., 1996
Salecan	*Agrobacterium* sp.	Wzx/Wzy	13	7		KT780309	Xu et al., 2017
Succinoglycan	*Sinorhizobium meliloti*	Wzx/Wzy	19	9		NC_003078.1	Finan et al., 2001
Xanthan	*Xanthomonas campestris*	Wzx/Wzy	12	6		AM920689	Vorhölter et al., 2008

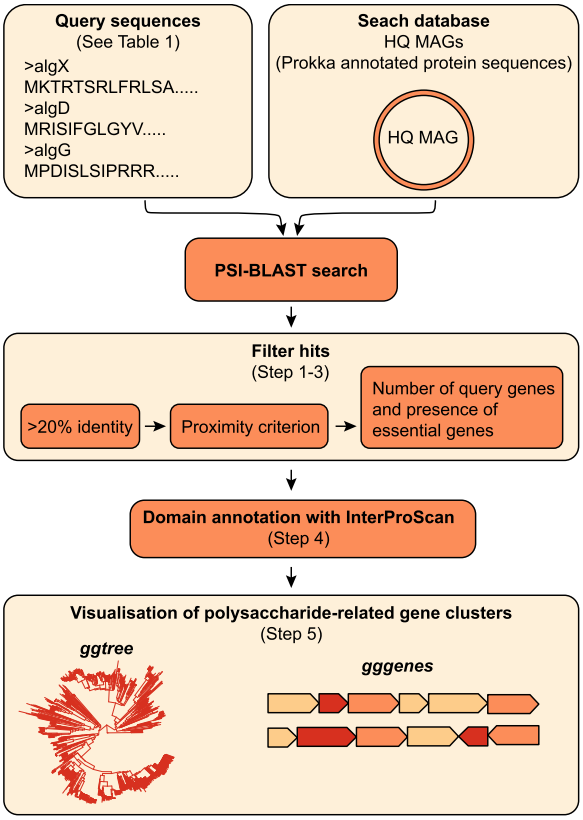


Fig. 1. Workflow of the polysaccharide-related gene cluster prediction pipeline. Protein homologs to query sequences in the HQ MAG database were identified with PSI-BLAST. The resulting homologs were filtered by percent identity, proximity, number of genes, and presence of essential genes. The remaining homologs, which constitute putative gene clusters, were domain annotated with InterProScan. Gene cluster structures and taxonomic distribution of the identified gene clusters were visualized in R.

enhanced in Adobe Illustrator 2022.

2.4. Data and code availability

R-scripts used for data analyses and figures are available at GitHub: https://github.com/msdueholm/Publications/tree/master/Dueholm2022b. Raw data for the scripts is available at figshare: https://doi.org/10.6084/m9.figshare.21287112.v1. A list of all identified EPS gene clusters is provided as Supplementary Data S1.

3. Results and discussion

3.1. Exopolysaccharide biosynthesis systems in AS bacteria

To identify potential exopolysaccharide biosynthesis systems in the HQ MAGs, we first searched for gene clusters associated with the biosynthesis of known exopolysaccharides. This resulted in a list of 16 validated exopolysaccharide systems, each consisting of a set of query sequences (Table 1). An initial homology search using the translated protein sequences from these query genes was conducted against the translated coding sequences (CDS) from the HQ MAGs using PSI-BLAST. However, after removing hits with low percent identity (<20%), homolog proteins related to all exopolysaccharide systems were still identified in almost all MAGs, suggesting that many of the query genes had homologs associated with unrelated metabolic pathways (Fig. 2a). This was also apparent from the observation that some query sequences from different exopolysaccharide systems targeted the same CDSs in the HQ MAGs (**Fig. S1**).

To increase the specificity of our search, we introduced a proximity criterion, utilizing the fact that functionally related genes, including those encoding exopolysaccharides, are generally organized in operons or BGCs in bacteria (Schmid et al., 2015) (Fig. 2b). This filtration proved to be highly effective in identifying complete or near-complete exopolysaccharide BGCs while removing hits, which are likely associated with unrelated metabolic pathways. To only include the most plausible hits, we considered a pathway to be present if all essential genes (e.g., the main synthase unit) were present while complying with the proximity criterion and the threshold for the number of genes for the given polysaccharide biosynthesis pathway. Generally, homologs of some genes related to a specific polysaccharide were found in many MAGs. In contrast, other genes were more specific and were mainly seen with a high percent identity in MAGs with a complete biosynthesis pathway (Fig. 2a **and Fig. S2-S8**). Thus, some genes are good predictors for a putative biosynthetic gene cluster. We identified homologs for 10 out of the 16 examined exopolysaccharide biosynthetic pathways (**Supplementary Data S1**).

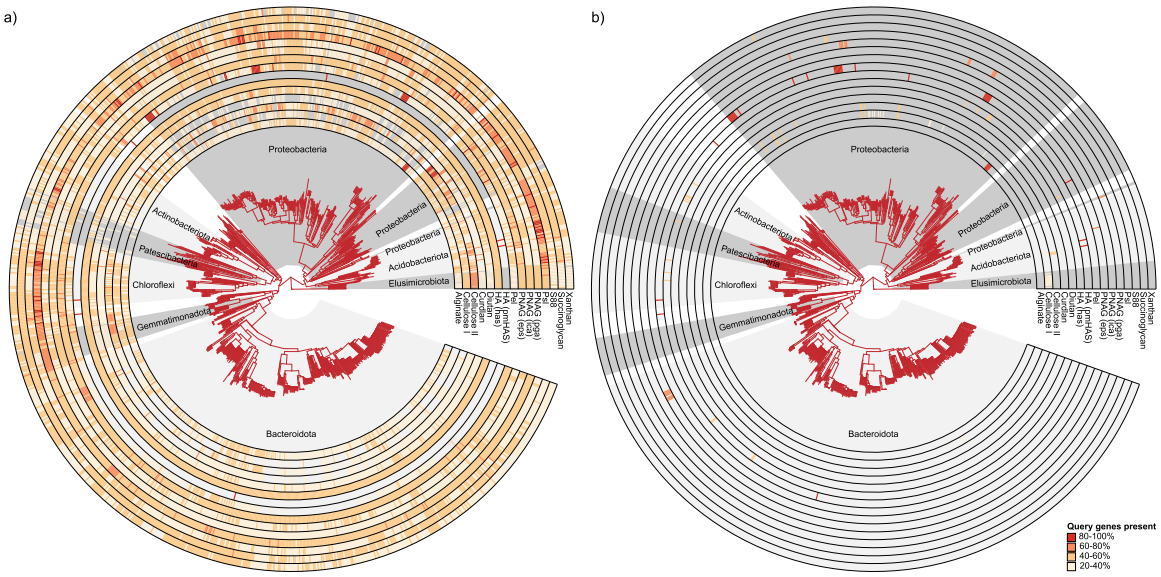


Fig. 2. Phylogenetic distribution of identified polysaccharide-related gene clusters. HQ MAG tree from Singleton et al. (2021) annotated with a heatmap of PSI-BLAST hits a) before and b) after automated proximity filtration using the criteria specified in Table 1.

3.2. Alginate BGCs are uncommon in AS bacteria

Alginate is an anionic linear polysaccharide composed of mannuronic acid and guluronic acid residues (Remminghorst and Rehm, 2006). Putative gene clusters for alginate synthesis were only identified in five MAGs. These belong to the two MiDAS placeholder genera midas_g_2943 (3 MAGs) and midas_g_4210 (2 MAGs), both in the Micavibrionales order of the Proteobacteria phylum (**Fig. S2**). Interestingly, several members of this order, including *Micavibrio*, are epiphytic, obligate bacterial parasites that, among others, prey on *Pseudomonas aeruginosa* from which the query operon originates (Wang et al., 2011). Accordingly, the alginate BGC may have been acquired by horizontal gene transfer because of this lifestyle. The alginate-related genes were organized as gene clusters consisting of both operons and individually transcribed genes (Fig. 3). This contrasts with the single-operon organization of the query genes in *P. aeruginosa*. The organization of the alginate gene clusters was genus specific. The midas_g_4210 gene clusters lacked the *algG* gene, proposed to be an epimerase (Schmid et al., 2015), whereas those for midas_g_2943 had homologs for all query genes. The latter also encoded additional genes, including three copies of the putative acetylation gene *algX*. Because the genetic potential for alginate was restricted to a few low-abundant bacteria, it is unlikely that alginate is an important polymer in WWTPs. Furthermore, it suggests that the main polysaccharide component in ALE is evolutionarily unrelated to alginate, although sharing physicochemical properties.

3.3. Pel operons are found in several orders within the Proteobacteria

Pel is a known constituent in biofilms produced by *P. aeruginosa* (Friedman and Kolter, 2003) and is composed of dimeric repeats of α−1, 4 linked galactosamine and *N*-acetylgalactosamine (Le Mauff et al., 2022). The genetic capacity to synthesize Pel is encoded in the seven-gene *pel*-operon (*pelA-G*), widespread among Gram-negative and Gram-positive bacteria (Whitfield and Howell, 2021). We identified 25 MAGs with the genetic capacity to produce Pel (**Fig. S3**). These MAGs were all, except for one Myxococcota MAG, from the Proteobacteria phylum, and represented the following orders: The Burkholderiales (18 MAGs), Pseudomonadales (5 MAGs), Oceanospirillales (1 MAG), and Myxococcales (1 MAG). The *pel* genes were for all MAGs organized as a single operon, and gene synteny matched that of the query operon from *P. aeruginosa* except for the Myxococcales MAG (Fig. 4). In the latter, the *pelA* and *pelB* genes were located downstream from *pelG* as the last genes in the operon compared to at the beginning in operons from the Burkholderiales and Pseudomonadales order. The presence of the Pel biosynthetic machinery in 25 MAGs representing 12 species, including a general core (*Lautropia* midas_s_135) and three conditional rare or abundant taxa (CRAT) species across the world (*Propionivibrio aalborgensis; Propionivibrio* midas_s_421, and *Hahella* midas_s_10,928) (**Table S1**) suggests that Pel could be important in global wastewater treatment systems.

3.4. Cellulose BGCs display diverse structures

Cellulose is a polymer of β−1–4-linked D-glucose and a common

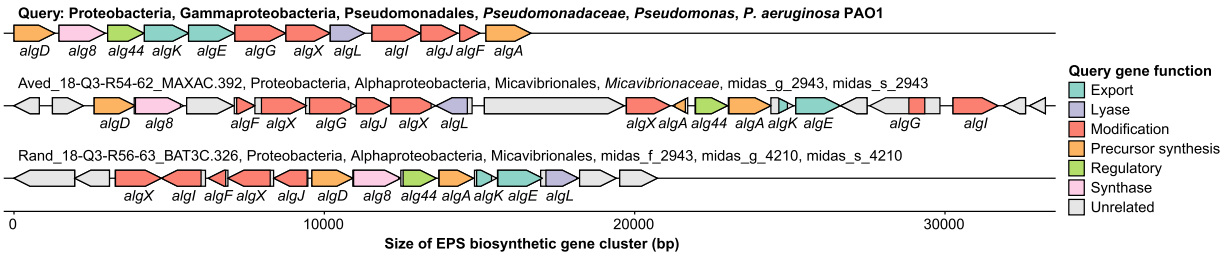


Fig. 3. Alginate biosynthetic gene clusters. Overview of the alginate query gene cluster and the two putative gene cluster types (midas_g_2943 and midas_g_4210), including two flanking genes on both sides. Each putative gene cluster is shown with MAG name and the assigned MiDAS 4.8.1 taxonomy. Genes are colored based on predicted functions according to the query sequence.

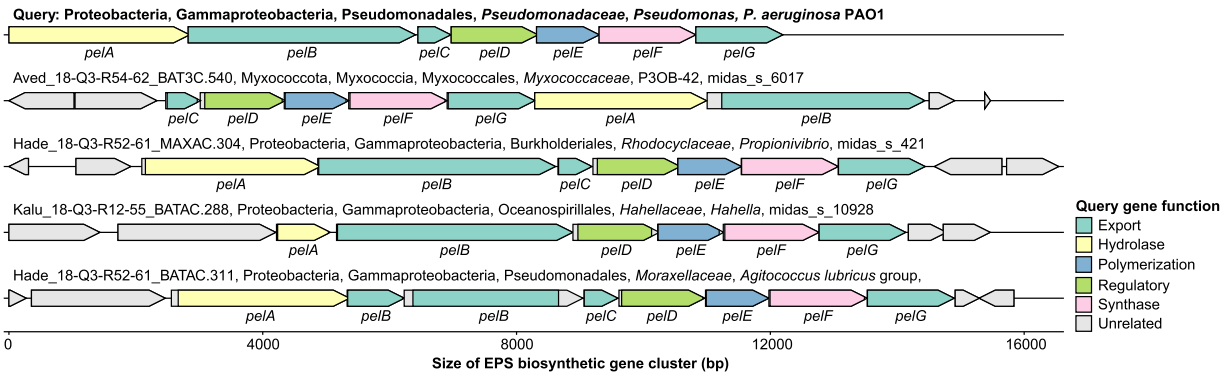


Fig. 4. Pel biosynthetic gene clusters. Overview of the Pel query gene cluster and an example of a putative gene cluster from each order (Myxococcales, Burkholderiales, Oceanospirillales, and Pseudomonadales), including two flanking genes on both sides. Each putative gene cluster is shown with MAG name and the assigned MiDAS 4.8.1 taxonomy. Genes are colored based on predicted functions according to the query sequence.

constituent of wastewater originating from, e.g., toilet paper (Ruiken et al., 2013). However, many bacteria can also produce cellulose as a component of their extracellular matrix (Römling and Galperin, 2015). The presence of cellulose in the inlet wastewater makes it challenging to differentiate microbial-produced cellulose, necessitating genomic or protein investigations to resolve its origin. Previous studies have shown that the AS isolate *Shinella zoogloeoides* ATCC 19,623 (formerly known as *Zoogloea ramigera* I-16-M) produces and uses cellulose for floc formation (Gao et al., 2022). Otherwise, little is known about microbial-produced cellulose in wastewater treatment systems.

Komagataeibacter medellinensis (formerly known as *Gluconacetobacter xylinus* and *Acetobacter xylinum*) is a model organism for bacterial cellulose, and it can produce two different forms (Matsutani et al., 2015). Cellulose I is a ribbon-like polymer composed of bundles of microfibrils, whereas cellulose II is an amorphous polymer with increased thermostability compared to cellulose I (Yu and Atalla, 1996). The two forms of cellulose are encoded by separate BGCs (Matsutani et al., 2015). We used both as queries to identify MAGs with the genetic potential for cellulose production. 43 MAGs were identified when a minimum threshold of two observed homologs, including the main synthase gene (*bcsAI* or *bcsABII*), were used for identification (**Fig. S4 and S5**). Out of these, 30 MAGs encoded only cellulose I, three MAGs only cellulose II, and ten MAGs both cellulose I and II. These MAGs included four loose core and two CRAT species, including *Acidovorax* midas_s_1484, *Dechloromonas* midas_s_173 recently classified as *D. phosphoritropha* (Petriglieri et al., 2021), and *Nitrotoga* midas_s_181, which all encode both cellulose types (**Table S1**).

Because the genomic potential for cellulose I expression was more common in the examined MAGs, we analyzed this system in greater detail (Fig. 5). Interestingly, critical Pfam domains present in the query genes were only detected in the homologs of 14 MAGs, which could indicate that some detected operons were non-functional. The operons from the 14 MAGs could be separated into three groups depending on gene organization and flanking genes: *Agrobacterium tumefaciens*-like (1 MAG), *K. medellinensis*-like (5 MAGs), and *Escherichia coli*-like (8 MAGs) (Fig. 5). The *E. coli*-like hits contained a *bcsQ* and a *bcsG* gene and had similar operon structures. BcsG has been identified as a phosphoethanolamine transferase involved in producing modified cellulose with these groups added (Römling and Galperin, 2015). The *K. medellinensis*-like operon structures contained a *bcsD* gene in two

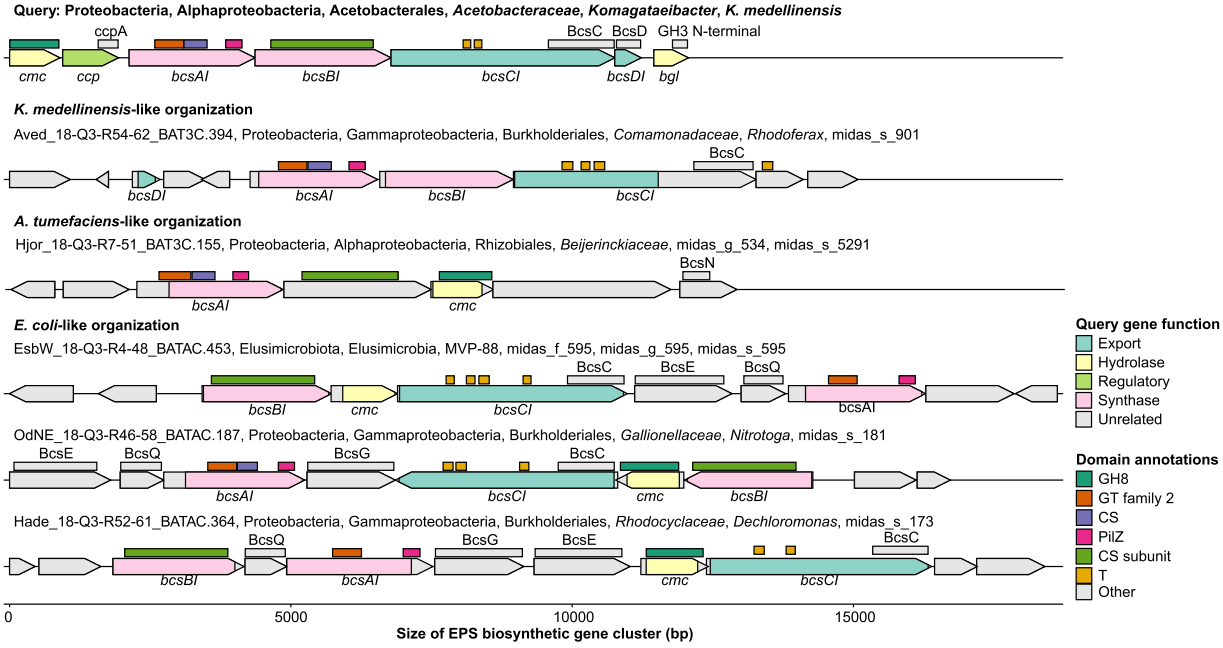


Fig. 5. Cellulose biosynthetic gene clusters. Overview of the cellulose I query operon with domains and examples of identified BGCs including two flanking genes on both sides. Each putative BGC is shown with MAG name and the assigned MiDAS 4.8.1 taxonomy. Genes are colored based on predicted functions according to the query sequence. Domain labels are abbreviated in line with Pfam. Domain labels unrelated to cellulose synthesis have been removed.

MAGs. BscD is often coupled to the crystalline cellulose structure found in *K. medellinensis* (Römling and Galperin, 2015). The *A. tumefaciens*-like operon uniquely encoded the *bcsN* gene, which was previously shown to be specific for Alphaproteobacteria (Römling and Galperin, 2015).

Cellulose is known to be regulated by c-di-GMP via its binding to conserved motifs in the PilZ domain in BcsA (McNamara et al., 2015). The *A. tumefaciens*-like operon is distinguished by the absence of a functional important amino acid in the PilZ domain, which has previously been coupled to c-di-GMP independent regulation (Matthysse et al., 1995b, 1995a), suggesting different regulatory strategies.

The diverse organization of the identified cellulose operons implies differentiation of their regulation, monomer composition, and the structural organization of cellulose fibrils. Due to their uncharacterized nature, it is currently a challenge to couple the physicochemical characteristics of microbial cellulose to the different cellulose synthase complexes (McNamara et al., 2015). However, the genomic potential for cellulose production in several MAGs, including core taxa, supports the role of microbial cellulose as an important EPS in AS floc formation.

3.5. The pga-operon is favored for PNAG biosynthesis in AS

Poly-N-acetylglucosamine (PNAG) is an EPS polysaccharide associated with microbial biofilms. It is a homopolysaccharide consisting of β−1,6-linked N-acetylglucosamine and is synthesized in a range of bacterial species, including both Gram-negative (*pgaABCD* operon in *E. coli*) and Gram-positive species (*icaADBC* operon in *Staphylococcus* species and *epsHIJK* operon in *B. subtilis*) (Atkin et al., 2014; Bundalovic-Torma et al., 2020; Roux et al., 2015).

The Gram-negative PNAG (*pga*) query identified putative operons in 12 MAGs (**Fig. S6**). No homologs to *pgaD* were found by PSI-BLAST, but in all operons, a 'PgaD-like protein' domain was identified based on InterProScan. Organization and domain annotation were similar across all identified operons. However, all hits from the Betaproteobacteriales order lacked an annotated GH13 domain within the *pgaB* gene, dividing the hits into two operon types: one with an assigned GH13 and one without (Fig. 6). The GH13 domain in PgaB has previously been proposed to play a role in binding and translocating PNAG through the periplasmic space (Little et al., 2014). However, it may also act as an extracellular hydrolase that can disperse PNAG-dependent biofilms (Little et al., 2018). All hits were identified in the Proteobacteria phylum within the Gammaproteobacteria (11 MAGs) and Desulfobacterota (1 MAG) classes. Similarly, Bundalovic-Torma et al. (2020) showed that PNAG is widely represented in the Proteobacteria phylum but is also found in other phyla. Among the identified MAGs were the two loose core species *Ca.* Accumulibacter phosphatis and *Sulfuritalea hydrogenivorans* and the CRAT species *Ca.* Accumulibacter midas_s_315, which was recently reclassified as *Ca.* Propionivibrio dominans (**Table S1**) (Petriglieri et al., 2022).

The Gram-positive PNAG (*eps*) genes are often observed in larger gene clusters (*epsA-O*), but only *epsHIJK* are required for biosynthesis. When filtered to three required genes, only eight hits remained. These belonged to the Bacteroidetes (7 MAGs) and Acidobacterium (1 MAG) phyla (**Fig. S7**). The Bacteroidetes MAGs included a single CRAT species representing a MiDAS placeholder order (midas_o_31) within the class SJA-28 (**Table S1**). The modification gene *epsI* was not identified in any of the MAGs, and in most, the annotated *epsK* genes were only partially aligned to the query, which questions the functionality of these operons. We also searched for homologous synthesis systems for the *Staphylococcus*-like PNAG (*ica*) homologous synthesis systems, but no operons were identified.

3.6. BGCs for several known Wzx/Wzy-dependent polysaccharides were not identified

Generally, our approach did not identify any complete biosynthetic pathway (≥75% of query genes) for xanthan, salecan, Psl, diutan, S88, and succinoglycan (Fig. 2). These all represent the Wzx/Wzy-dependent pathway and include at least 12 genes. The large number of genes provides increased room for diversification and gene shuffling, which may have prevented us from detecting such operons. In support of this, we found that the genetic loci associated with the only three Wzx/Wzy-dependent pathways detected in the MAGs were rearranged entirely (Fig. 7). Previous studies, carried out on *Lactobacillus,* have furthermore demonstrated that diversification of Wzx/Wzy-dependent EPS gene cluster can occur at the strain level (Deo et al., 2019; Martino et al., 2016). If this is also the case for strains in AS, we might expect that such gene clusters may not be correctly assembled in some of the HQ MAGs as these may represent species-level co-assemblies of several strains (Meziti et al., 2021). This questions whether the larger Wzx/Wzy-dependent gene clusters are not detected due to strain diversity or are simply not present in the AS bacteria.

3.7. Limited potential for hyaluronic acid and curdlan biosynthesis

HA-like polymers are produced by HA-synthases and have previously been reported in aerobic granular sludge, where they are assumed to form hydrogels (Felz et al., 2020). HA-synthases are divided into two distinct classes based on their configuration, the HA (*has*) and HA (*pmHAS*). We identified homologs of proteins associated with HA (*has*) in 20 MAGs. However, the organization of the genes questions whether they are functional for HA synthesis. The query is a short three-gene operon (Sze et al., 2016), whereas the identified homologs were separated by multiple genes and often present on separate strands of the DNA. Accordingly, the identified hits might originate from conserved domains also found in genes related to other pathways utilizing the same substrates. HA (*pmHAS*) only contained a single query gene. We identified homologs in seven MAGs from different phyla. However, Pfam domain annotation did not identify all the domains required for HA-synthesis in any of the hits, and it may therefore be speculated whether these homologs are functional or not. We also searched for curdlan BGCs but did not detect any.

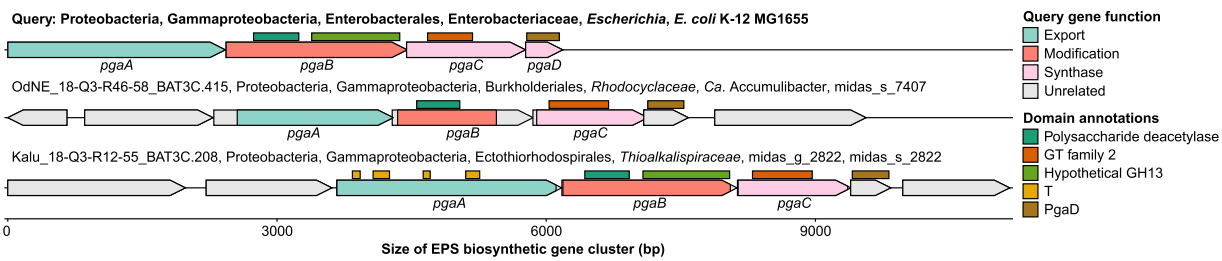


Fig. 6. PNAG biosynthetic gene clusters. Overview of a PNAG query operon and examples of the two types of putative operons (with or without the GH13 domain of *pgaB*), including two flanking genes on both sides. Each putative gene cluster is shown with MAG name and the assigned MiDAS 4.8.1 taxonomy. Genes are colored after the predicted function. Domain labels are abbreviated in line with Pfam. Domain labels unrelated to PNAG synthesis have been removed.

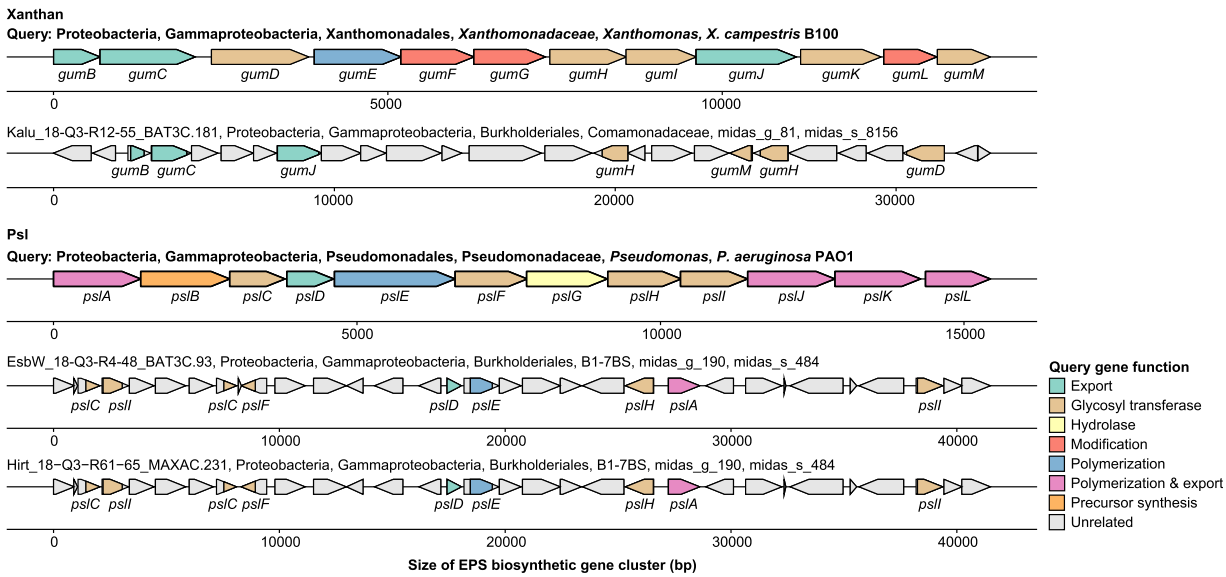


Fig. 7. Xanthan and Psl biosynthetic gene clusters. Overview of the xanthan and Psl query operon and detected putative BGCs, including two flanking genes on both sides. Each putative BGC is shown with MAG name and the assigned MiDAS 4.8.1 taxonomy. Genes are colored based on predicted functions according to the query sequence.

3.8. A genetic approach to identifying exopolysaccharide potential

The experimental validation of EPS polysaccharides in complex bacterial communities such as WWTPs is highly challenging (Seviour et al., 2019). Thus, approaches to identifying genomic potentials may provide a possible starting point for a better understanding of biofilm and floc formation in the context of WWTPs. Furthermore, the genomic potentials can be linked to bacterial taxonomies, which could facilitate a more refined regulation of, e.g., floc formation by altering conditions in WWTPs to select for specific microbial taxa.

In this work, we used a bioinformatic approach to identify potential exopolysaccharide synthesis operons or gene clusters incorporating genomic proximity. The strategy proved extremely powerful in finding homologous systems of the synthase-dependent exopolysaccharide synthesis pathways and relies only on the availability of HQ MAGs or genomes and that the relevant pathways have clustered and conserved genes. Thus, this method can easily be extended to similar pathways in any ecosystem.

3.9. Limited genomic potential for biosynthesis of known exopolysaccharides

We identified biosynthetic gene clusters associated with known exopolysaccharides in 121 MAGs, corresponding to 11% of the HQ MAGs obtained by Singleton et al. (2021). This suggests that it is either a subset of the abundant bacteria in Danish WWTPs that can produce exopolysaccharides, or that there are exopolysaccharide systems that were not detected using our approach. In support of the latter, it should be noted that while this study provides information on the genomic presence of known polysaccharide synthesis pathways, the method does not capture potential novel polysaccharide synthesis systems not covered by the queries. To identify novel polysaccharide synthesis systems, the search can be expanded with additional genes for general synthesis proteins from the three main mechanisms of polysaccharide synthesis (ABC transporter, synthase-, and Wzx/Wzy-dependent) as new systems could be expected to have conserved key genes homologous to those already known. However, this approach requires fine-tuning of the method.

Putative gene clusters related to 10 out of the 16 examined exopolysaccharide BGCs were identified in our MAGs, including MAGs that represent core or CRAT species found in WWTPs across the globe (Dueholm et al., 2022) (**Table S1**). The most common type of exopolysaccharide gene cluster was cellulose (40 MAGs) followed by Pel (25 MAGs). Most MAGs encoded only a single type of exopolysaccharide, except for cellulose, where several MAGs encoded both type I and II. Interestingly, gene clusters for specific exopolysaccharides were generally conserved across MAGs from the same species (**Table S1**). Accordingly, we may be able to link the potential for exopolysaccharide biosynthesis to specific species. It should be stressed that the current study only determined genomic potentials. Therefore, future research must confirm that the predicted exopolysaccharide systems are important for the AS floc formation and, if they are, how they are regulated in response to process and environmental factors.

3.10. A platform for in situ studies

Our approach has revealed the genomic potential for known exopolysaccharide biosynthesis systems in AS. Thus, it represents an excellent starting point for in situ studies, which potentially can verify the systems and help characterize their biofilm dynamics. Various imaging techniques can link the production of specific exopolysaccharides with specific bacterial taxa and spatially locate key producers in the AS flocs. This knowledge can prove important in understanding and characterizing the role of the particular exopolysaccharides in floc formation to improve solid-liquid separation.

To potentially regulate the floc assembly in the future, transcriptomics can be used to study the triggers of exopolysaccharide synthesis and identify the most important producers. Temporal studies with altering conditions can potentially reveal regulatory mechanisms, which have huge potential in controlling flocculation and settlement. Furthermore, proteomics can be used to verify the expression and quantification of relevant proteins, which can indicate the importance of a given exopolysaccharide for the biofilm matrix, which is critical to the process management of the WWTPs.

4. Conclusions

We have here applied genome mining and gene synteny analyses to uncover the genetic potential for exopolysaccharide synthesis in more than 1000 HQ MAGs, representing typical species found in AS systems

globally. This provided a direct link between individual taxa, their potential exopolysaccharides, and the associated BGCs. However, as only 11% of the examined MAGs encoded homologs to known exopolysaccharide gene clusters, and most only carried the genes for one system, there might be an unexplored potential for novel exopolysaccharide systems in the many poorly studied taxa found in WWTPs. Characterization of the genomic potential for EPS biosynthesis represents a foundation for further exploration into what exopolysaccharides are expressed in situ by which bacteria, and how expression is regulated in response to process and environmental factors. This information may ultimately be used to regulate exopolysaccharide production to improve the properties of AS flocs or enhance the value of the EPS as extractable renewable bioresources.

Declaration of Competing Interest

The authors declare no conflict of interest.

Data availability

Links to all data and codes used are provide in the materials and methods.

Acknowledgments

The project has been funded by the Danish Research Council (grant 6111–00617A), the Villum Foundation (Dark Matter, grant 16558) and the Novo Nordisk Foundation (REThiNk, grant NNF22OC0071498)

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2022.119485.

References

Ajao, V., Bruning, H., Rijnaarts, H., Temmink, H., 2018. Natural flocculants from fresh and saline wastewater: comparative properties and flocculation performances. Chem. Eng. J. 349, 622–632. https://doi.org/10.1016/j.cej.2018.05.123.

Allen, M.S., Welch, K.T., Prebyl, B.S., Baker, D.C., Meyers, A.J., Sayler, G.S., 2004. Analysis and glycosyl composition of the exopolysaccharide isolated from the floc-forming wastewater bacterium *Thauera* sp. MZ1T. Environ. Microbiol. 6, 780–790. https://doi.org/10.1111/j.1462-2920.2004.00615.x.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic. Acids. Res. 25, 3389–3402. https://doi.org/10.1093/nar/25.17.3389.

An, W., Guo, F., Song, Y., Gao, N., Bai, S., Dai, J., Wei, H., Zhang, L., Yu, D., Xia, M., Yu, Y., Qi, M., Tian, C., Chen, H., Wu, Z., Zhang, T., Qiu, D., 2016. Comparative genomics analyses on EPS biosynthesis genes required for floc formation of *Zoogloea resiniphila* and other activated sludge bacteria. Water Res. 102, 494–504. https://doi.org/10.1016/j.watres.2016.06.058.

Atkin, K.E., MacDonald, S.J., Brentnall, A.S., Potts, J.R., Thomas, G.H., 2014. A different path: revealing the function of staphylococcal proteins in biofilm formation. FEBS Lett. 588, 1869–1872. https://doi.org/10.1016/j.febslet.2014.04.002.

Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A., Tringe, S.G., Ivanova, N. N., Copeland, A., Clum, A., Becraft, E.D., Malmstrom, R.R., Birren, B., Podar, M., Bork, P., Weinstock, G.M., Garrity, G.M., Dodsworth, J.A., Yooseph, S., Sutton, G., Glöckner, F.O., Gilbert, J.A., Nelson, W.C., Hallam, S.J., Jungbluth, S.P., Ettema, T.J. G., Tighe, S., Konstantinidis, K.T., Liu, W.T., Baker, B.J., Rattei, T., Eisen, J.A., Hedlund, B., McMahon, K.D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Tyson, G.W., Rinke, C., Lapidus, A., Meyer, F., Yilmaz, P., Parks, D.H., Eren, A.M., Schriml, L., Banfield, J.F., Hugenholtz, P., Woyke, T., 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat. Biotechnol. 35, 725–731. https://doi.org/10.1038/nbt.3893.

Bundalovic-Torma, C., Whitfield, G.B., Marmont, L.S., Howell, P.L., Parkinson, J., 2020. A systematic pipeline for classifying bacterial operons reveals the evolutionary landscape of biofilm machineries. PLOS Comput. Biol. 16, e1007721 https://doi.org/10.1371/journal.pcbi.1007721.

Coleman, R.J., Patel, Y.N., Harding, N.E., 2008. Identification and organization of genes for diutan polysaccharide synthesis from *Sphingomonas* sp. ATCC 53159. J. Ind. Microbiol. Biotechnol. 35, 263–274. https://doi.org/10.1007/s10295-008-0303-3.

DeAngelis, P.L., Jing, W., Drake, R.R., Achyuthan, A.M., 1998. Identification and molecular cloning of a unique hyaluronan synthase from *Pasteurella multocida*. J. Biol. Chem. 273, 8454–8458. https://doi.org/10.1074/jbc.273.14.8454.

Deo, D., Davray, D., Kulkarni, R., 2019. A diverse repertoire of exopolysaccharide biosynthesis gene clusters in *Lactobacillus* revealed by comparative analysis in 106 sequenced genomes. Microorganisms 7, 444. https://doi.org/10.3390/microorganisms7100444.

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. Nat. Commun. 13, 1908. https://doi.org/10.1038/s41467-022-29438-7.

Felz, S., Neu, T.R., van Loosdrecht, M.C.M., Lin, Y., 2020. Aerobic granular sludge contains Hyaluronic acid-like and sulfated glycosaminoglycans-like polymers. Water Res. 169, 115291 https://doi.org/10.1016/j.watres.2019.115291.

Feng, C., Lotti, T., Canziani, R., Lin, Y., Tagliabue, C., Malpei, F., 2021. Extracellular biopolymers recovered as raw biomaterials from waste granular sludge and potential applications: a critical review. Sci. Total Environ. 753, 142051 https://doi.org/10.1016/j.scitotenv.2020.142051.

Finan, T.M., Weidner, S., Wong, K., Buhrmester, J., Chain, P., Vorhölter, F.J., Hernandez-Lucas, I., Becker, A., Cowie, A., Gouzy, J., Golding, B., Pühler, A., 2001. The complete sequence of the 1,683-kb pSymB megaplasmid from the N_2-fixing endosymbiont *Sinorhizobium meliloti*. Proc. Natl. Acad. Sci. 98, 9889–9894. https://doi.org/10.1073/pnas.161294698.

Flemming, H.-C., van Hullebusch, E.D., Neu, T.R., Nielsen, P.H., Seviour, T., Stoodley, P., Wingender, J., Wuertz, S., 2022. The biofilm matrix: multitasking in a shared space. Nat. Rev. Microbiol. 1–17. https://doi.org/10.1038/s41579-022-00791-0.

Flemming, H.-C., Wingender, J., 2010. The biofilm matrix. Nat. Rev. Microbiol. 8, 623–633. https://doi.org/10.1038/nrmicro2415.

Flemming, H.-C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S.A., Kjelleberg, S., 2016. Biofilms: an emergent form of bacterial life. Nat. Rev. Microbiol. 14, 563–575. https://doi.org/10.1038/nrmicro.2016.94.

Franklin, M.J., Nivens, D.E., Weadge, J.T., Lynne Howell, P., Howell, P.L., 2011. Biosynthesis of the *Pseudomonas aeruginosa* extracellular polysaccharides, alginate, Pel, and Psl. Front. Microbiol. 2, 1–16. https://doi.org/10.3389/fmicb.2011.00167.

Friedman, L., Kolter, R., 2003. Genes involved in matrix formation in *Pseudomonas aeruginosa* PA14 biofilms. Mol. Microbiol. 51, 675–690. https://doi.org/10.1046/j.1365-2958.2003.03877.x.

Frølund, B., Palmgren, R., Keiding, K., Nielsen, P.H., 1996. Extraction of extracellular polymers from activated sludge using a cation exchange resin. Water Res. 30, 1749–1758. https://doi.org/10.1016/0043-1354(95)00323-1.

Gao, N., Dai, J., Liu, Y., Li, S., Wang, J., Lu, W., Qiu, D., 2022. Cellulose-mediated floc formation by the activated sludge bacterium *Shinella zoogloeoides* ATCC 19623. BMC Microbiol. 22, 104. https://doi.org/10.1186/s12866-022-02516-y.

Gao, N., Xia, M., Dai, J., Yu, D., An, W., Li, S., Liu, S., He, P., Zhang, L., Wu, Z., Bi, X., Chen, S., Haft, D.H., Qiu, D., 2018. Both widespread PEP-CTERM proteins and exopolysaccharides are required for floc formation of *Zoogloea resiniphila* and other activated sludge bacteria. Environ. Microbiol. 20, 1677–1692. https://doi.org/10.1111/1462-2920.14080.

Gill, S.R., Fouts, D.E., Archer, G.L., Mongodin, E.F., DeBoy, R.T., Ravel, J., Paulsen, I.T., Kolonay, J.F., Brinkac, L., Beanan, M., Dodson, R.J., Daugherty, S.C., Madupu, R., Angiuoli, S.V., Durkin, A.S., Haft, D.H., Vamathevan, J., Khouri, H., Utterback, T., Lee, C., Dimitrov, G., Jiang, L., Qin, H., Weidman, J., Tran, K., Kang, K., Hance, I.R., Nelson, K.E., Fraser, C.M., 2005. Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. J. Bacteriol. 187, 2426–2438. https://doi.org/10.1128/JB.187.7.2426-2438.2005.

Goldstein, S., Beka, L., Graf, J., Klassen, J.L., 2019. Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing. BMC Genomics 20, 23. https://doi.org/10.1186/s12864-018-5381-7.

Gouy, M., Milleret, F., Mugnier, C., Jacobzone, M., Gautier, C., 1984. ACNUC: a nucleic acid sequence data base and analysis system. Nucleic. Acids. Res. 12, 121–127. https://doi.org/10.1093/nar/12.1Part1.121.

Ikeda, F., Shuto, H., Saito, T., Fukui, T., Tomita, K., 1982. An extracellular polysaccharide produced by *Zoogloea ramigera* 115. Eur. J. Biochem. 123, 437–445. https://doi.org/10.1111/j.1432-1033.1982.tb19787.x.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

Karakas, I., Sam, S.B., Cetin, E., Dulekgurgen, E., Yilmaz, G., 2020. Resource recovery from an aerobic granular sludge process treating domestic wastewater. J. Water Process Eng. 34, 101148 https://doi.org/10.1016/j.jwpe.2020.101148.

Karnezis, T., Epa, V.C., Stone, B.A., Stanisich, V.A., 2003. Topological characterization of an inner membrane (1→3)-β-D-glucan (curdlan) synthase from *Agrobacterium* sp. strain ATCC31749. Glycobiology 13, 693–706. https://doi.org/10.1093/glycob/cwg093.

Karr, P.R., Keinath, T.M., 1978. Influence of particle size on sludge dewaterability. J. Water Pollut. Control Fed. 50, 1911–1930.

Kim, N.K., Mao, N., Lin, R., Bhattacharyya, D., van Loosdrecht, M.C.M., Lin, Y., 2020. Flame retardant property of flax fabrics coated by extracellular polymeric substances recovered from both activated sludge and aerobic granular sludge. Water Res. 170, 115344 https://doi.org/10.1016/j.watres.2019.115344.

Koren, S., Phillippy, A.M., 2015. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. Curr. Opin. Microbiol., Host–microbe interactions: bacteria • Genomics 23, 110–120. https://doi.org/10.1016/j.mib.2014.11.014.

Le Mauff, F., Razvi, E., Reichhardt, C., Sivarajah, P., Parsek, M.R., Howell, P.L., Sheppard, D.C., 2022. The Pel polysaccharide is predominantly composed of a dimeric repeat of α-1,4 linked galactosamine and N-acetylgalactosamine. Commun. Biol. 5, 1–7. https://doi.org/10.1038/s42003-022-03453-2.

Lin, Y., de Kreuk, M., van Loosdrecht, M.C.M., Adin, A., 2010. Characterization of alginate-like exopolysaccharides isolated from aerobic granular sludge in pilot-plant. Water Res. 44, 3355–3364. https://doi.org/10.1016/j.watres.2010.03.019.

Lin, Y.M., Sharma, P.K., van Loosdrecht, M.C.M., 2013. The chemical and mechanical differences between alginate-like exopolysaccharides isolated from aerobic flocculent sludge and aerobic granular sludge. Water Res. 47, 57–65. https://doi.org/10.1016/j.watres.2012.09.017.

Little, D.J., Li, G., Ing, C., DiFrancesco, B.R., Bamford, N.C., Robinson, H., Nitz, M., Pomès, R., Howell, P.L., 2014. Modification and periplasmic translocation of the biofilm exopolysaccharide poly-β-1,6-N-acetyl-D-glucosamine. Proc. Natl. Acad. Sci. U. S. A. 111, 11013–11018. https://doi.org/10.1073/pnas.1406388111.

Little, D.J., Pfoh, R., Le Mauff, F., Bamford, N.C., Notte, C., Baker, P., Guragain, M., Robinson, H., Pier, G.B., Nitz, M., Deora, R., Sheppard, D.C., Howell, P.L., 2018. PgaB orthologues contain a glycoside hydrolase domain that cleaves deacetylated poly-β(1,6)-N-acetylglucosamine and can disrupt bacterial biofilms. PLoS Pathog. 14, e1006998 https://doi.org/10.1371/journal.ppat.1006998.

Liu, Y., Fang, H.H.P., 2003. Influences of extracellular polymeric substances (EPS) on flocculation, settling, and dewatering of activated sludge. Crit. Rev. Environ. Sci. Technol. 33, 237–273. https://doi.org/10.1080/10643380390814479.

Ma, L., Jackson, K.D., Landry, R.M., Parsek, M.R., Wozniak, D.J., 2006. Analysis of *Pseudomonas aeruginosa* conditional Psl variants reveals roles for the Psl polysaccharide in adhesion and maintaining biofilm structure postattachment. J. Bacteriol. 188, 8213–8221. https://doi.org/10.1128/JB.01202-06.

Ma, Z., Geng, J., Zhang, H., Yu, H., Yi, L., Lei, M., Lu, C., Fan, H., Hu, S., 2011. Complete genome sequence of *Streptococcus equi* subsp. *zooepidemicus* strain ATCC 35246. J. Bacteriol. 193, 5583–5584. https://doi.org/10.1128/JB.05700-11.

Martino, M.E., Bayjanov, J.R., Caffrey, B.E., Wels, M., Joncour, P., Hughes, S., Gillet, B., Kleerebezem, M., van Hijum, S.A.F.T., Leulier, F., 2016. Nomadic lifestyle of *Lactobacillus plantar*um revealed by comparative genomics of 54 strains isolated from different habitats. Environ. Microbiol. 18, 4974–4989. https://doi.org/10.1111/1462-2920.13455.

Matsutani, M., Ito, K., Azuma, Y., Ogino, H., Shirai, M., Yakushi, T., Matsushita, K., 2015. Adaptive mutation related to cellulose producibility in *Komagataeibacter medellinensis* (*Gluconacetobacter xylinus*) NBRC 3288. Appl. Microbiol. Biotechnol. 99, 7229–7240. https://doi.org/10.1007/s00253-015-6598-x.

Matthysse, A.G., Thomas, D.L., White, A.R., 1995a. Mechanism of cellulose synthesis in *Agrobacterium tumefaciens*. J. Bacteriol. 177, 1076–1081. https://doi.org/10.1128/jb.177.4.1076-1081.1995.

Matthysse, A.G., White, S., Lightfoot, R., 1995b. Genes required for cellulose synthesis in *Agrobacterium tumefaciens*. J. Bacteriol. 177, 1069–1075. https://doi.org/10.1128/jb.177.4.1069-1075.1995.

McNamara, J.T., Morgan, J.L.W., Zimmer, J., 2015. A molecular description of cellulose biosynthesis. Annu. Rev. Biochem. 84, 895–921. https://doi.org/10.1146/annurev-biochem-060614-033930.

Meziti, A., Rodriguez-R, L.M., Hatt, J.K., Peña-Gonzalez, A., Levy, K., Konstantinidis, K. T., 2021. The reliability of metagenome-assembled genomes (MAGs) in representing natural populations: insights from comparing MAGs against isolate genomes derived from the same fecal sample. Appl. Environ. Microbiol. 87, e02593. https://doi.org/10.1128/AEM.02593-20. -20.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., Bateman, A., 2021. Pfam: the protein families database in 2021. Nucleic. Acids. Res. 49, D412–D419. https://doi.org/10.1093/nar/gkaa913.

Monteiro, C., Saxena, I., Wang, X., Kader, A., Bokranz, W., Simm, R., Nobles, D., Chromek, M., Brauner, A., Brown, R.M., Romling, U., 2009. Characterization of cellulose production in *Escherichia coli* Nissle 1917 and its biological consequences. Environ. Microbiol. 11, 1105–1116.

Nierychlo, M., Andersen, K.S., Xu, Y., Green, N., Jiang, C., Albertsen, M., Dueholm, M.S., Nielsen, P.H., 2020. MiDAS 3: an ecosystem-specific reference database, taxonomy and knowledge platform for activated sludge and anaerobic digesters reveals species-level microbiome composition of activated sludge. Water Res. 182, 115955 https://doi.org/10.1016/j.watres.2020.115955.

Nouha, K., Kumar, R.S., Balasubramanian, S., Tyagi, R.D., 2018. Critical review of EPS production, synthesis and composition for sludge flocculation. J. Environ. Sci. 66, 225–245. https://doi.org/10.1016/j.jes.2017.05.020.

Petriglieri, F., Singleton, C.M., Kondrotaite, Z., Dueholm, M.K.D., McDaniel, E.A., McMahon, K.D., Nielsen, P.H., 2022. Reevaluation of the phylogenetic diversity and global distribution of the genus "*Candidatus* Accumulibacter. mSystems 7, e00016–22. https://doi.org/10.1128/msystems.00016-22.

Petriglieri, F., Singleton, C., Peces, M., Petersen, J.F., Nierychlo, M., Nielsen, P.H., 2021. *Candidatus* Dechloromonas phosphoritropha" and "*Ca*. D. phosphorivorans", novel polyphosphate accumulating organisms abundant in wastewater treatment systems. ISME J. 15, 3605–3614. https://doi.org/10.1038/s41396-021-01029-2.

Rehman, Z.U., Wang, Y., Moradali, M.F., Hay, I.D., Rehm, B.H.A., 2013. Insights into the assembly of the alginate biosynthesis machinery in *Pseudomonas aeruginosa*. Appl. Environ. Microbiol. 79, 3264–3272. https://doi.org/10.1128/AEM.00460-13.

Remminghorst, U., Rehm, B.H.A., 2006. Bacterial alginates: from biosynthesis to applications. Biotechnol. Lett. 28, 1701–1712. https://doi.org/10.1007/s10529-006-9156-x.

Römling, U., Galperin, M.Y., 2015. Bacterial cellulose biosynthesis: diversity of operons, subunits, products, and functions. Trends Microbiol. 23, 545–557. https://doi.org/10.1016/j.tim.2015.05.005.

Roux, D., Cywes-Bentley, C., Zhang, Y.-F., Pons, S., Konkol, M., Kearns, D.B., Little, D.J., Howell, P.L., Skurnik, D., Pier, G.B., 2015. Identification of poly-N-acetylglucosamine as a major polysaccharide component of the *Bacillus subtilis* biofilm matrix. J. Biol. Chem. 290, 19261–19272. https://doi.org/10.1074/jbc.M115.648709.

Ruiken, C.J., Breuer, G., Klaversma, E., Santiago, T., van Loosdrecht, M.C.M., 2013. Sieving wastewater – Cellulose recovery, economic and energy evaluation. Water Res. 47, 43–48. https://doi.org/10.1016/j.watres.2012.08.023.

Schmid, J., Sieber, V., Rehm, B., 2015. Bacterial exopolysaccharides: biosynthesis pathways and engineering strategies. Front. Microbiol. 6, 1–24. https://doi.org/10.3389/fmicb.2015.00496.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Seviour, R.J., Nielsen, P.H., 2010. Microbial Ecology of Activated Sludge. IWA Publishing.

Seviour, T., Derlon, N., Dueholm, M.S., Flemming, H.C., Girbal-Neuhauser, E., Horn, H., Kjelleberg, S., van Loosdrecht, M.C.M., Lotti, T., Malpei, M.F., Nerenberg, R., Neu, T. R., Paul, E., Yu, H., Lin, Y., 2019. Extracellular polymeric substances of biofilms: suffering from an identity crisis. Water Res. 151, 1–7. https://doi.org/10.1016/j.watres.2018.11.020.

Seviour, T., Lambert, L.K., Pijuan, M., Yuan, Z., 2010. Structural determination of a key exopolysaccharide in mixed culture aerobic sludge granules using NMR spectroscopy. Environ. Sci. Technol. 44, 8964–8970. https://doi.org/10.1021/es102658s.

Seviour, T., Yuan, Z., van Loosdrecht, M.C.M.M., Lin, Y., 2012. Aerobic sludge granulation: a tale of two polysaccharides? Water Res. 46, 4803–4813. https://doi.org/10.1016/j.watres.2012.06.018.

Siddharth, T., Sridhar, P., Vinila, V., Tyagi, R.D., 2021. Environmental applications of microbial extracellular polymeric substance (EPS): a review. J. Environ. Manage. 287, 112307 https://doi.org/10.1016/j.jenvman.2021.112307.

Singleton, C.M., Petriglieri, F., Kristensen, J.M., Kirkegaard, R.H., Michaelsen, T.Y., Andersen, M.H., Kondrotaite, Z., Karst, S.M., Dueholm, M.S., Nielsen, P.H., Albertsen, M., 2021. Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. Nat. Commun. 12, 2009. https://doi.org/10.1038/s41467-021-22203-2.

Smoot, J.C., Barbian, K.D., Gompel, J.J.V., Smoot, L.M., Chaussee, M.S., Sylva, G.L., Sturdevant, D.E., Ricklefs, S.M., Porcella, S.F., Parkins, L.D., Beres, S.B., Campbell, D.S., Smith, T.M., Zhang, Q., Kapur, V., Daly, J.A., Veasy, L.G., Musser, J. M., 2002. Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. Proc. Natl. Acad. Sci. 99, 4668–4673. https://doi.org/10.1073/pnas.062526099.

Sze, J.H., Brownlie, J.C., Love, C.A., 2016. Biotechnological production of hyaluronic acid: a mini review. 3 Biotech 6, 67. https://doi.org/10.1007/s13205-016-0379-9.

Vorhölter, F.-J., Schneiker, S., Goesmann, A., Krause, L., Bekel, T., Kaiser, O., Linke, B., Patschkowski, T., Rückert, C., Schmid, J., Sidhu, V.K., Sieber, V., Tauch, A., Watt, S. A., Weisshaar, B., Becker, A., Niehaus, K., Pühler, A., 2008. The genome of *Xanthomonas campestris* pv. *campestris* B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis. J. Biotechnol. 134, 33–45. https://doi.org/10.1016/j.jbiotec.2007.12.013.

Wang, X., Preston, J.F., Romeo, T., 2004. The *pgaABCD* locus of *Escherichia coli* promotes the synthesis of a polysaccharide adhesin required for biofilm formation. J. Bacteriol. 186, 2724–2734. https://doi.org/10.1128/JB.186.9.2724-2734.2004.

Wang, Z., Kadouri, D.E., Wu, M., 2011. Genomic insights into an obligate epibiotic bacterial predator: *micavibrio aeruginosavorus* ARL-13. BMC Genomics 12, 453. https://doi.org/10.1186/1471-2164-12-453.

Whitfield, G.B., Howell, P.L., 2021. The matrix revisited: opening night for the Pel polysaccharide across eubacterial kingdoms. Microbiol. Insights 14, 11786361209885888. https://doi.org/10.1177/11786361209885888.

Whitney, J.C., Howell, P.L., 2013. Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria. Trends Microbiol. 21, 63–72. https://doi.org/10.1016/j.tim.2012.10.001.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. J. Open Source Softw. 4, 1686. https://doi.org/10.21105/joss.01686.

Wilén, B.-M., Keiding, K., Nielsen, P.H., 2000. Anaerobic deflocculation and aerobic reflocculation of activated sludge. Water Res. 34, 3933–3942. https://doi.org/10.1016/S0043-1354(00)00274-8.

Wilkins, D., 2020. gggenes: draw gene arrow maps in "ggplot2.".

Xu, L., Cheng, R., Li, J., Wang, Y., Zhu, B., Ma, S., Zhang, W., Dong, W., Wang, S., Zhang, J., 2017. Identification of substituent groups and related genes involved in salecan biosynthesis in *Agrobacterium* sp. ZX09. Appl. Microbiol. Biotechnol. 101, 585–598. https://doi.org/10.1007/s00253-016-7814-z.

Xu, S., Dai, Z., Guo, P., Fu, X., Liu, S., Zhou, L., Tang, W., Feng, T., Chen, M., Zhan, L., Wu, T., Hu, E., Jiang, Y., Bo, X., Yu, G., 2021. ggtreeExtra: compact visualization of richly annotated phylogenetic data. Mol. Biol. Evol. 38, 4039–4042. https://doi.org/10.1093/molbev/msab166.

Yamazaki, M., Thorne, L., Mikolajczak, M., Armentrout, R.W., Pollock, T.J., 1996. Linkage of genes essential for synthesis of a polysaccharide capsule in *Sphingomonas* strain S88. J. Bacteriol. 178, 2676–2687. https://doi.org/10.1128/jb.178.9.2676-2687.1996.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.-Y., 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8, 28–36. https://doi.org/10.1111/2041-210X.12628.

Yu, X., Atalla, R.H., 1996. Production of cellulose II by *Acetobacter xylinum* in the presence of 2,6-dichlorobenzonitrile. Int. J. Biol. Macromol. 19, 145–146. https://doi.org/10.1016/0141-8130(96)01109-9.

Zhang, F., Qian, D.-K., Geng, Z.-Q., Dai, K., Zhang, W., van Loosdrecht, M.C.M., Zeng, R. J., 2021. Enhanced methane recovery from waste-activated sludge by alginate-degrading consortia: the overlooked role of alginate in extracellular polymeric substances. Environ. Sci. Technol. Lett. 8, 86–91. https://doi.org/10.1021/acs.estlett.0c00784.