Metagenomic analysis of sludge and early-stage biofilm communities of a submerged membrane bioreactor

Zahid Ur Rehman, Luca Fortunato, Tuoyuan Cheng, TorOve Leiknes

PII: S0048-9697(19)34673-X
DOI: <https://doi.org/10.1016/j.scitotenv.2019.134682>
Reference: STOTEN 134682

To appear in: *Science of the Total Environment*

Received Date: 15 September 2019
Revised Date: 24 September 2019
Accepted Date: 25 September 2019

Please cite this article as: Z.U. Rehman, L. Fortunato, T. Cheng, T. Leiknes, Metagenomic analysis of sludge and early-stage biofilm communities of a submerged membrane bioreactor, *Science of the Total Environment* (2019), doi: <https://doi.org/10.1016/j.scitotenv.2019.134682>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

For Submission to Science of The Total Environment

**Metagenomic analysis of sludge and early-stage biofilm communities
of a submerged membrane bioreactor**

Zahid Ur Rehman^{a*}, Luca Fortunato^a, Tuoyuan Cheng^a, TorOve Leiknes^a

Category: Original research paper

^aWater Desalination and Reuse Center (WDRC), King Abdullah University of Science and Technology (KAUST),
Thuwal 23955-6900, Kingdom of Saudi Arabia

*Corresponding authors: Tel. +966 12 808 0192; Email: Zahid.urrehman@kaust.edu.sa/ Tel. +966 12 808 2193;

Abstract

Biofilm formation on membranes in activated sludge membrane bioreactors (MBR), commonly identified as biofouling, is a significant problem for MBR operations. A better understanding of microbial species involved in the biofilm formation is needed to develop anti-biofilm measures. A read-based and genome-resolved shotgun metagenomic approach was applied to characterize the composition and functional potential of the sludge and early stage biofilm microbial communities in an MBR process. Read-based analysis revealed that the prevalence of different phyla are relatively similar in both the sludge and biofilm samples, with *Proteobacteria* as the most dominant, followed by *Chloroflexi*, *Bacteroidetes* and *Planctomycetes*. However, the relative abundance of these phyla slightly varies between the sludge and biofilm. Phyla such as *Actinobacteria*, bacterial candidate phyla, *Chlamydiae*, *Cyanobacteria*/Melainabacteria and *Firmicutes* are 2 to 4 times more abundant in the biofilm than in the sludge. At the genus level, genera belonging to *Proteobacteria* (*Legionella*, *Caulobacter*, *Sphingomonas*, *Acinetobacter* and *Rhizobium*), *Cyanobacteria* (*Hassallia*), and *Spirochaetes* (*Turneriella*) are at least twice more abundant in the biofilm. These genera, especially those belonging to Phylum *Proteobacteria*, are known to play an important role in the formation of biofilms on surfaces. The Alpha diversity is found slightly higher in the biofilm, compared with sludge samples. Functional classification of reads through the SEED subsystem shows that functional classes such as those involved in the metabolism of various molecules are significantly different in the biofilm and sludge.

A phylogenomic analysis of the six extracted metagenome assembled genomes (MAGs) shows that three MAGs belong to *Proteobacteria*, and one MAG belong to each of *Chloroflexi*, *Bacteroidetes* and *Planctomycetes*. The relative abundance of the MAG belonging to *Alphaproteobacteria* is higher in the biofilm. A functional potential analysis of the MAGs reveals

24 their potential to metabolize carbon and nitrogen sources, as well as the prevalence of antibiotic
25 resistance genes.

26

27 Key words: biofilms, membrane bioreactor, metagenomics, activated sludge, wastewater

28

1. Introduction

Water scarcity, caused by climate change and the growth in population is a significant problem faced by mankind. Reusing treated wastewater is becoming increasingly important to relieve the pressure on already dwindling water sources. Membrane bioreactors (MBR) that couple the conventional activated sludge with membrane filtration, are widely used to treat municipal and industrial wastewater for reuse in a variety of applications. However, membrane biofouling remains a significant problem in MBR processes, as it lowers the productivity and increases operational costs (Wang et al., 2009).

Membrane biofouling in an MBR for treating wastewater is governed by complex physicochemical and biological processes. These include MBR operating conditions, composition of wastewater and membrane properties. Generally, biofouling involves conditioning of the surface of the membrane with organic substances (Lorite et al., 2011), and the attachment of pioneering bacterial species that form micro-colonies. These events facilitate the attachment of other microbes, *i.e.* secondary colonizers, and subsequently the formation of mature biofilms consisting of multiple microbial species (Nagaraj et al., 2017). Identifying key microbial species involved in the formation and sustenance of a biofilm allows us to design rational strategies to mitigate biofouling (Campbell et al., 1999).

Recently, studies have been conducted to explore the microbial diversity of sludge and membrane biofilms, however, contradicting results were reported. Some studies suggest different microbial communities develop in the sludge compared to the biofilm on the membrane surface (Jinhua et al., 2006; Miura et al., 2007b), while others report similar community profiles, especially in the case of mature biofilms (Ziegler et al., 2016). The differences reported might be due to different operational parameters of the systems studied, and maybe also the age of the biofilm. There is

therefore a need to investigate the microbial communities of the sludge and the biofilm at various stages of growth to get a better understanding of the development and maturation of these microbial community in the biofilm.

Moreover, the above-mentioned studies used different techniques to investigate the composition of the microbial communities, (e.g. culture-based or 16S-rRNA gene-based analysis). Although these studies provide valuable information, they also have important limitations due to the techniques applied, e.g. cultivation and primer bias (**Dang and Lovell, 2000**). There is therefore a need for other techniques that can overcome these biases in order to truly get a better understanding of the composition of microbial communities of the activated sludge and the biofilm. Shotgun metagenomic sequencing provides an alternative to traditional techniques used to characterize microbial communities. This technique has an added advantage of yielding information about the functional potential of a microbial community. Shotgun metagenomics has been used to study the microbial communities of activated sludge processes treating different types of influent wastewater and at different geographical locations (Ibarbalz et al., 2016; Johnson et al., 2015). However, studies using such a technology to compare the composition and functional potential of microbial communities of activated sludge and biofilm are scarce.

In this study, a shotgun metagenomic approach was used to compare the composition and functional potential of the sludge and biofilm microbial communities of a lab-scale MBR system with a submerged ultrafiltration (UF) membrane configuration. The MBR was operated for eight days, fed with primary treated wastewater obtained from a full-scale wastewater treatment plant. This allows any changes to the microbial community profile of the sludge and membrane biofilm formed during early-stage operation to be investigated. The shotgun sequencing approach allows extraction of the metagenome-assembled genomes (MAGs) from the most abundant bacterial

species in the sludge and biofilm samples. Extracted MAGs were annotated and their functional potentials investigated.

2. Materials and methods

2.1 MBR setup and operation

Experiments were conducted using a lab-scale activated sludge membrane bioreactor (AS-MBR) for the treatment of municipal wastewater collected from King Abdullah University of Science and Technology's wastewater treatment plant (WWTP). The details of reactor setup and operations are provided elsewhere (Luca Fortunato, 2018). Briefly, the lab-scale AS-MBR setup consisted of a 20 L feeding tank and a membrane bioreactor of 10.5 L. Primary treated wastewater from an equalization tank at the WWTP was collected every 2 days, transported and stored at 4 °C and used as influent. Composition of the wastewater included; chemical oxygen demand (COD) of 200 ±30 mg/L, total suspended solids (TSS) of 70 ±30 mg/L, total dissolved solids (TDS) of 350 ±33 mg/L, total nitrogen (TN) of 14.5 ±3.5 mg/L, a PO₄ concentration of 5.0 ±0.2 mg/L and a pH of 7.4 ±0.2. The feed was pumped into the AS-MBR reactor and operated for 8 days under a constant flux of 10 L/m²h. The membrane was an UF flat sheet membrane (MemSis, Turkey) with an area of 0.04 m² and 20 KDa molecular weight cut-off (MWCO) submerged in the biological reactor. Activated sludge was collected from the WWTP to seed the reactor and acclimatized in the reactor for 21 days before starting the experiment. The mixed liquor suspended solids (MLSS) was set to 4 g/L. Sludge retention time (SRT) was fixed at 70 days during the experiment by daily draining 150 mL of mixed liquor from the bioreactor. The hydraulic retention time (HRT) of the AS-MBR system was set at 26 hours.

2.2 Sampling

Sludge and biofilm samples (on 1 cm² membrane pieces) were collected and immediately processed for DNA extraction. Membrane samples were vortexed in 1X TE buffer to separate the biomass from the membrane. For DNA extraction, a QIAamp® DNA mini kit (Qiagen, Germany) was used according to the DNA purification protocol for tissues. The extracted DNA quality was measured using NanoDrop (ThermoFisher scientific, USA). The 260/280 ratio of samples was between 1.8 and 2.0, while 260/230 was ≥ 2.0 which are required in order to proceed with library preparation. The DNA concentration was quantified using the Qubit BR DNA assay kit (Invitrogen, USA).

2.3 Sequencing of DNA

Libraries were prepared using a TruSeq® Nano DNA LT kit for sequencing, following the manufacturer's instructions (Illumina, USA). The prepared libraries were quantified and normalized to 10 nM, pooled and sequenced on a HiSeq 4000 (Illumina, USA) paired-end platform.

2.4 Read- based analysis

Raw reads in FASTQ format were checked for their quality, where low-quality reads and adapter sequences were removed using Cutadapt (Martin, 2011). Reads with a quality score lower than 20, and shorter than 130 bp were removed. Quality-filtered paired-end reads were merged using Pandaseq (Masella et al., 2012). DIAMOND was used to align merged reads against NCBI's non-redundant protein database (July 12, 2018) using the BlastX tool (Buchfink et al., 2015). A default e-value cutoff of 0.001 and a bit score greater than 50 was used. The file produced by DIAMOND was "meganized" and subsequently imported into MEGAN 6 to perform taxonomic and functional binning of the reads (Huson et al., 2007). For comparison, sample reads were normalized to those with lowest read count, and all the unassigned reads were discarded.

MEGAN uses the Lowest Common Ancestor (LCA) algorithm for taxonomic binning of reads. The LCA uses NCBI taxonomy and places reads onto the node (taxa), which is at higher taxonomic level than for the species for which the read has significant alignment (Huson et al., 2016). Functional analysis of reads was performed using the SEED subsystem classification (Overbeek et al., 2014). MEGAN provides a hierarchical representation for both the taxonomy and SEED functional classes. Relative abundance of reads belonging to functional categories of SEED subsystem, for both sludge and biofilm samples, were obtained from MEGAN. The significant difference in the relative abundance of reads was calculated using Tukey's Honest Significant Difference test; p-value less than 0.05 was considered significant.

MEGAN was used to calculate Alpha diversity measures, Shannon-weaver index and Simpson's reciprocal index. To perform these analyses, all the taxonomic nodes (reads belonging to bacteria and archaea only) up to the genus level were included for calculations. In total 390 taxonomic nodes were included to calculate diversity indices.

2.5 Genome-resolved analysis

Quality filtered reads from all the samples were combined and assembled, using meta-Spades that was developed specifically for metagenomic reads (Bankevich et al., 2012). Assembled contigs smaller than 1000 bp were removed from the assembly. To assess coverage, reads were mapped back to the contigs, using bowtie2 that produced SAM files (Langmead and Salzberg, 2012). The Sequence Alignment/Map (SAM) files were converted to a Binary Alignment/Map (BAM) format, using SAM tools (Li et al., 2009). Contigs were binned using COCOCT v 0.4.1 available on the anvi'o platform v3 (Alneberg et al., 2014). CONCOCT uses nucleotides composition and relative abundance of contigs across samples for binning into individual genomes. Individual bins (MAGs) were manually refined using anvi'o's interactive graphical interface (Eren et al., 2015). This

interactive interface allows users to add or remove contigs from a bin (MAG), in real time. Contigs were removed from a MAG if the coverage, GC content and taxonomy were not consistent with other contigs in the bin. Completeness and contamination of MAGs were calculated using the presence/absence of a set of 139 bacterial single-copy core genes (BSGs) (Campbell et al., 2013). Phylogenomic analysis of the extracted MAGs was performed as follows; the MAGs were first annotated using Prokka (Seemann, 2014), and the resulting amino- acid sequence file was blasted against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database using GhostKOALA (Kanehisa et al., 2016b). Taxonomy information produced by KEGG was used to identify and download reference genomes from the GenBank (Benson et al., 2013). The FASTA files of the reference genomes, and our MAGs, were processed via anvi'o phylogenomic workflow, and a contig database for each genome was created. The HMM profile of the Campbell *et al* (Campbell et al., 2013) collection was used to identify single-copy core genes in the extracted MAGs, and to create a FASTA file containing concatenated amino acid sequences of ribosomal proteins. The concatenated amino acid sequences were imported into MEGA7 for ClustalW- based alignment (Kumar et al., 2016). The phylogenetic tree was computed using the neighbor-joining statistical method.

2.6 Module Completion Ratio (MCR)

The completion ratios of KEGG modules in the extracted MAGs were calculated using MAPLE-2.3.1, as described previously (Arai et al., 2018). MCR ratios between the different MAGs were compared using the MCR comparison option.

2.7 Data Availability

The DNA sequence data generated and analyzed during the current study are available in NCBI Sequence Read Archive (SRA) (BioProject Accession = PRJNA516038) (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA516038>). The sequences of MAGs were submitted to the DDBJ/ENA/GenBank. The Accession numbers of MAGs are given in Fig. 5.

2.8 Statistics

Tukey test for HSD was performed using RStudio v. 1.1.463. Multivariate analysis such as PCA-biplot was performed in RStudio using packages FactoMineR(Le et al., 2008) and factoextra (R package version 1.0.5).

3. Results and discussion

3.1 Composition of the microbial community (Phylum and Class level)

Shotgun metagenomic sequencing generated 59Gb of data for all the four samples. After quality filtering of reads we were left with 44Gb of data. Total numbers of raw, quality-filtered and merged reads obtained for each sample are given in (Supp. Table 1). Approximately 96% of the reads, from all samples, were assigned to cellular organisms belonging to bacterial taxa (Supp. Table 1). Recently, we published the bacterial phyla and Proteobacterial classes detected in the same sludge and biofilm samples (Luca Fortunato, 2018). In that study, Kaiju was used to taxonomically classify the reads, which also uses the NCBI taxonomy. However, in the present study, DIAMOND was used for the protein sequence alignment. Even though different methodologies were used for the taxonomic characterization of microbial communities in the sludge and the biofilm, the distribution of bacterial phyla were found to be similar in both studies (Supp. Fig 1) (Luca Fortunato, 2018).

Proteobacteria were the most dominant phyla in both the sludge and the biofilm (Supp. Fig 1) (Luca Fortunato, 2018). These results are in agreement with previous studies that show *Proteobacteria* are the dominant bacterial phyla in the sludge and biofilms of wastewater treatment systems (Bedoya et al., 2019; Jo et al., 2016; Miura et al., 2007a). However, *Proteobacteria* were found to be more abundant in the biofilm (~46%) than in the sludge (~38%). The higher abundance of *Proteobacteria* in the biofilm substantiates previous reports that suggested that members of this Phyla are involved in initiating biofilms (Elifantz et al., 2013; Nagaraj et al., 2017). *Chloroflexi* were found to be the second most abundant Phylum in the sludge and biofilm, followed by *Bacteroidetes* and *Planctomycetes* (Supp. Fig 1). *Chloroflexi* and *Bacteroidetes* are involved in the degradation of polymers and complex organic substances, resulting in simpler organic molecules that are more readily used by other microbes (Zhang et al., 2017). *Chloroflexi* were observed to be more abundant in the sludge (20.5%) than in the biofilm (13.1%). *Chloroflexi* are filamentous bacteria, and various species of this phylum have been detected in several wastewater treatment plants (Miura et al., 2007a). It was suggested that filaments of these bacteria might adhere to and penetrate between membranes and foulants (Jo et al., 2016). The higher abundance of *Chloroflexi* in the sludge and the biofilm detected in this study supports this hypothesis. The other bacterial phyla, such as *Actinobacteria*, *Acidobacteria*, *Chlorobi*, *Firmicutes* and *Nitrospirae*, which are commonly detected in wastewater treatment plants were also found (Supp. Fig. 1). Overall, the distribution and relative abundance of bacterial phyla found in this study were similar to previous reports in the literature, albeit with some differences (Miura et al., 2007a; Wang et al., 2012).

The greatest difference in the relative abundance of bacterial phyla between the sludge and the biofilm was observed for *Chlamydiae*, *Cyanobacteria*, bacterial candidate phyla, *Actinobacteria* and *Firmicutes*. All of these phyla were at least two times higher in the biofilm than in the sludge

(Supp. Fig. 1). Besides these differences, the distribution of bacterial phyla in the sludge and the biofilm were found to be similar, which was expected, as microbes in the sludge act as inoculum for the formation of the biofilm on the membrane. Furthermore, it has been speculated that microbial communities of sludge flocs and membrane biofilms are similar, as they originate from, and are seeded by, microbial aggregates in the wastewater (Laspidou and Rittmann, 2002; Sheng et al., 2010).

Since *Proteobacteria* were the most dominant bacterial Phylum found in the sludge and the biofilm, the relative abundance of Proteobacterial classes was therefore investigated (Supp. Table 2). In the sludge samples, *Betaproteobacteria* were the most dominant (~50%), however, in the biofilm samples, *Gammaproteobacteria* and *Alphaproteobacteria* were found to be the dominant Proteobacterial classes, followed by *Betaproteobacteria* (Supp. Table 2). These results support findings from previous studies suggesting that these bacteria are involved in the early colonization of surfaces (Elifantz et al., 2013; Nagaraj et al., 2017). Another study investigating a 55 day-old biofilm in MBR, showed that *Betaproteobacteria* were dominant in biofilm samples, followed by *Gammaproteobacteria* (Miura et al., 2007b). These findings suggest that, as biofilms mature, *Betaproteobacteria* gradually replace other bacteria, and are thus characteristic of stable and mature biofilms (Nagaraj et al., 2017).

3.2 Composition of the microbial community (genus level)

The relative abundance of different genera in the sludge and biofilm was also investigated. A heatmap analysis shows that duplicate sludge and biofilm samples are clustered more closely together (Fig. 1). Similar to PCoA analysis, the dendrogram presented above the heatmap, shows that the composition of sludge samples was more uniform, as sludge samples clustered more

233 closely than the biofilm samples (Fig. 1). Most of the genera were present at a relative abundance
234 of approximately 1% or less. Some genera such as *Thauera*, *Nitrospira* and *Legionella* and
235 *Hassallia* were present at more than 10% of relative abundance (*Hassallia* only for one replicate).
236 The greatest increase was observed in the biofilm for *Hassallia* and *Legionella*, which were 11.6
237 and 13.7 times higher respectively (Fig. 1). *Acinetobacter*, *Caulobacter*, *Fluviicola*, *Hassallia*,
238 *Legionella*, *Rhizobium*, *Sphingomonas* and *Turneriella* were at least twice more abundant in the
239 biofilm than in the sludge. However, *Thauera*, *Nitrospira* and *Legionella* and *Hassallia* were the
240 key bacterial genera responsible for dissimilarity between sludge and biofilm samples as shown
241 by PCA-biplot analysis (Supp. Fig. 2)

242 *Caulobacter*, along with glycosphingolipid-producing bacteria such as *Rhizobium* and
243 *Sphingomonas* are known as primary colonizers of membrane surfaces (Nagaraj et al., 2017). The
244 sticky extracellular polysaccharides (EPS) produced by these bacteria help in membrane
245 attachment and protect these bacteria from adverse environmental conditions (Nagaraj et al.,
246 2017). Genus *Hassallia* belong to *Cyanobacteria*, and their significance in membrane biofilm is
247 unknown. *Legionella* is an important waterborne pathogen that can colonize and persist in biofilms
248 in a variety of man-made water systems (Abu Khweek and Amer, 2018), however, information
249 regarding the role of *Legionella* in biofilms on membranes used in MBR is scarce. Our results
250 suggest that *Legionella* can colonize and populate membrane surfaces and may have important
251 implications for the spread of legionellosis.

252 Bacteria genera such as *Nitrospira*, *Thauera* and *Zoogloea*, were respectively 5.8, 8 and 12.8-
253 times higher in the sludge than in the biofilm (Fig. 1). These are typical genera detected in activated
254 sludge processes (Wang et al., 2012), where these bacteria are involved in the degradation of
255 organic and inorganic compounds commonly detected in wastewater. For example, *Nitrospira* are

involved in the nitrification process, and *Thauera* are responsible for the degradation of aromatic compounds in wastewater (Daims et al., 2015; Zhang et al., 2017). Members of genus *Zoogloea* are considered to be responsible for the formation of flocs in activated sludge, and therefore improve the water purification process (Rossellomora et al., 1995).

3.3 Biofilm community is more diverse than in suspended sludge

Alpha diversity measures such as the Shannon-Weaver's and Simpsons's reciprocal index, were calculated based on taxonomy. Results show that the biofilm microbial diversity was slightly higher than the sludge community. The average Shannon-Weaver index for the biofilm was 5.1 (± 0.03) and 4.8 (± 0.01) for the sludge. Similarly, the average Simpson's reciprocal index calculated for the biofilm was 10.4 (± 0.12) and 9.9 (± 0.03) for the sludge. Although these differences are small, they show that the alpha diversity of the biofilm is greater than the diversity in the sludge. These results are consistent with a recent study performed on five full-scale MBRs that showed that the microbial diversity of an early stage biofilm (5 hrs) was slightly higher than that of an activated sludge (Matar et al., 2017). Based on these results, it appears that biofilm communities formed during the early stages of operation undergoes changes leading to a higher microbial diversity.

3.4 Biofilm and sludge communities cluster separately in ordination plots

A principal coordinate analysis was performed to investigate the similarity/dissimilarity of the biofilm and sludge microbial communities. These results show that the duplicate sludge samples had similar values along the 1st axis of PCoA and the same was found for the duplicate biofilm

samples (Fig. 2). However, there were small differences between the duplicates along the 2nd axis of PCoA. Such differences were greater in the biofilm than in the sludge samples (Fig. 2), suggesting that the composition of the sludge microbial community is more homogenous than the biofilm one. These results further support the dendrogram in Fig. 1 that shows that duplicate sludge samples are more similar compared to duplicate biofilm samples (Fig.1 and 2). A possible reason for this may be the aeration in the MBR tank, which leads to a continuous mixing of the sludge inside the reactor and hence a more uniform composition of the microbial community. However, the bacteria that adhere to the membrane surface in the MBR experience slightly different environmental conditions such as suction pressure and nutrient availability. This might lead to differences in the composition of the community at various localized areas of the membrane surface. Similar results were reported in a recent study comparing the microbial diversities between the biofilm and activated sludge in ten different MBRs (Jo et al., 2016).

Results found in this study suggest that an eight-days old biofilm microbial community is not merely the concentration of microbes on the membrane caused by the convective forces generated during filtration. Differences in the biofilm and sludge microbial community observed are most probably the result of different micro-environments experienced by the microbes. In this scenario, only microbes that can thrive in the biofilm are retained and enriched on the membranes. Findings in this study are consistent with previous reports that showed that the biofilm microbial diversity was different from that in the sludge community (Jo et al., 2016; Miura et al., 2007a).

3.5 Genes involved in the production/metabolism of carbohydrates, proteins, amino-acids, and cofactors, vitamins are the most abundant

Reads from sludge and biofilm samples were functionally classified, using a SEED subsystem approach, where genes are classified based on their functional roles. The classified genes that are part of a related metabolic pathway are grouped into “subsystems”. The level 1 SEED subsystems detected in this study were found similar to those described in previous studies conducted on WWTP (Tian et al., 2015; Ye et al., 2012). We performed Tukey’s honest significance test to confirm the statistically significant difference in the abundance of reads belonging to SEED subsystems. These analyses showed that reads annotated to SEED subsystems for carbohydrates, and amino acid and derivatives were significantly higher in the sludge. However, SEED subsystems such as cofactors, vitamins, prosthetic groups, pigments and RNA metabolism were found significantly abundant in biofilm samples compared to sludge. Differences in other SEED subsystems were observed, such as those related to metabolism of DNA and fatty acid and lipids (Fig. 3). However, the distribution of the various functional classes in the sludge and the biofilm was very similar. Overall, results show that genes related to the metabolism of carbohydrates, proteins, amino acids and fatty acids were highly abundant in both the sludge and the biofilm (Fig. 3). Wastewater contains higher concentrations of biological products, such as carbohydrates, amino acids, and proteins, so the genes involved in the metabolism of these molecules are abundant in microbes (Ibarbalz et al., 2016). Moreover, these molecules play an important role in the physiology of microbes. For example, carbohydrates play an important role in the formation of sludge flocs and biofilms (McSwain et al., 2005). It has been suggested that a high number of genes involved in the production of macromolecules detected in wastewater water bacteria may be the result of a high selective pressure on these microbes to form flocs and biofilms that are required for retaining biomass in the system (Albertsen et al., 2012). Carbohydrates also serve as an important source of carbon and energy for microorganisms, and many environmental bacteria carry

enzymes capable of degrading a large number of diverse carbohydrates (El Kaoutari et al., 2013; Gunsalus et al., 1955). Similarly, proteins also play an important role in the flocculation of sludge granules and found to be the dominant fraction of EPS in the sludge (Guo et al., 2016). Although statistically significant differences were observed in the relative abundance of functional categories in the sludge and biofilm samples, visually there were no major differences (Fig. 3). It is possible that these small differences will become more pronounced as the biofilm matures. It is noted that the above analysis only reveals the presence of genes, however, a metatranscriptomic analysis of the sludge and the biofilm would reveal the genes that are functionally important and differentially expressed. Identifying the functions that play key roles in sustaining the biofilm microbial community on the membrane may provide suitable targets for rational anti-biofilm and biofouling mitigation strategies in MBR systems.

3.6 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways related to carbon and nitrogen metabolism are important for sludge and biofilm communities

It has been suggested that the accuracy of the annotation is reduced when short reads are used and that read-based annotations provide only a general view of data (Ibarbalz et al., 2016). To address this concern, reads were assembled into contigs that were binned to generate individual genomes called “MAGs”. In total 40 MAGs were obtained by co-assembly of sludge and biofilm samples (Supp. Table 3). For further analysis, we selected MAGs with less than 5% redundancy (contamination) and more than 95% completion. Only 8 MAGs were found to meet this inclusion criterion. Their various features and characteristics are presented in Table 1.

The extracted MAGs were annotated using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database as reference (Kanehisa et al., 2016a). The functional potential of the extracted

346 MAGs was evaluated, using a metabolic and physiological potential evaluator (MAPLE), and
 347 represented as a module completion ratio (MCR) of KEGG modules. KEGG modules are manually
 348 defined functional units based on KEGG pathways. Modules with q values (false-discovery rate)
 349 less than 0.5, at the individual taxonomic rank level, were considered functional. In total, we
 350 observed 100, 28, 46, 11, 66 and 44 functional modules ($q < 0.5$) for MBR_BPTB, MBR_CLF,
 351 MBR_BCT, MBR_PLN, MBR_GPTB, MBR_APTB, respectively (Fig. 4). The
 352 *Betaproteobacterial* MAG (MBR_BPTB) showed higher prevalence of pathways modules
 353 involved in carbohydrate metabolism. Entner-Doudoroff pathway module (M00008) and pentose
 354 phosphate pathway, non-oxidative phase, module (M00007) were exclusively detected in
 355 MBR_BPTB. Overall, all the MAGs showed higher prevalence of pathway modules involved in
 356 metabolism of carbohydrates and fatty acids (Fig. 4, Supp. Table 4). These functions are important
 357 in a sludge environment in which cells are continuously lysing, releasing organic molecules that
 358 can be used by the surrounding microbes as an energy and carbon source. The higher abundance
 359 of genes related to degradation of organic molecules was also suggested by the read-based analysis
 360 through the SEED subsystem (Fig. 3). Interestingly, most of the extracted MAGs did not contain
 361 KEGG modules involved in nitrogen metabolism (Fig. 4). However, one *Betaproteobacterial*
 362 MAG (MBR_BPTB) contained KEGG modules (M00529 and M00530) related to nitrate
 363 reduction (Fig. 4). The *Planctomycetal* MAG (MBR_PLN) contained module (M00804), with
 364 33.3% completion ratio, which is involved in the complete nitrification of ammonia to nitrate
 365 defined as comammox (Supp. Table 4). However, the working probability of this module was low
 366 ($q = 1$). *Planctomycetes* have been detected in diverse environments, and studies are beginning to
 367 highlight their pivotal role in nitrogen metabolism (Delmont et al., 2018; Rehman et al., 2019). In
 368 this study, the nitrifying genes were found to be relatively less abundant, which is consistent with

a previous study of nitrifying genes in the activated sludge microbial community (Yu and Zhang, 2012). The same study, however, showed that nitrifying bacteria have a high expression activity of nitrifying genes, which efficiently reduced the concentration of ammonia.

The extracted MAGs also carry genes that confer resistance to commonly used antibiotics, such as aminoglycosides, beta-lactams and fluoroquinolones (Fig. 4). It is not surprising, as municipal wastewater contains antibiotics that can exert selective pressure on sludge microbes to develop resistance. Furthermore, studies have shown that WWTPs are a major source of antibiotic resistance in environment (Pazda et al., 2019).

Interestingly, MBR_APTB, abundant in biofilm communities, contains a module for the transport of adhesin proteins required for the formation of biofilms (Hinsa et al., 2003). The production of adhesin proteins might be the reason for a higher abundance of MBR_APTB in the biofilm. A complete list of all the modules, their module IDs and the MCR is given in the Supplementary Table 4.

3.7 Extracted MAGs belong to diverse phyla of bacteria

A phylogenomic analysis of the MAGs showed that three MAGs belonged to *Proteobacteria* (one each to Alpha, Beta and *Gammaproteobacteria*) (Fig. 5). One MAG belonged to *Planctomycetes*, one to *Chloroflexi* and one to *Bacteroidetes*. Two MAGs belonged to bacterial candidate phyla. These MAGs were included, even though their completion ratio was less than 75%, because of their importance in wastewater treatment. The higher prevalence of proteobacterial MAGs detected in this study is in agreement with the read-based analysis that showed that *Proteobacteria* were dominant bacteria in both the sludge and biofilm communities (Supp. Figure 1). The *Betaproteobacterial* MAG (MBR_BPTB) was not detected in the biofilm, but it was found abundant in the sludge (Table 1). *Betaproteobacteria*, such as the *Azoarcus*-*Thauera* cluster, are

known to degrade aromatic compounds and therefore play an important role in the microbial ecology of the sludge (Rabus et al., 2002). This implies that the metabolism of aromatic compounds occurs primarily in the sludge. The relative abundance of *Gammaproteobacterial* MAG (MBR_GPTB) was found to be slightly higher in the sludge than in the biofilm. However, *Alphaproteobacterial* MAG (MBR_APTB) was more abundant in the biofilm than in the sludge (Table 1). Similar observations were made in previous studies, in which *Alphaproteobacteria* were found to be more abundant in the biofilm, and slightly more than *Gammaproteobacteria* (Elifantz et al., 2013). *Alphaproteobacteria* and *Gammaproteobacteria* are generally considered to be the primary colonizers of biofilms formed on membranes, and *Betaproteobacteria* the secondary colonizer (Nagaraj et al., 2017). Results from this study support this assumption as they show that *Alphaproteobacteria* were enriched in the eight-days old biofilm. Members of this group of bacteria can strongly attach to the surfaces through their ability to produce strong adhesion molecules (Tsang et al., 2006). The bacteria candidate phyla MAGs (MBR_CSP_1, 2) were found to be slightly higher in the biofilm than in the sludge (Table 1), and they are phylogenetically closely related to *Saccharimonas aalborgensis* (Fig. 5). In a previous study, the genome of these bacteria was extracted thorough metagenomic sequencing of activated sludge (Albertsen et al., 2013). It was suggested that these bacteria carry genes involved in the production of type IV pili that play a role in adhesion to surfaces. The slightly higher relative abundance of these bacteria in the biofilm (Table 1) and their ability to grow in microaerophilic conditions (Albertsen et al., 2013) suggests that these bacteria are important members of biofilms that form in activated sludge reactors.

4. Conclusions

In this study, a shotgun metagenomic sequencing was applied to study the composition and functional potential of activated sludge and an eight-days old membrane biofilm harvested from a lab-scale membrane bioreactor (MBR) treating domestic wastewater. Our analyses showed that primary colonizers that produce sticky glycosphingolipids and bacteria that possess adhesin proteins and pili were abundant in biofilm. This is not surprising as such appendages would allow these bacteria to attach and enrich on the surfaces.

Assuming that the abundance of reads/genes reflects the functional activity, it appears that the sludge and biofilm microbial community has a preference to degrade specific kind of macromolecules- for example, carbohydrates and lipids in case of sludge and nucleic acids in case of the biofilm community. However, a metatranscriptomic or metaproteomic analysis of the sludge and biofilm would allow a better understanding of the functions performed by these communities.

Overall the results show that the microbial community of the biofilm resembles that of the sludge, both taxonomically and functionally. This is understandable as the biofilm community originates from the sludge community. However, localized conditions in the biofilm select for microbes that are more suitable for the maintenance and development of the biofilm. We believe that the functional characterization of membrane biofilms, at different levels of maturity, may provide valuable information about the biological processes that are critical to the sustenance of the biofilms. Such information can be applied develop rational approaches and better strategies for biofilm control and biofouling mitigation which is a major challenge in MBR operations. For example, quorum-quenching bacteria have been used as anti-biofilm agents in pilot-scale MBRs treating real municipal wastewater and other systems (Lee et al., 2016; Rehman and Leiknes, 2018). A better knowledge of quorum sensing enzymes prevalent in MBRs would allow us to

choose quorum quenching bacteria and enzymes that can degrade specific type of AHLs important in biofilm formation in MBRs.

Acknowledgements

This work was supported by the King Abdullah University of Science and Technology (KAUST) under award number BAS/1/1061-01-01.

Competing Interests

The authors declare that there are no competing interests

Author Contribution

ZUR, and TL designed the study. LF performed the MBR operations. ZUR processed and analyzed the sequence data. TC performed statistical analysis. ZUR wrote the paper. LF and TL proofread the article.

Fig. 1. Heat map showing the relative abundance of bacteria genera. Heatmap of relative abundance of genera in duplicate biofilm (B1, B2) and sludge (S1, S2) samples is given. A color key shows relative abundance (percentage) is given in upper left corner. Dendrogram representing the clustering of sludge and biofilm samples is shown on top of heatmap. A multicolor panel next to heatmap shows the Phyla to which the genera belong. Green arrows on

left shows the genera with relative abundance of more than 10%, while black arrows show the genera with two-fold higher abundance in the biofilm compared to the sludge.

Fig. 2. Principal coordination analysis of microbial communities. The clustering of duplicate biofilms (B1, B2) and sludge (S1, S2) are analyzed by PCoA. B1 and B2 have similar PC1 values, and duplicate S1 and S2 have similar PC1 values. However, S1 and S2 clustered more closely compared to B1 and B2 along the second axis, PC2.

Fig. 3. Relative abundance of SEED subsystems (level 1) in the sludge and the biofilm. Each bar represents the average (two samples) relative abundance of SEED subsystems detected in the sludge and the biofilm. The error bar represents the standard deviation. Asterisks show the significant difference between sludge and biofilm as determined by Tukey's HSD, p-value < 0.05 was considered significant.

Fig. 4. Functional potential analysis of MAGs. The comparison of module completion ratios (MCR) for selected KEGG pathway modules between eight extracted MAGs is shown. The functional classes, pathways, and their module IDs are given. Heatmap of MCR is shown (blue=0 MCR, yellow =50 MCR, and red = 100 MCR). All the modules with MCR 50 or higher have Q-value < 0.5.

Fig. 5. The phylogenomic tree was inferred by using maximum likelihood method. The consensus tree inferred from 500 bootstrap replicates is shown. Branches with bootstrap values less than 50% are collapsed. The values at the nodes show the percentage of replicates in which the associated taxa clustered together. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. All positions containing gaps and missing data were eliminated. There were total of 1162 positions in the final dataset. Evolutionary analysis was conducted in MEGA7. The MAGs extracted in this study are highlighted in red color. The accession number of all the genomes/MAGs are given in small brackets.

Table 1. Table showing the characteristics of MAGs that met inclusion criteria and were functionally annotated.

Supp. Fig. 1. Relative abundance of bacterial phyla in sludge and biofilm samples. The stacked bar chart shows the relative abundance of phyla in sludge (S1 and S2) and biofilm (B1 and B2) samples.

Supp. Fig. 2. The PCA-biplot of relative abundances at Genera level. The angles between samples represent their difference, with the scattered Genera points quantifying their contribution to the overall variation. The sharp angles between samples from the same groups, together with the almost-right angles between samples from different groups indicate that the two groups are different at community structure level via multivariate analysis. The key Genera that contributed to their differences are identified and highlighted as those far from the origin.

Supp. Table 1. Statistics of reads: number of raw, quality filtered, and merged reads. Number of reads assigned to cellular organisms, bacteria and SEED subsystems are also provided. Cellular organisms include bacteria, archaea and eukaryotes.

Supp. Table 2. Relative abundance of Proteobacterial classes in duplicate sludge (S1 and S2) and biofilm (B1 and B2) samples.

Supp. Table 3. Characteristics of all the extracted MAGs. The eight MAGs that were further analyzed in this study are given in bold letters.

Supp. Table 4. Excel table showing the KEGG pathways classified under different functional classes. Module IDs of modules corresponding to KEGG pathways, their completion ratio, and Q-values are given for MAGs investigated in this study.

References

- Abu Khweek A, Amer AO. Factors Mediating Environmental Biofilm Formation by *Legionella pneumophila*. *Frontiers in Cellular and Infection Microbiology* 2018; 8.
- Albertsen M, Hansen LBS, Saunders AM, Nielsen PH, Nielsen KL. A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. *Isme Journal* 2012; 6: 1094-1106.
- Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology* 2013; 31: 533-+.
- Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. *Nature Methods* 2014; 11: 1144-1146.
- Arai W, Taniguchi T, Goto S, Moriya Y, Uehara H, Takemoto K, et al. MAPLE 2.3.0: an improved system for evaluating the functionomes of genomes and metagenomes. *Bioscience Biotechnology and Biochemistry* 2018; 82: 1515-1517.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012; 19: 455-77.
- Bedoya K, Coltell-Simon O, Cabarcas F, Alzate JF. Metagenomic assessment of the microbial community and methanogenic pathways in biosolids from a municipal wastewater treatment plant in Medellin, Colombia. *Science of the Total Environment* 2019; 648: 572-581.

- 531 Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank.
532 Nucleic Acids Research 2013; 41: D36-D42.
- 533 Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nature
534 Methods 2015; 12: 59-60.
- 535 Campbell JH, O'Donoghue P, Campbell AG, Schwientek P, Sczyrba A, Woyke T, et al. UGA is
536 an additional glycine codon in uncultured SR1 bacteria from the human microbiota.
537 Proceedings of the National Academy of Sciences of the United States of America 2013;
538 110: 5540-5545.
- 539 Campbell P, Srinivasan R, Knoell T, Phipps D, Ishida K, Safarik J, et al. Quantitative structure-
540 activity relationship (QSAR) analysis of surfactants influencing attachment of a
541 Mycobacterium sp to cellulose acetate and aromatic polyamide reverse osmosis
542 membranes. Biotechnology and Bioengineering 1999; 64: 527-544.
- 543 Daims H, Lebedeva EV, Pjevac P, Han P, Herbold C, Albertsen M, et al. Complete nitrification
544 by Nitrospira bacteria. Nature 2015; 528: 504-+.
- 545 Dang HY, Lovell CR. Bacterial primary colonization and early succession on surfaces in marine
546 waters as determined by amplified rRNA gene restriction analysis and sequence analysis
547 of 16S rRNA genes. Applied and Environmental Microbiology 2000; 66: 467-475.
- 548 Delmont TO, Quince C, Shaiber A, Esen OC, Lee STM, Rappe MS, et al. Nitrogen-fixing
549 populations of Planctomycetes and Proteobacteria are abundant in surface ocean
550 metagenomes. Nature Microbiology 2018; 3: 804-+.
- 551 El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B. The abundance and variety of
552 carbohydrate-active enzymes in the human gut microbiota. Nature Reviews Microbiology
553 2013; 11: 497-504.

- Elifantz H, Horn G, Ayon M, Cohen Y, Minz D. Rhodobacteraceae are the key members of the microbial community of the initial biofilm formed in Eastern Mediterranean coastal seawater. *Fems Microbiology Ecology* 2013; 85: 348-357.
- Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *Peerj* 2015; 3.
- Gunsalus IC, Horecker BL, Wood WA. Pathways of Carbohydrate Metabolism in Microorganisms. *Bacteriological Reviews* 1955; 19: 79-128.
- Guo X, Wang X, Liu JX. Composition analysis of fractions of extracellular polymeric substances from an activated sludge culture and identification of dominant forces affecting microbial aggregation. *Scientific Reports* 2016; 6.
- Hinsa SM, Espinosa-Urgel M, Ramos JL, O'Toole GA. Transition from reversible to irreversible attachment during biofilm formation by *Pseudomonas fluorescens* WCS365 requires an ABC transporter and a large secreted protein. *Molecular Microbiology* 2003; 49: 905-918.
- Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Research* 2007; 17: 377-386.
- Huson DH, Beier S, Flade I, Gorska A, El-Hadidi M, Mitra S, et al. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *Plos Computational Biology* 2016; 12.
- Ibarbalz FM, Orellana E, Figuerola ELM, Erijman L. Shotgun Metagenomic Profiles Have a High Capacity To Discriminate Samples of Activated Sludge According to Wastewater Type. *Applied and Environmental Microbiology* 2016; 82: 5186-5196.

- 576 Jinhua P, Fukushi K, Yamamoto K. Bacterial community structure on membrane surface and
577 characteristics of strains isolated from membrane surface in submerged membrane
578 bioreactor. *Separation Science and Technology* 2006; 41: 1527-1549.
- 579 Jo SJ, Kwon H, Jeong SY, Lee CH, Kim TG. Comparison of microbial communities of activated
580 sludge and membrane biofilm in 10 full-scale membrane bioreactors. *Water Research*
581 2016; 101: 214-225.
- 582 Johnson DR, Lee TK, Park J, Fenner K, Helbling DE. The functional and taxonomic richness of
583 wastewater treatment plant microbial communities are associated with each other and
584 with ambient nitrogen and carbon availability. *Environmental Microbiology* 2015; 17:
585 4851-4860.
- 586 Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for
587 gene and protein annotation. *Nucleic Acids Research* 2016a; 44: D457-D462.
- 588 Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for
589 Functional Characterization of Genome and Metagenome Sequences. *Journal of*
590 *Molecular Biology* 2016b; 428: 726-731.
- 591 Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version
592 7.0 for Bigger Datasets. *Mol Biol Evol* 2016; 33: 1870-4.
- 593 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 2012; 9:
594 357-U54.
- 595 Laspidou CS, Rittmann BE. A unified theory for extracellular polymeric substances, soluble
596 microbial products, and active and inert biomass. *Water Research* 2002; 36: 2711-2720.
- 597 Le S, Josse J, Husson F. FactoMineR: An R package for multivariate analysis. *Journal of*
598 *Statistical Software* 2008; 25: 1-18.

- 599 Lee S, Park SK, Kwon H, Lee SH, Lee K, Nahm CH, et al. Crossing the Border between
600 Laboratory and Field: Bacterial Quorum Quenching for Anti-Biofouling Strategy in an
601 MBR. *Environmental Science & Technology* 2016; 50: 1788-1795.
- 602 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
603 Alignment/Map format and SAMtools. *Bioinformatics* 2009; 25: 2078-2079.
- 604 Lorite GS, Rodrigues CM, de Souza AA, Kranz C, Mizaikoff B, Cotta MA. The role of
605 conditioning film formation and surface chemical changes on *Xylella fastidiosa* adhesion
606 and biofilm evolution. *Journal of Colloid and Interface Science* 2011; 359: 289-295.
- 607 Luca Fortunato NP, Zahid Ur Rehman, Hokyong Shon, TorOve Leiknes. Real-time monitoring
608 of membrane fouling development during early stages of activated sludge membrane
609 bioreactor operation. *Process Safety and Environmental Protection* 2018; 120: 313-320.
- 610 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
611 *EMBnet.journal* 2011; 17.
- 612 Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: PAired-eND
613 Assembler for Illumina sequences. *Bmc Bioinformatics* 2012; 13.
- 614 Matar GK, Bagchi S, Zhang K, Oerther DB, Saikaly PE. Membrane biofilm communities in full-
615 scale membrane bioreactors are not randomly assembled and consist of a core
616 microbiome. *Water Research* 2017; 123: 124-133.
- 617 McSwain BS, Irvine RL, Hausner M, Wilderer PA. Composition and distribution of extracellular
618 polymeric substances in aerobic flocs and granular sludge. *Applied and Environmental*
619 *Microbiology* 2005; 71: 1051-1057.

- 620 Miura Y, Hiraiwa MN, Ito T, Itonaga T, Watanabe Y, Okabe S. Bacterial community structures
621 in MBRs treating municipal wastewater: Relationship between community stability and
622 reactor performance. *Water Research* 2007a; 41: 627-637.
- 623 Miura Y, Watanabe Y, Okabe S. Membrane biofouling in pilot-scale membrane bioreactors
624 (MBRs) treating municipal wastewater: Impact of biofilm formation. *Environmental*
625 *Science & Technology* 2007b; 41: 632-638.
- 626 Nagaraj V, Skillman L, Ho G, Li D, Gofton A. Characterisation and comparison of bacterial
627 communities on reverse osmosis membranes of a full-scale desalination plant by bacterial
628 16S rRNA gene metabarcoding. *Npj Biofilms and Microbiomes* 2017; 3.
- 629 Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid
630 Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids*
631 *Research* 2014; 42: D206-D214.
- 632 Pazda M, Kumirska J, Stepnowski P, Mulkiewicz E. Antibiotic resistance genes identified in
633 wastewater treatment plant systems - A review. *Sci Total Environ* 2019; 697: 134023.
- 634 Rabus R, Kube M, Beck A, Widdel F, Reinhardt R. Genes involved in the anaerobic degradation
635 of ethylbenzene in a denitrifying bacterium, strain EbN1. *Archives of Microbiology*
636 2002; 178: 506-516.
- 637 Rehman ZU, Ali M, Iftikhar H, Leiknes T. Genome-resolved metagenomic analysis reveals roles
638 of microbial community members in full-scale seawater reverse osmosis plant. *Water Res*
639 2019; 149: 263-271.
- 640 Rehman ZU, Leiknes T. Quorum-Quenching Bacteria Isolated From Red Sea Sediments Reduce
641 Biofilm Formation by *Pseudomonas aeruginosa*. *Frontiers in Microbiology* 2018; 9.

- 642 Rossellomora RA, Wagner M, Amann R, Schleifer KH. The Abundance of Zoogloea-Ramigera
643 in Sewage-Treatment Plants. *Applied and Environmental Microbiology* 1995; 61: 702-
644 707.
- 645 Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; 30: 2068-9.
- 646 Sheng GP, Yu HQ, Li XY. Extracellular polymeric substances (EPS) of microbial aggregates in
647 biological wastewater treatment systems: A review. *Biotechnology Advances* 2010; 28:
648 882-894.
- 649 Tian M, Zhao FQ, Shen X, Chu KH, Wang JF, Chen S, et al. The first metagenome of activated
650 sludge from full-scale anaerobic/anoxic/oxic (A2O) nitrogen and phosphorus removal
651 reactor using Illumina sequencing. *Journal of Environmental Sciences* 2015; 35: 181-190.
- 652 Tsang PH, Li GL, Brun YV, Ben Freund L, Tang JX. Adhesion of single bacterial cells in the
653 micronewton range. *Proceedings of the National Academy of Sciences of the United*
654 *States of America* 2006; 103: 5764-5768.
- 655 Wang XH, Hu M, Xia Y, Wen XH, Ding K. Pyrosequencing Analysis of Bacterial Diversity in
656 14 Wastewater Treatment Systems in China. *Applied and Environmental Microbiology*
657 2012; 78: 7042-7047.
- 658 Wang ZW, Wu ZC, Tang SJ. Extracellular polymeric substances (EPS) properties and their
659 effects on membrane fouling in a submerged membrane bioreactor. *Water Research*
660 2009; 43: 2504-2512.
- 661 Ye L, Zhang T, Wang TT, Fang ZW. Microbial Structures, Functions, and Metabolic Pathways
662 in Wastewater Treatment Bioreactors Revealed Using High-Throughput Sequencing.
663 *Environmental Science & Technology* 2012; 46: 13244-13252.

- 664 Yu K, Zhang T. Metagenomic and Metatranscriptomic Analysis of Microbial Community
665 Structure and Gene Expression of Activated Sludge. Plos One 2012; 7.
- 666 Zhang B, Xu XY, Zhu L. Structure and function of the microbial consortia of activated sludge in
667 typical municipal wastewater treatment plants in winter. Scientific Reports 2017; 7.
- 668 Ziegler AS, McIlroy SJ, Larsen P, Albertsen M, Hansen AA, Heinen N, et al. Dynamics of the
669 Fouling Layer Microbial Community in a Membrane Bioreactor. Plos One 2016; 11.

670

671

672

Bin_ID	Bin_name	Relative abundance (S1, S2)	Relative abundance (B1, B2)	Length	Con tigs	N50	GC content	Completion (%)	Redundancy (%)
Bin3_1	MBR_BPTB	0.57, 0.43	0, 0	3.87E+06	294	2.13E+04	67.27	97.84	3.60
Bin6	MBR_CLF	0.46, 0.48	0.04, 0.03	4.95E+06	80	1.34E+05	53.69	96.40	0.00
Bin7_1	MBR_BCT	0.44, 0.39	0.08, 0.09	3.87E+06	56	1.31E+05	39.44	97.12	0.72
Bin8	MBR_PLN	0.37, 0.4	0.15, 0.09	4.12E+06	18	5.24E+05	70.93	97.12	2.88
Bin4_1	MBR_GPTB	0.3, 0.25	0.23, 0.21	3.47E+06	197	3.29E+04	61.83	97.12	0.72
Bin10_2	MBR_CSP1	0.23, 0.24	0.29, 0.24	1.28E+06	35	5.39E+04	51.69	70.50	7.91
Bin10_3	MBR_CSP2	0.24, 0.22	0.28, 0.26	1.09E+06	14	1.52E+05	50.43	73.38	3.60
Bin14_1	MBR_APTB	0.2, 0.18	0.31, 0.32	1.95E+06	37	1.14E+05	48.51	97.12	0.00

673

674

- 675 • Membrane biofouling is a considerable problem in wastewater treatment processes
- 676 • Microbial community of the sludge and eight days old biofilm was investigated
- 677 • Proteobacteria and Bacterial candidate phyla are abundant in biofilm
- 678 • Functional potential of microorganism differs in sludge and biofilm
- 679 • Adhesin protein pathway was only detected in genome abundant in biofilm
- 680
- 681

682
683

Journal Pre-proofs

Functional class	Pathway	Module ID	MBR_EPTB	MBR_CLF	MBR_BCT	MBR_PLN	MBR_OPTB	MBR_APTB	MCR (%)
Central carbohydrate metabolism	Citrate cycle (TCA cycle, Krebs cycle)	M00009							
	Citrate cycle, first carbon oxidation, oxaloacetate => 2-oxoglutarate	M00010							
	Citrate cycle, second carbon oxidation, 2-oxoglutarate => oxaloacetate	M00011							
	Entner-Doudoroff pathway, glucose-6P => glyceraldehyde-3P + pyruvate	M00008							
	Gluconeogenesis, oxaloacetate => fructose-6P	M00003							
	Glycolysis, core module involving three-carbon compounds	M00002							
	Pentose phosphate pathway, non-oxidative phase, fructose 6P => ribose 5P	M00007							
	PRPP biosynthesis, ribose 5P => PRPP	M00005							
	Pyruvate oxidation, pyruvate => acetyl-CoA	M00307							
Nitrogen metabolism	Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	M00001							
	Denitrification, nitrate => nitrogen	M00529							
	Dissimilatory nitrate reduction, nitrate => ammonia	M00530							
Metabolic capacity	Nitrate assimilation	M00615							
	Sulfate-sulfur assimilation	M00616							
Sulfur metabolism	Assimilatory sulfate reduction, sulfate => H ₂ S	M00176							
Fatty acid metabolism	beta-Oxidation	M00087							
	beta-Oxidation, acyl-CoA synthesis	M00086							
	Fatty acid biosynthesis, elongation	M00083							
	Fatty acid biosynthesis, initiation	M00082							
Bacterial secretion system	Sec (secretion) system	M00335							
	Twin-arginine translocation (Tat) system	M00336							
	Adhesin protein transport system	M00330							
Drug resistance	Aminoglycoside resistance, protease FtsH	M00742							
	Aminoglycoside resistance, protease HtpX	M00743							
	Imipenem resistance, repression of porin OprD	M00745							
	Multidrug resistance, efflux pump MexJK-OprM	M00642							
Two-component regulatory system	AlgZ-AlgR (alginate production) two-component regulatory system	M00493							
	EnvZ-OmpR (osmotic stress response) two-component regulatory system	M00445							
	FlrB-FlrC (polar flagellar synthesis) two-component regulatory system	M00515							
	PhoR-PhoB (phosphate starvation response) two-component regulatory system	M00434							
	PilS-PilR (type 4 fimbriae synthesis) two-component regulatory system	M00501							
	ChvG-ChvI (acidity sensing) two-component regulatory system	M00520							

MCR (%)

100

75

50

25

0

686

687