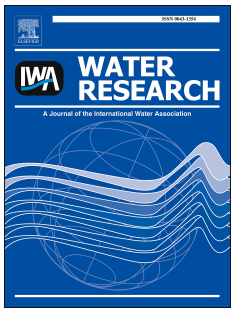
Membrane biofilm communities in full-scale membrane bioreactors are not randomly assembled and consist of a core microbiome

Accepted Manuscript



Membrane biofilm communities in full-scale membrane bioreactors are not randomly assembled and consist of a core microbiome

Gerald K. Matar, Samik Bagchi, Kai Zhang, Daniel B. Oerther, Pascal E. Saikaly

PII: S0043-1354(17)30529-8

DOI: 10.1016/j.watres.2017.06.052

Reference: WR 13007

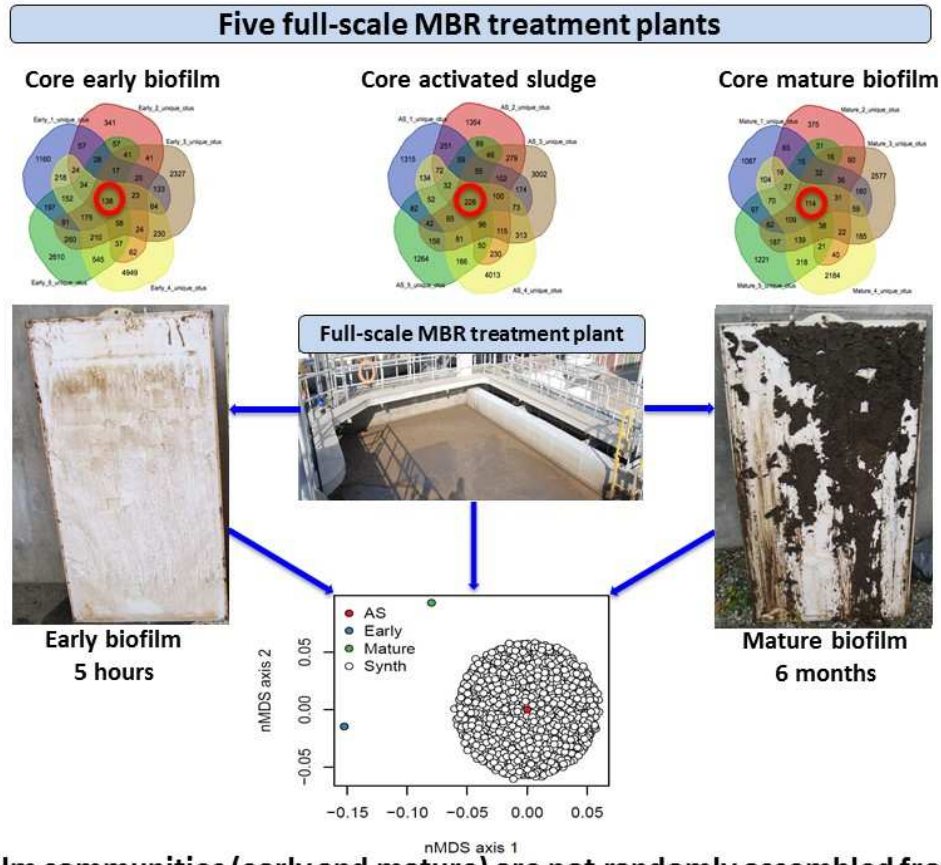
To appear in: *Water Research*

Received Date: 6 February 2017

Revised Date: 12 May 2017

Accepted Date: 19 June 2017

Please cite this article as: Matar, G.K., Bagchi, S., Zhang, K., Oerther, D.B., Saikaly, P.E., Membrane biofilm communities in full-scale membrane bioreactors are not randomly assembled and consist of a core microbiome, *Water Research* (2017), doi: 10.1016/j.watres.2017.06.052.



Biofilm communities (early and mature) are not randomly assembled from AS

Membrane biofilm communities in full-scale membrane bioreactors are not randomly assembled and consist of a core microbiome

Gerald K. Matar[a†], Samik Bagchi[a†¥], Kai Zhang[b], Daniel B. Oerther[c], Pascal E. Saikaly[a,*]

[a]King Abdullah University of Science and Technology, Biological and Environmental Sciences and Engineering Division, Water Desalination and Reuse Research Center, Thuwal 23955-6900, Saudi Arabia

[b]Baswood Corporation, Allen, Texas 75013, USA

[c]Department of Civil, Architectural, and Environmental Engineering, and Environmental Research Center, Missouri University of Science and Technology, Rolla, Missouri 65409, USA

*Corresponding author: Pascal E. Saikaly, pascal.saikaly@kaust.edu.sa; Tel.: +966–2–808-4903

[†]Contributed equally to this work.

¥ Present address: University of Kansas, Department of Civil, Environmental, and Architectural Engineering, Lawrence, KS, USA 66045

Abstract

Finding efficient biofouling control strategies requires a better understanding of the microbial ecology of membrane biofilm communities in membrane bioreactors (MBRs). Studies that characterized the membrane biofilm communities in lab-and pilot-scale MBRs are numerous, yet similar studies in full-scale MBRs are limited. Also, most of these studies have characterized the mature biofilm communities with very few studies addressing early biofilm communities. In this study, five full-scale MBRs located in Seattle (Washington, U.S.A.) were selected to address two questions concerning membrane biofilm communities (early and mature): (i) Is the assembly of biofilm communities (early and mature) the result of random immigration of species from the source community (i.e. activated sludge)? and (ii) Is there a core membrane biofilm community in full-scale MBRs? Membrane biofilm (early and mature) and activated sludge (AS) samples were collected from the five MBRs, and 16S rRNA gene sequencing was applied to investigate the bacterial communities of AS and membrane biofilms (early and mature). Alpha and beta diversity measures revealed clear differences in the bacterial community structure between the AS and biofilm (early and mature) samples in the five full-scale MBRs. These differences were mainly due to the presence of large number of unique but rare operational taxonomic units (~13% of total reads in each MBR) in each sample. In contrast, a high percentage (~87% of total reads in each MBR) of sequence reads was shared between AS and biofilm samples in each MBR, and these shared sequence reads mainly belong to the dominant taxa in these samples. Despite the large fraction of shared sequence reads between AS and biofilm samples, simulated biofilm communities from random sampling of the respective AS community revealed

that biofilm communities differed significantly from the random assemblages ($P < 0.001$ for each MBR), indicating that the biofilm communities (early and mature) are unlikely to represent a random sample of the AS community. In addition to the presence of unique operational taxonomic units in each biofilm sample (early or mature), comparative analysis of operational taxonomic units and genera revealed the presence of a core biofilm community in the five full-scale MBRs. These findings provided insight into the membrane biofilm communities in full-scale MBRs. More comparative studies are needed in the future to elucidate the factors shaping the core and unique biofilm communities in full-scale MBRs.

Keywords Biofouling; membrane bioreactor; activated sludge; early biofilm; mature biofilm; 16S rRNA gene sequencing

1. Introduction

Providing adequate supply of clean fresh water as the world's population increases is one of the grand challenges facing society in the current century. One possible solution to address this challenge is to recover clean water for reuse from wastewater using membrane bioreactors (MBRs). The MBR offers several advantages over conventional activated sludge (AS) process, such as producing less sludge and providing high quality permeate without encountering a large footprint (Miura et al., 2007; Wang et al., 2009). Despite these advantages, membrane fouling, particularly biofouling remains a major hindrance to the wide spread application of MBRs. Several control strategies have been suggested to mitigate biofouling in MBRs including physical cleaning (e.g. back-washing, back-pulsing, air sparging), chemical cleaning (e.g. acids, bases, oxidants, chelating agents, polymeric coagulants, surfactants), membrane modification (e.g. charge, hydrophobicity, roughness), and biological-based antifouling strategies (e.g. quorum quenching, enzymatic disruption, energy uncoupling, and biofilm disruption by adding bacteriophage) (Malaeb et al., 2013). However, these strategies often fail to adequately control biofouling. Finding more efficient strategies to control biofouling requires a more fundamental understanding of the factors that shape membrane biofilm community assembly in MBRs.

Several sequential steps are generally considered to be involved in the progression of biofilm formation on surfaces, beginning with the formation of a conditioning film followed by a series of ordered processes: (i) attachment of pioneer colonizers onto surfaces; (ii) growth of pioneer colonizers, which change the surface characteristics of the substratum and facilitate the attachment of new organisms resulting in early biofilm

formation; and (iii) subsequent development to mature biofilms (Dang and Lovell, 2000; Zhang et al., 2006; Bereschnko et al., 2010). This sequence of events in the colonization of surfaces is well understood for human dental plaque and other solid surfaces (Dang and Lovell, 2000; Davey et al., 2000; Costerton, 2007; Kjelleberg et al., 2007), and it has been observed in the colonization of reverse osmosis (RO) membrane and spacer surfaces (Bereschnko et al., 2010). However, this detailed level of understanding on biofilm formation on membrane surfaces in MBRs is less studied.

Studies in lab- and full-scale AS process (Ofiţeru et al., 2010; Ayarza and Erijman 2011; Valentin-Vargas et al., 2012; Bagchi et al., 2015; Vuono et al., 2015; Meerburg et al., 2016; Saunders et al., 2016) suggest that both local (environmental and operational condtions, biotic interactions) and regional (dispersal or ~~the~~ propagation and immigration of biota) processes regulate the assembly of AS microbial community. By viewing biofilms as microbial landscapes and adopting metacommunity ecology as a framework to elucidate the mechanisms underlying biofilm community assembly in streams, Besemer et al. (2012) and Wilhelm et al. (2013) showed that stochastic dispersal from the source community was unlikely to shape biofilm communities in streams, and species sorting by local environmental conditions was the key mechanism underlying biofilm community assembly. Biofilms in streams assemble from different sources (e.g. soil and groundwater) in the catchment (Besemer et al., 2012). In contrast, biofilms in biological wastewater treatment plants such as MBRs mainly assemble from one source (i.e. AS). However, a large knowledge gap exists whether biofilm communities on membrane surfaces of MBRs assemble because of local or regional processes.

Previous studies characterizing the biofilm microbial community in MBRs (Lim et al., 2004; Choi et al., 2006; Jinhua et al., 2006; Zhang et al., 2006; Miura et al., 2007; Huang et al., 2008; Fontanos et al., 2010; Lim et al., 2012; Piasecka et al., 2012; Lee et al., 2014, Jo et al., 2016), showed that the biofilm microbial community was distinct from the AS community. Despite these numerous studies, we still know little of the bacteria that form biofilms on membrane surfaces of MBRs. This is mainly due to the fact that the majority of these studies were conducted in lab-scale MBRs where conditions are not as complex as in full-scale systems, and with very few on pilot-scale MBRs (Jinhua et al., 2006; Miura et al., 2007) and one study in full-scale MBRs (Jo et al., 2016). Most community ecology studies of full-scale biological wastewater treatment plants have been limited to microbial diversity surveys of AS communities, and a major finding of these studies was the existence of some core AS communities shared between geographically distributed biological wastewater treatment plants (Xia et al., 2010; Zhang et al., 2012; Wang et al., 2012), which raises the question of whether a core biofilm community also exists in geographically distributed full-scale MBRs.

While most previous studies have characterized the microbial communities in mature biofilms (Jinhua et al., 2006; Miura et al., 2007; Huang et al., 2008; Fontanos et al., 2010; Lee et al., 2014; Jo et al., 2016), some researchers claim that characterizing the early colonizers on membrane surfaces might help develop better biofouling control strategies; yet, few studies have addressed these early colonizers (Choi et al., 2006; Zhang et al., 2006; Lim et al., 2012; Piasecka et al., 2012). Also, studies characterizing both the early colonizers and mature biofilm communities in MBRs are lacking.

This study was motivated by two questions concerning the membrane biofilm communities (early and mature) in full-scale MBRs: (i) Is the assembly of biofilm communities (early and mature) the result of random immigration of species from the AS community or the result of specific selection of certain species due to local conditions? (ii) Is there a core membrane biofilm community in full-scale MBRs? To address these questions, 16S rRNA gene sequencing combined with multivariate statistical analysis was applied to characterize the biofilm (early and mature) and AS bacterial communities in five full-scale MBRs located in the same city (Seattle, Washington, U.S.A.), and equipped with the same membrane type and treating predominantly domestic wastewater. To the best of our knowledge this is the first study to characterize both early and mature biofilm communities in full-scale MBRs.

2. Materials and methods

2.1. Full-scale MBRs and sample collection

Five full-scale MBRs were identified in the region of Seattle (Washington, U.S.A.) (Fig. S1). The five MBRs (referred to herein as MBR 1, 2, 3, 4 and 5) were equipped with KUBOTA flat-sheet microfiltration (MF) membranes (KUBOTA Membranes, USA) and treated predominantly domestic wastewater. Details of influent wastewater characteristics and operational parameters of the five MBRs were provided by the plant operators and are listed in Table S1. Old membrane modules that have been in operation for at least six months were removed from the membrane basin with a crane and duplicate membrane samples (5 cm^2 each) were sectioned from different locations on the membrane surfaces, on which mature biofilms have already been developed.

Immediately afterwards, new membrane modules were deployed in the membrane basin to replace the old membrane modules and after 5 hours of filtration, the new membrane modules were removed and duplicate membrane samples (5 cm^2 each) were sectioned from different locations on the membrane surfaces, on which early biofilms or colonizers have already been developed. The 5 hours of filtration was sufficient to observe a visible biofilm on the surface of the new membrane modules. The AS samples (20 mL each) were collected from the membrane basin of each MBR treatment plant at the same time when the early and biofilm samples were collected. All membrane and AS samples from the five full-scale MBRs were collected over a period of one week during the month of December (December 6 to 12) (Table S1). In total, 30 samples were collected including duplicate samples of early biofilms, mature biofilms and AS from the five full-scale MBR plants. All samples were immediately stored on ice and transported to the laboratory, where they were stored at -80^{o}C until further analysis.

2.2. DNA extraction, PCR and 16S rRNA gene sequencing

Before DNA extraction the membrane samples were rinsed with 1 × PBS (phosphate-buffered saline: 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, and 0.24 g KH_2PO_4 per liter distilled water, pH 7.4) to remove loosely deposited sludge (Huang et al., 2008). Genomic DNA was extracted from the mature biofilms, early biofilms and AS samples using the PowerSoil DNA extraction kit (MO BIO Laboratories, inc., Carlsbad, CA) according to the manufacturer's protocol. The quality (A260/A280) and quantity (A260) of the extracted genomic DNA was determined with a Nanodrop® 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA).

For each DNA sample, triplicate PCR reactions was performed in a 25–µl reaction volume using the HotStarTaq Plus Master Mix (QIAGEN, Valencia, CA), 0.5 µM of each primer and 100–200 ng of template DNA. The extracted DNA samples were amplified using the forward primer 8F (5'–AGAGTTTGATCCTGGCTCAG–3') and reverse primer 533R (5'–TTACCGCGGCTGCTGGCAC–3') (Lu et al., 2012). These primers targeted the V1–V3 region of the bacterial 16S rRNA gene. Barcodes that allow sample multiplexing during pyrosequencing were incorporated between the 454 adapter and the forward primer. PCR was performed using a C1000 Thermal Cycler (BIO-RAD, Hercules, CA) with the following PCR conditions: initial denaturation at 95^{o}C for 5 minutes, followed by 25 cycles of denaturation at 95^{o}C for 30 seconds, annealing at 55^{o}C for 30 seconds, and extension at 72^{o}C for 30 seconds. The PCR was completed with a final extension at 72^{o}C for 5 minutes (Lu et al., 2012).

The triplicate PCR products from each sample were pooled and confirmed by gel electrophoresis. Then, gel bands were excised and purified using the Qiaquick gel extraction kit (QIAGEN, Valencia, CA) according to the manufacturer's protocol. The concentration of the PCR products was measured on Qubit® 2.0 Fluorometer using the PicoGreen® dsDNA quantitation assay (Invitrogen, Carlsbad, CA). The purified barcoded amplicons were pooled in equimolar concentrations and sequenced on the Roche 454 FLX Titanium genome sequencer (Roche, Indianapolis, IN) at the Bioscience Core Laboratory at King Abdullah University of Science and Technology, according to the manufacturer's instructions.

2.3. Processing of sequencing data

189 The 16S rRNA gene amplicons were processed using the Quantitative Insights Into
190 Microbial Ecology (QIIME v1.7.0) pipeline (Caporaso et al., 2010b). All raw reads were
191 first denoised, filtered for quality check and demultiplexed to trim the barcodes, primers
192 and to remove low-quality sequence reads, such as sequences outside the bounds of 200
193 and 600 bp, sequences containing ambiguous bases, sequences with 6 homopolymers and
194 sequences with quality score below 25 (El-Chakhtoura et al., 2015). Chimeric sequences
195 were identified and removed from the sequences using Chimera Slayer as implemented in
196 QIIME. The sequences were clustered into operational taxonomic units (OTUs) using
197 UCLUST (Edgar, 2010), with 97% sequence identity threshold. Representative sequence
198 from each OTU was phylogenetically aligned using PyNAST (Caporaso et al., 2010a)
199 and assigned to a taxonomic identity using the Greengenes 13_5 database (DeSantis et al.,
200 2006).
201 To compensate for stochastic sampling efforts and reduce effects of variation among
202 replicates (Andrew et al., 2012), duplicate samples were pooled together to create
203 combined OTU files, resulting in 15 pooled samples. The OTU table was further
204 clustered based on biomass category into three subsets i.e. Early, Mature and AS or by
205 MBR plant into five subsets. Shared OTUs within each of the three subsets (i.e. AS,
206 Early or Mature) or each MBR plant was visualized by Venn diagram in R 'vegan scalpel'
207 program. The distribution of the different bacterial phyla and proteobacterial classes was
208 visualized in a heatmap using R 'vegan scalpel' program.
209 The sequencing reads were deposited into the Sequence Read Archive (SRA) of the
210 National Center for Biotechnology Information (NCBI) under study accession number
211 SRP064009.

212

213 *2.4. Alpha and beta diversity estimates*

214 For alpha diversity measurements, both non-phylogeny based metrics (observed
215 OTUs, Shannon diversity index (H) and Chao 1 richness estimator) and phylogeny based
216 metric (phylogenetic diversity (PD_whole)) were calculated with QIIME at the 3%
217 distance level for each pooled sample using rarefied OTU dataset. Community
218 comparisons between samples (beta diversity) was performed with unweighted UniFrac
219 and Bray-Curtis distance and visualized by principal coordinate analysis (PCoA) in
220 QIIME. To remove inherent heterogeneity of sampling depth, we subsampled the dataset
221 (normalized abundance values) to an even depth of 4,000 sequences across the pooled
222 samples. This number was chosen, as it was slightly less than the pooled sample with the
223 lowest reads (i.e. Early biofilm from MBR 2, which had 4,007 reads). We also assessed
224 the beta diversity of total, dominant and rare taxa for the pooled samples. Rare taxa were
225 defined as OTUs that encompass ≤ 20 sequences (i.e. $\leq 0.5\%$) (Bagchi et al., 2015).
226 Unweighted UniFrac distance was calculated for the total, dominant and rare OTUs and
227 visualized by nonmetric multidimensional scaling (NMDS) using the software PRIMER
228 6 (version 6.1.13) and PERMANOVA+ add on (version 1.0.3) (PRIMER-E LTD, United
229 Kingdom).

230 The OTU table was separated based on biomass category into three subsets i.e. Early,
231 Mature and AS. Average unweighted UniFrac distance within and between Early, Mature
232 and AS communities was calculated for each category by distance comparison command
233 in QIIME.

234

2.5. Statistical analysis

Reproducibility between duplicate samples was evaluated by one way pairwise analysis of similarity (ANOSIM) based on Spearman's rank correlation at a 999 permutation using the statistical software PRIMER 6 (version 6.1.13) and PERMANOVA+ add on (version 1.0.3) (PRIMER-E LTD, United Kingdom). ANOSIM produces a test statistic (R) which can range from -1 to 1 (Rees et al., 2004). An R value of 0 indicates no separation in community structure and a value of 1 indicates separation (Ramette, 2007).

To estimate the probability that a biofilm community (early or mature) represents a random sample of the respective suspended community (i.e. AS), a random subsampling of the AS community from each MBR was done as described in Besemer et al. (2012). In brief, OTUs from each AS community were sampled with replacement until the number of OTUs in this randomly assembled community equaled the richness of the respective biofilm community. This procedure was repeated to yield 1,000 random subsamples of each AS community. The probability of the biofilm community to fall within the distribution of these random subsamples was calculated as the percentage of the distances of the random subsamples to their centroid (Besemer et al., 2012). The results of the random sampling procedure were visualized in NMDS.

3. Results

3.1. Alpha diversity measures

16S rRNA gene sequencing was conducted on 30 samples including duplicates from each type of sample (i.e., AS, early and mature biofilms). One-way pairwise analysis of

similarity (ANOSIM) showed high similarity (98.4%; R: -0.25) between duplicate samples. Spearman correlation coefficient was 92% ($p < 0.001$). A total of 743,970 high-quality reads were generated for the 15 pooled samples after denoising, quality filtering and removal of chimeric sequences. The sequences were clustered into 22,877 OTUs at a 97% sequence identity threshold.

The alpha diversity values of the pooled 15 samples using rarefied OTUs ranged as follows: observed OTUs (939-6,943), Chao 1 (1,765-8,113), H (6.41-8.22) and PD (74.59-226.01) (Table 1). All four indices (i.e. observed OTUs, Chao 1, H and PD) demonstrated that the early biofilm samples have higher diversity than the mature biofilm samples among the five MBR plants. Similarly, AS samples had higher diversity than mature biofilm samples except for MBR 3 where Chao 1, H and PD were higher for mature biofilms than AS. No clear trend in diversity was observed between AS and early biofilm samples. Good's coverage (84.20-97.91%, averaging 95%) revealed that the 16S rRNA gene sequences identified in these samples represent the majority of bacterial diversity present in each sample.

3.2. Beta diversity measures

The bacterial communities in the five MBRs were compared using both phylogenetic (unweighted UniFrac) and non-phylogenetic (Bray-Curtis distance) measures. The PCoA results based on unweighted UniFrac distance revealed that the bacterial communities in the 15 pooled samples were clustered into five groups with AS and biofilm (early and mature) samples from the same MBR plant grouped together (Fig. 1). Similar results were obtained using Bray-Curtis distance at 3% cutoff-OTU level (Fig. S2).

281 To compare the bacterial communities in the five MBR plants based on total,
282 dominant and rare OTUs, the bacterial community in each sample was separated into rare
283 (blue triangles), dominant (red squares) and total taxa (green triangles) and visualized in
284 NMDS plot generated based on unweighted UniFrac distance (Fig. 2). Rare OTUs were
285 defined as OTUs with relative abundance ≤ 0.5% (Bagchi et al., 2015). The NMDS
286 results showed that the bacterial communities in the five MBR plants were more
287 dispersed based on rare OTUs than the total and dominant OTUs as can be seen by their
288 wide distribution in the NMDS plot (Fig. 2). Also, the total and dominant bacterial taxa
289 were clustered together. These results suggested that the difference in the bacterial
290 communities in the five MBR plants was mainly due to differences in the community
291 structure of the rare OTUs.

292 Although PCoA (Fig. 1 and Fig. S2) and NMDS analysis (Fig. 2) showed that the AS
293 and biofilm samples from each MBR were clustered together, comparison of unweighted
294 UniFrac distance between samples in different categories revealed that AS samples from
295 the five MBRs were highly dissimilar from the biofilm samples (early and mature), and
296 early biofilms were dissimilar from mature biofilms (Fig. 3). Also, high dissimilarity was
297 observed between samples within the same category (i.e. AS, Early or Mature) (Fig. 3).
298

299 *3.3. Effect of source community*

300 To estimate the probability that the biofilm communities (early and mature) represent
301 random samples of their respective AS communities, the biofilm communities were
302 compared to 1,000 random subsamples of the AS communities and the results were
303 visualized on NMDS plot based on the Horn Index (Fig. 4). In all five MBRs, the biofilm

304 differed significantly from the random assemblages ($P < 0.001$ for each MBR), indicating
305 that the biofilm communities (early and mature) are unlikely to represent a random
306 sample of the AS community.

307

308 *3.4. Shared and core genera/OTUs*

309 Using PyNAST with the Greengenes database as a reference, 100%, 87%, 68%, 52%
310 and 32% of the V1-V3 16S rRNA gene pyrotags could be assigned to the phylum, class,
311 order, family and genus level, respectively. The AS and biofilm samples (early and
312 mature) were allocated to 13 phyla, 21 classes and 382 genera. The dominant phylum
313 across the 15 samples was *Proteobacteria* (47.4%), followed by *Bacteroidetes* (13.9%),
314 *Actinobacteria* (9.7%), *Acidobacteria* (6.0%), *Chloroflexi* (5.7%), *Nitrospira* (3.8%),
315 *OD1* (3.3%), *TM7* (2.8%), *Firmicutes* (2.4%), *Gemmatimonadetes* (2.0%) and
316 *Planctomycetes* (2.0%) (Fig. S3). The numbers in parentheses represent the averages of
317 all 15 samples (i.e. AS, early and mature) collected from the five MBRs. The phylum
318 *Bacteroidets* was relatively more dominant in MBR 4 and 5, while the phylum
319 *Chloroflexi* was more abundant in MBR 4. The phylum *Actinobacteria* was relatively
320 more dominant in early (13.6%) and mature (9.9%) biofilms than AS (5.5%) samples
321 (Fig. S3). Within *Proteobacteria*, *Alphaproteobacteria* (20.9%; 18.7%; 28.8%) was the
322 dominant class, followed by *Betaproteobacteria* (18.4%; 18.1%; 13.7%),
323 *Gammaproteobacteria* (4.3%; 3.1%; 3.6%) and *Deltaproteobacteria* (1.9%; 1.9%; 1.5%)
324 (Fig. S4). The numbers in parentheses represent the averages of AS, early and mature
325 samples collected from the five MBRs, respectively.

In the current study, core indicate shared membership (genera or OTUs) across all 5 samples in the same category (i.e. AS, early or mature) (Wang et al., 2012; Zhang et al., 2012). Of the 382 classified genera (32% of sequence reads), 83, 63 and 50 core genera were detected in AS, early biofilm and mature biofilm samples, respectively (Table S2). The relative abundance of the dominant core genera in each category are presented in Fig. 5. This resulted in 30 genera that were common to all 15 samples, but their relative abundance varied between the different sample categories (i.e. AS, early or mature). The 30 core genera mainly belonged to the *Proteobacteria* (*Alphaproteobacteria* and *Betaproteobacteria*), *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Chloroflexi*, *Gemmatimonadetes*, *Nitrospira* and *Planctomycetes* phyla.

To assess the number of core OTUs within the same category (i.e. AS, early or mature), the five AS, five early or five mature biofilm samples, from the five different MBRs were combined together. Of the 14,090, 14,323 and 9,518 total observed OTUs, only 228 OTUs (1.62%), 138 OTUs (0.96%) and 114 OTUs (1.20%) were shared respectively by the five combined AS, early biofilm and mature biofilm samples (Table 2, Fig. S5). However, these core OTUs comprise a high fraction of the total number of sequence reads in the AS (35.17%), early biofilm (27.94%) and mature biofilm (25.80%) samples, respectively (Table 2). Based on the Venn diagrams (Fig. S5), the unique OTUs (i.e. those found in only one sample) for the five combined AS, early biofilm and mature biofilm samples were 10,948 OTUs, 11,387 OTUs and 7,444 OTUs, respectively. These correspond to 77.70% (AS), 79.50% (early biofilm) and 78.20% (mature biofilm) of the total observed OTUs in each category (Table 2). By comparing the ratio of the number of sequence reads to the number of core or unique OTUs within the same category (i.e. AS,

early or mature), the core OTUs corresponded to the dominant OTUs (332 to 629 reads per OTU; averaging 468 reads per OTU), whereas the unique OTUs corresponded to the rare OTUs (15 to 20 reads per OTU; averaging 17 reads per OTU) (Andrew et al., 2012).

Despite the fact that the AS, early and mature biofilm in each MBR harbored a large number of unique OTUs (Fig. S6), the percentage of shared OTUs between the three samples (i.e. AS, early and mature) within each MBR was high ranging from 17.15%-41.46% (Table 2). These shared OTUs comprise a high fraction (52.01%-94.99%, averaging 85.74%) of the total number of reads (Table 2). Also, the shared OTUs correspond to the dominant OTUs (22 to 72 reads per OTU; averaging 49 reads per OTU) in each MBR plant, whereas the unique OTUs correspond to the rare OTUs (3 to 10 reads per OTU; averaging 4 reads per OTU).

4. Discussion

The aim of this study was to evaluate if the assembly of biofilm communities in full-scale MBRs is random or the result of species sorting, and to determine if a core biofilm (early and mature) community exists in full-scale MBRs.

4.1. Is the assembly of biofilm community (early and mature) the result of random immigration of species from the AS community or the result of specific selection of certain species due to local conditions?

Alpha (Table 1) and beta diversity measures using unweighted UniFrac distance (Fig. 3) revealed clear differences in the bacterial community diversity between the AS and early (after only 5 h of filtration) and mature biofilm samples in the full-scale MBRs.

372 This difference was due to the difference in the assembly mechanism of AS and biofilm
373 communities, which are two distinct forms of microbial aggregates. Similar results were
374 reported in previous studies where the biofilm community was distinct from the AS
375 community in lab- (Lim et al., 2004; Choi et al., 2006; Zhang et al., 2006; Huang et al.,
376 2008; Fontanos et al., 2010; Lim et al., 2012; Piasecka et al., 2012; Lee et al., 2014),
377 pilot- (Jinhua et al., 2006; Miura et al., 2007) and full-scale (Jo et al., 2016) MBRs.
378 Based on the Venn diagrams, the difference between AS and biofilm communities (early
379 and mature) was mainly attributed to the presence of a large number of unique OTUs in
380 each sample (Fig. S6). These unique OTUs represent the rare OTUs in the community as
381 they correspond to a small fraction (averaging 12.9%) of the total sequence reads in each
382 MBR plant. It has been theorized that rare species are regarded as a 'seed bank' (i.e. a
383 reserve of taxa that survive in an ecosystem at low abundance and low activities) that
384 may become abundant when the conditions are favorable (Pedrós-Alió, 2006; Saikaly and
385 Oerther, 2011). However, this should not be taken as a rule to suggest that these rare
386 OTUs are of little importance to the community. For example, Musat et al. (2008)
387 showed that the least abundant species (~0.3% of the total cell number) contributed to
388 more than 40% and 70% of the total uptake of ammonium and carbon, respectively in the
389 oligotrophic, meromictic Lake Cadagno. In contrast, a high percentage (averaging
390 87.1%) of sequence reads was shared between the AS and biofilm samples (early or
391 mature) in each MBR plant (Table 2) and these shared sequence reads mainly belong to
392 the dominant OTUs in these samples. These results agree with previous studies in full-
393 scale MBRs (Jo et al., 2016), and freshwater (Bereschenko et al., 2008) and seawater

394 (Zhang et al., 2011) RO plants, where the biofilm communities on membrane surfaces
395 best resembled the source community (i.e. AS, freshwater or seawater microbes).
396 Biofilm community in MBRs may also assemble from the microbial community
397 present in the influent domestic wastewater. The influent wastewater community was not
398 sampled in the current study. However, the 16S rRNA gene sequencing results clearly
399 showed that a large fraction of sequence reads (averaging 87.1%) was shared between the
400 AS and biofilm (early and mature) communities suggesting that the AS community
401 mainly contributed to the assembly of biofilms on the membrane surfaces of the full-scale
402 MBRs in the current study. Saunders et al. (2016) showed that immigration from the
403 influent wastewater had a modest impact on activated sludge community in full-scale AS
404 wastewater treatment plants. Also, Vuono et al. (2016) showed in a full-scale AS study
405 that only during disturbance (lowering the SRT by increasing the biomass wasting rate)
406 some of the most abundant bacteria in the immigrant community (i.e. influent
407 wastewater) colonized the AS community and in few cases, became dominant.
408 The fact that a large fraction of sequence reads was shared between the AS and
409 biofilm (early and mature) communities does not suggest that the biofilm community is a
410 mere reflection of the AS community or a simple concentration of bacteria present in the
411 AS. Battin et al. (2007) suggested viewing biofilms as microbial landscapes, which
412 offered an opportunity to microbial ecologists to study biofilm community assembly
413 according to the metacommunity ecology theory, which states that local and regional
414 processes regulate the assembly of local communities (Leibold et al., 2004; Holyoak et al.,
415 2005). In this context, we found that stochastic dispersal or immigration from AS was
416 unlikely to shape the biofilm (early or mature) community structure on membrane

surfaces (Fig. 4). This suggests that species sorting by the local environmental, operational and biotic conditions likely selected microorganisms from AS for biofilm formation. This species sorting by local conditions resulted in the presence of unique OTUs (rare taxa) in the early and mature biofilms (Fig. S6) and in different relative abundances of shared genera (dominant taxa) between the AS and biofilms (Fig. 5).

It has been suggested that initial colonization of surfaces in natural environments such as lakes and streams is likely to be stochastic (Jackson et al., 2001; Besemer et al., 2007), as it mainly depends on immigration from the source community. However, this might not be a general rule as we showed in the current study that biofilm formation on virgin membrane surfaces in MBRs after a short period of filtration (5 h) was not stochastic (Fig. 4). Bereschenko et al. (2008, 2010) identified *Sphingomonas* spp. as the key microorganisms responsible for initiating membrane surface colonization in full-scale freshwater RO treatment plant because of their competitive advantage in this environment, suggesting that initial colonization is not stochastic. Tan et al. (2014) reported that initiation of granulation from AS in aerobic granular biofilm reactor is not random, and was positively correlated with quorum sensing (QS) signaling. Besemer et al. (2012) and Wilhelm et al. (2013) showed that species sorting by local environmental conditions was the major mechanism for shaping biofilm community structure in natural environments such as streams (Besemer et al., 2012; Wilhelm et al., 2013). Collectively, these results indicate that local conditions rather than regional processes regulate assembly of biofilm communities in natural and engineered ecosystems.

4.2. Is there a core membrane biofilm community in full-scale MBRs?

Comparative analysis of OTUs (Table 2) and genera (Table S2) revealed the presence of a core biofilm (early and mature) and AS community across the five full-scale MBRs. Although the shared OTUs between the biofilm samples (early or mature) in the 5 MBRs was < 2%, these shared OTUs represented the dominant taxa and corresponded to a high fraction (averaging 26.9%) of shared sequence reads between the biofilm communities (Table 2). Classification of the 30 abundant core genera (AS and biofilm) across the 5 full-scale MBRs (Fig. 5) showed the presence of genera that were also observed in full-scale AS systems in Asia (China, Hong Kong and Singapore), North America (Canada and United States) and Europe (Denmark), including *Dechloromonas*, *Flavobacterium*, *Gordonia*, *Galdilinea*, *Gemmatimonas*, *Mycobacterium*, *Nitrospira*, *Tetrasphaera*, *Thauera*, and *Zooglea* (Wang et al., 2012; Zhang et al., 2012; Jo et al., 2016; Saunders et al., 2016). Zhang et al. (2012) identified a set of core genera (AS) shared by 14 full-scale AS treatment plants from distinct geographic locations (Asia and North America), and operated using different process configurations and used to treat sewage with different characteristics (i.e. chemical oxygen demand, total nitrogen, total phosphorous, pH and conductivity). Similarly, Wang et al (2012) identified 60 core genera (AS) shared by 14 full-scale AS treatment plants from different cities in China and operated under different conditions (dissolved oxygen, temperature, SRT and MLSS) and treated sewage with different characteristics. Jo et al. (2016) detected 20 dominant core genera in the biofilm and AS community in 10 full-scale MBRs in China despite significant differences in environmental factors (e.g. flux, hydraulic retention time, solid retention time, specific aeration demand, membrane type, wastewater characteristics, and mixed liquor suspended solids). In the current study, the 5 full-scale MBRs were selected from the

463 same city (Seattle, U.S.A.), equipped with the same type of membrane (KUBOTA flat-
464 sheet MF membranes) that were designed to operate under the same flux and air-scouring
465 rate, sampled during the same period (December 6 to 12) (Table S1), and treated
466 predominately domestic wastewater. Therefore, it was not surprising to observe a large
467 number of common genera between the samples (AS or biofilm) (Table S2). It should be
468 noted that 13 (*Arcrobacter*, *Caldilinea*, *Dechloromonas*, *Flavobacterium*, *Gordonia*,
469 *Haliscomenobacter*, *Iamia*, *Mycobacterium*, *Nitrospira*, *Novosphingobium*, *Rhodobacter*,
470 *Trichoccus*, and *Steroidobacter*) out of the 20 core genera detected on the biofilm of 10
471 full-scale MBRs in China (Jo et al., 2016) were also detected on the biofilm (early and
472 mature) in the current study (Fig. 5) despite differences in geographic location (North
473 America vs. China), wastewater characteristics, plant operation, membrane type, flux, etc.
474 This further supports that a core biofilm community exists in geographically distributed
475 full-scale MBRs.

476 In the current study, the abundant core community in AS was also present as an
477 abundant core community in the biofilm, but their relative abundance varied between the
478 AS and biofilm samples. This is not surprising since the AS community is the main
479 source of inoculum for the biofilm. Nevertheless, this is not to say that the biofilm
480 community is a mere reflection of the AS community, and our results showed that the
481 assembly of biofilm communities from AS was not random, and was the result of species
482 sorting by local conditions (environmental and operational condtions, biotic interactions).
483 However, the specific local conditions driving the assembly of the abundant core
484 community in the full-scale MBRs were outside the scope of the current study. In MBRs,
485 several operating parameters have been shown to influence the microbial community

structure on membrane surfaces. For example, Huang et al. (2008) compared the biofouling communities of identical membranes operated under different fluxes (15 and 30 $L/m^2.h$) and solid retention times (SRTs, 8 and 30 d), and they concluded that the imposed membrane flux affected the community structure and composition of biofouling microorganisms. Miura et al. (2007) reported that the shear force induced by aeration over the membrane surface directly influenced the biofouling community composition where high shear forces selected for *Betaproteobacteria*. Also, studies have shown that the biofilm community structure may be affected by the physicochemical properties of polymeric membranes such as hydrophobicity, roughness and surface charge (Fontanos et al., 2010; Lee et al., 2014). The aforementioned studies were conducted in lab- or pilot-scale MBRs where conditions are different from full-scale MBRs. As the only published study in full-scale MBRs, Jo et al. (2016) showed that mixed liquor suspended solids, hydraulic retention time, food to microorganism ratio and specific aeration demand are important factors affecting the biofilm bacterial composition, whereas flux, temperature in the membrane tank, influent wastewater characteristics and membrane type are not important factors affecting biofilm community.

In addition to the presence of a dominant core membrane biofilm community in the five full-scale MBRs, there was a presence of a high number of unique OTUs (rare taxa) in each biofilm (early or mature) sample (Fig. S5), and these unique OTUs were mainly responsible for the difference in the community structure between the 5 MBRs (Fig. 1 and Fig. S2). This was evidenced in the NMDS analysis which showed that the bacterial communities in the five MBRs were more dispersed based on rare OTUs than the total and dominant OTUs (Fig. 2). The rare OTUs on the membrane surfaces may have an

509 important ecosystem function, but their importance could not be evaluated with the
510 current experimental design, and other criteria are needed to evaluate to what extent they
511 should be considered important. The unique OTUs in the different biofilm (early or
512 mature) samples could be due to differences in the environmental and operational
513 parameters between the 5 MBRs. A recent study reported that core AS communities in
514 full-scale AS systems are more shaped by deterministic factors than the rare members,
515 which are more shaped by neutral factors (Meerburg et al., 2016). Although it was out of
516 the scope of the current study, a more comprehensive and systematic study is needed in
517 the future to elucidate the factors shaping the core and rare biofilm communities in full-
518 scale MBRs.

519 It has been suggested that targeting the early colonizers in MBRs could help in
520 preventing biofouling (Choi et al., 2006; Zhang et al., 2006; Piasecka et al., 2012). This
521 is based on the premise that early colonizers determine the composition and nature of the
522 mature biofilm (Dang and Lovell, 2000; Davey et al., 2000; Kolenbrander et al., 2005;
523 Zhang et al., 2006; Costerton, 2007; Kjelleberg et al., 2007). For example, Lu et al.
524 (2016) observed that initial colonizers (*Nitrosomonas*, *Nitrospira*, *Nitrobacter*,
525 *Pseudomonas* and *Acinetobacter* species) profoundly affected the fouling behavior and
526 bacterial succession in a lab-scale nitrification MBR. Similarly, Bereschenko et al. (2008,
527 2010) identified *Sphingomonas* spp. as key organism responsible for the initiation of
528 membrane surface colonization that facilitates the attachment of other bacteria and
529 encourages the formation of mature biofilm in full-scale freshwater RO treatment facility.
530 Nevertheless, the large number of core early colonizers (63 genera) detected in the 5 full-
531 scale MBRs (Table S2) renders the application of innovative biological-based fouling

control strategies (e.g. quorum quenching, enzymatic disruption, energy uncoupling, and biofilm disruption by adding bacteriophage) challenging. For example, quorum quenching (QQ) has been suggested to be an effective method for mitigating biofouling in MBRs (Lee et al., 2016). However, of the 63 core genera detected in the early biofilm samples across the 5 full-scale MBRs, only 7 genera (*Acidovorax*, *Arcobacter*, *Bradyrhizobium*, *Flavobacterium*, *Nitrobacter*, *Nitrospira*, and *Rhodobacter*) have been classified in the literature as QS related bacteria (Jo et al., 2016). Jo et al. (2016) detected only 11.6% of QS bacterial genera in the biofilm of 10 full-scale MBRs in China. These results suggest that a single approach might not be effective in controlling biofouling in MBRs, and a combination of approaches might be more effective. For example, combinations of phage enzymes and disinfectants by adding the phage and then the disinfectant have been found to be more effective in biofilm eradication than adding either alone (Tait et al., 2002). Combination of QQ and chemically enhanced backwashing with chlorine injection was more effective in controlling fouling in MBR than adding either alone (Weerasekara et al., 2016). The current study is the first to characterize the early colonizers in full-scale MBRs. Due to technical reasons, only five MBR plants were selected in this study, and more comparative studies on full-scale MBRs are needed in the future to characterize the bacterial community structure of early colonizers with the aim of developing an effective global approach for mitigating biofouling in MBRs.

5. Conclusions

The main outcomes of this study can be summarized as follows:

- Alpha and beta diversity measures showed clear differences in the community structure between activated sludge and biofilm communities (early and mature) in the five full-scale MBRs. This difference was mainly attributed to the presence of large number of unique but rare operational taxonomic units (~13% of total reads in each MBR) in each sample.
- Despite the large fraction of sequence reads (~87% of total reads in each MBR) shared between activated sludge and biofilm communities (early and mature), simulated biofilm communities from random sampling of the respective activated sludge community revealed that stochastic immigration from the source community (i.e. activated sludge) was unlikely to shape the biofilm community assembly in MBRs.
- In addition to the presence of unique operational taxonomic units in each biofilm sample (early or mature), comparative analysis of operational taxonomic units and genera revealed the presence of a core biofilm community in the five full-scale MBRs. These core genera and operational taxonomic units represented the dominant taxa in the community.

Acknowledgments

This work was sponsored by King Abdullah University of Science and Technology (KAUST). We thank Prof. Regina Lamendella for NMDS analysis (Fig. 4).

References

577 Andrew, D.R., Fitak, R.R., Vega, A.M., Racolta, A., Martinson, V.G., Donstova, K.,
578 2012. Abiotic factors shape microbial diversity in Sonoran desert soils. Appl
579 Environ Microbiol. 78, 7527-7537.
580 Ayarza, J.M., Erijman, L., 2011. Balance of neutral and deterministic components in the
581 dynamics of activated sludge floc assembly. Microbial Ecol. 61, 486-495.
582 Bagchi, S., Tellez, B.G., Ananda Rao, H., Lamendella, R., and Saikaly, P.E., 2015.
583 Diversity and dynamics of dominant and rare bacterial taxa in replicate sequencing
584 batch reactors operated under different solids retention time and seeded with either
585 acclimated or non-acclimated sludge. Appl Microbiol Biotechnol. 99, 2361-2370.
586 Battin. T.J., Sloan, W.T., Kjelleberg, S., Daims, H., Head, I.M., Curtis, T.P., Eberl, L.,
587 2007. Microbial landscapes: new paths to biofilm research. Nat Rev Microbiol. 5,
588 76-81.
589 Bereschenko, L.A., Heilig, G.H.J., Nederlof, M.M., van Loosdrecht, M.C.M., Stams,
590 A.J.M., Euverink, G.J.W., 2008. Molecular characterization of the bacterial
591 communities in the different compartments of a full-scale reverse-osmosis water
592 purification plant. Appl Environ Microbiol. 74, 5297-5304.
593 Bereschenko, L.A., Stams, A.J.M., Euverink, G.J.W., van Loosdrecht, M.C.M., 2010.
594 Biofilm formation on reverse osmosis membranes is initiated and dominated by
595 Sphingomonas spp. Appl Environ Microbiol. 76, 2623-2632.
596 Besemer, K., Peter, H., Logue, J.B., Langenheder ,S., Lindström, E.S., Tranvik, L.J.,
597 Battin, T.J., 2012. Unraveling assembly of stream biofilm communities. ISME J. 6,
598 1459-1468.

599 Besemer, K., Singer, G., Limberger, R., Chlup, A.K., Hochedlinger, G., Hodl, I., Baranyi,
600 C., Battin, T.J., 2007. Biophysical controls on community succession in stream
601 biofilms. Appl Environ Microbiol. 73, 4966-4974.
602 Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis T.Z., Andersen, G.L., Knight, R.,
603 2010a. PyNAST: a flexible tool for aligning sequences to a template alignment.
604 Bioinformatics. 26, 266-267.
605 Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello,
606 E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley,
607 S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge,
608 B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A.,
609 Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010b. QIIME allows
610 analysis of high-throughput community sequencing data. Nat Meth. 7, 335-336.
611 Choi, H., Zhang, K., Dionysiou, D.D., Oerther, D.B., Sorial, G.A., 2006. Effect of
612 activated sludge properties and membrane operation conditions on fouling
613 characteristics in membrane bioreactors. Chemosphere. 63, 1699-1708.
614 Costerton, J.W., 2007. The biofilm primer. Springer, New York, NY.
615 Dang, H.Y., Lovell, C.R., 2000. Bacterial primary colonization and early succession on
616 surfaces in marine waters as determined by amplified rRNA gene restriction
617 analysis and sequence analysis of 16S rRNA gene. Appl Environ Microbiol. 6, 467-
618 475.
619 Davey, M.E., O'Toole, O.A., 2000. Microbial biofilms. From ecology to molecular
620 genetics. Microbiol Mol Bio Rev. 64, 847-867.

621 DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T.,
622 Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a Chimera-Checked 16S
623 rRNA Gene Database and Workbench Compatible with ARB. Appl Environ
624 Microbiol. 72, 5069-5072.
625 Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST.
626 Bioinformatics. 26, 2460-2461.
627 El-Chakhtoura, J., Prest, E.I., Saikaly, P.E., van Loosdrecht, M.C.M., Hammes, F.,
628 Vrouwenvelder, J.S., 2015. Dynamics of bacterial communities before and after
629 distribution in a full-scale drinking water network. Water Res. 74, 180-190.
630 Fontanos, P., Yamamoto, K., Nakajima, F., Fukushi, K., 2010. Identification and
631 quantification of the bacterial community on the surface of polymeric membranes at
632 various stages of biofouling using fluorescence in situ hybridization. Separ Sci
633 Technol. 45, 904–910.
634 Holyoak, M., Leibold, M.A., Holt, R.D., 2005. Metacommunities: Spatial Dynamics and
635 Ecological communities. University of Chicago Press: Chicago.
636 Huang, L.N., Wever, H.D., Diels, L., 2008. Diverse and distinct bacterial communities
637 induced biofilm fouling in membrane bioreactors operated under different
638 conditions. Environ Sci Technol. 42, 8360-8366.
639 Jackson, C.R., Churchill, P.F., Roden, E.E., 2011. Successional changes in bacterial
640 assemblage structure during epilithic biofilm development. Ecology. 82, 555-566.
641 Jinhua, P., Fukushi, K., Yamamoto, K., 2006. Bacterial community structure on
642 membrane surface and characteristics of strains isolated from membrane surface in
643 submerged membrane bioreactor. Separ Sci Technol. 4, 1527–1549.

644 Jo, S.J., Kwon, H., Jeong, S-Y., Lee, C-H., Kim, T.G., 2016. Comparison of microbial
645 communities of activated sludge and membrane biofilm in 10 full-scale membrane
646 bioreactors. Water Res. 101, 214-225.
647 Kjelleberg, S., Givskov, M., 2007. The biofilm mode of life, p. 5-21. Horizon Bioscience,
648 Wymonham, United Kingdom.
649 Kolenbrander, P.E., Egland, P.G., Diaz, P.I., Palmer Jr, R.J., 2005. Genome-genome
650 interactions: bacterial communities in initial dental plaque. Trends Microbiol. 13,
651 11-15.
652 Lee, S., Park, S-K., Kwon, H., Lee, S.H., Lee, K., Nahm, C.H., Jo, S.J., Oh, H-S., Park,
653 P.K., Choo, K-H., Lee, C-H., Yi, T., 2016. Crossing the border between laboratory
654 and field: bacterial quorum quenching for anti-biofouling strategy in an MBR.
655 Environ Sci Technol. 50, 1788-1795.
656 Lee, S.H., Hong, T.I., Kim, B., Hong, S., Park, H.D., 2014. Comparison of bacterial
657 communities of biofilms formed on different membrane surfaces. World J
658 Microbiol Biotechnol. 30, 777-782.
659 Leibold, M., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J., Hoopes, M.F., Holt,
660 R.D., Shurin, J.B., Law, R., Tilman, D., Loreau, M., Gonzalez, A., 2004. The
661 metacommunity concept: a framework for multi-scale community ecology. Ecol
662 Lett. 7, 601-613.
663 Lim, B., Ahn, K., Songprasert, P., Cho, J., Lee, S., 2004. Microbial community structure
664 of membrane fouling film in an intermittently and continuously aerated submerged
665 membrane bioreactor treating domestic wastewater. Water Sci Technol. 49, 255-
666 261.

667 Lim, S.Y., Kim, S., Yeon, K.M., Sang, B.I., Chun, J., Lee, C.H., 2012. Correlation
668 between microbial community structure and biofouling in a laboratory scale
669 membrane bioreactor with synthetic wastewater. Desalination. 287, 209-215.
670 Lu, H., Xue, Z., Saikaly P.E., Nunes, S., Bluver, T.R., Liu, W.T., 2016. Membrane
671 biofouling in a wastewater nitrification reactor: microbial succession from
672 autotrophic colonization to heterotrophic domination. Water Res. 88, 337-345.
673 Lu, L., Xing, D., Ren, N., 2012. Pyrosequencing reveals highly diverse microbial
674 communities in microbial electrolysis cells involved in enhanced H_2 production
675 from waste activated sludge. Water Res. 46, 2425-2434.
676 Malaeb, L., Le-Clech, P., Vrouwenvelder, J.S., Ayoub, G.M., Saikaly, P.E., 2013. Do
677 biological-based strategies hold promises to biofouling control in MBRs? Water
678 Res. 47, 5447-5463.
679 Meerburg, F.A., Vlaeminek, S.E., Roume, H., Seuntjens, D., Pieper, D.H., Jauregui, R.,
680 Vilchez-Vargas, R., Boon, N., 2016. High-rate activated sludge communities have a
681 distinctly different structure compared to low-rate sludge communities, and are less
682 sensitive towards environmental and operational variables. Water Res. 100, 137-145.
683 Miura, Y., Watanabe, Y., Okabe, S., 2007. Membrane biofouling in pilot-scale membrane
684 bioreactors (MBRs) treating municipal wastewater: Impact of biofilm formation.
685 Environ Sci Technol. 41, 632-638.
686 Musat, N., Halm, H., Winterholler, B., Hoppe, P., Peduzzi, S., Hillion, F., Horreard, F.,
687 Amann, R., Jorgensen, B.B., Kuypers, M.M.M., 2008. A single-cell view on the
688 ecophysiology of anaerobic phototrophic bacteria. Proc Natl Acad Sci USA. 105,
689 17861-17866.

690 Ofiţeru, I.D., Lunn, M., Curtis, T.P., Wells, G.F., Criddle, C.S., Francis, C.A., Sloan,
691 W.T., 2010. Combined niche and neutral effects in a microbial wastewater
692 treatment community. Proc Natl Acad Sci USA. 107(35), 15345-15350.
693 Pedrós-Alió, C., 2006. Marine microbial diversity: can it be determined? Trends
694 Microbiol. 14, 257-263.
695 Piasecka, A., Souffreau, C., Vandepitte, K., Vanysacker, L., Bilad, R., Bie, T., Hellemans,
696 B., Meester, L., Yan, X., Declerck, P., Vankelecom, I., 2012. Analysis of the
697 microbial community structure in a membrane bioreactor during initial stages of
698 filtration. Biofouling. 28, 225-238.
699 Ramette, A., 2007. Multivariate analyses in microbial ecology. FEMS Microbiol Ecol. 62,
700 142-160.
701 Rees, G.N., Baldwin, D.S., Watson, G.O., Perryman, S., Nielsen, D.L., 2004. Ordination
702 and significance testing of microbial community composition derived from terminal
703 restriction fragment length polymorphisms: application of multivariate statistics.
704 Antonie van Leeuwenhoek. 86, 339-347.
705 Saikaly, P. E., Oerther, D. B., (2011). Bacterial diversity in activated sludge promotes
706 functional resistance following toxic shock loading. Microb Ecol. 61, 557-567.
707 Saunders, A.M., Albertsen, M., Vollertsen, J., Nielsen, P.H., 2016. The activated sludge
708 ecosystem contains a core community of abundant organisms. ISME J. 10, 11-20.
709 Tait, K., Skillman, L., Sutherland, I., 2002. The efficacy of bacteriophage as a method of
710 biofilm eradication. Biofouling. 18, 305–311.

711 Tan, C.H., Koh, K.S., Xie, C., Tay, M., Zhou, Y., Williams, R., Ng, W.J., Rice, S.A,.
712 Kjelleberg, S., 2014. The role of quorum sensing signaling in EPS production and
713 the assembly of a sludge community into aerobic granules. ISME J. 8, 1186-1197.
714 Valentín-Vargas, A., Toro-Labrador, G., Massol-Deyá, A.A., 2012. Bacterial Community
715 Dynamics in Full-Scale Activated Sludge Bioreactors: Operational and Ecological
716 Factors Driving Community Assembly and Performance. PLoS ONE. 7(8):e42524.
717 Vuono, D.C., Munakata-Marr, J., Spear, J.R. Drewes, J.E., 2016. Disturbance opens
718 recruitment sites for bacterial colonization in activated sludge. Environ Microbiol.
719 18, 87-99.
720 Wang, X., Hu, M., Xia, Y., Wen, X., Ding, K., 2012. Pyrosequencing analysis of
721 bacterial diversity in 14 wastewater treatment systems in China. Appl Environ
722 Microbiol. 78, 7042-7047.
723 Wang, Z., Wu, Z., Tang, S., 2009. Extracellular polymeric substances (EPS) properties
724 and their effects on membrane fouling in a submerged membrane bioreactor. Water
725 Res. 43, 2504–2512.
726 Weerasekara, N.A., Choo, K.H., Lee, C.H., 2016. Biofouling control: bacterial quorum
727 quenching versus chlorination in membrane bioreactors. Water Res. 102, 293-301.
728 Wilhelm, L., Singer, G.A., Fasching, C., Battin, T.J., Besemer, K., 2013. Microbial
729 biodiversity in glacier-fed streams. ISME J. 7, 1651–1660.
730 Xia, S., Duan, L., Song, Y., Li, J., Piceno, Y.M., Andersen, G.L., Alvarez-Cohen, L.,
731 Moreno-Andrade, I., Huang, C-L., Hermanowicz, S.W., 2010. Bacterial community
732 structure in geographically distributed biological wastewater treatment reactors.
733 Environ Sci Technol. 44, 7391-7396.

734 Zhang, K., Choi, H., Dionysiou, D., Sorial, G., Oerther, D., 2006. Identifying pioneer
735 bacterial species responsible for biofouling membrane bioreactors. Environ
736 Microbiol. 8, 433-440.
737 Zhang, M., Jiang, S., Tanuwidjaja, D., Voutchkov, N., Hoek, E., Cai, B., 2011.
738 Composition and variability of biofouling organisms in seawater reverse osmosis
739 desalination plants. Appl Environ Microbiol. 77, 4390-4398.
740 Zhang, T., Shao, M., Ye, L., 2012. 454 Pyrosequencing reveals bacterial diversity of
741 activated sludge from 14 sewage treatment plants. ISME J. 6, 1137-1147.
742

Figure Captions

Fig. 1. Principal coordinate analysis (PCoA) of the 15 pooled samples based on unweighted UniFrac distance showing the relatedness of the bacterial community structure of AS and biofilms (Early and Mature). The numbers from 1 to 5 refer to the five different full-scale MBRs.

Fig. 2. Nonmetric multidimensional scaling (NMDS) plot of the 15 pooled samples based on unweighted UniFrac distance showing the total (green triangles), dominant (red squares) and rare taxa (blue triangles). The numbers from 1 to 5 correspond to the five different full-scale MBRs.

Fig. 3. Box plot showing unweighted UniFrac distance within and between Early, Mature and AS communities in all five full-scale MBRs. The red lines within the box represent the median while the plus signs are for outliers.

Fig. 4. Nonmetric multidimensional scaling (NMDS) analysis, visualizing the results of a random sampling procedure, to estimate the probability that the biofilm communities (Early and Mature) represented random samples of their respective AS communities. A total of 1,000 random subsamples of the AS communities were assembled for each MBR. Five examples A) MBR 1 B) MBR 2 C) MBR 3 D) MBR 4 and E) MBR 5 are shown, to illustrate the distribution of the randomly produced AS communities in relation to the biofilm community. White, red, blue, and green circles represent the random subsamples of the AS community, the AS community, the early biofilm community, and the mature biofilm community. NMDS was calculated based on the Horn Index. Plotted NMDS values were selected from ten independent random starting positions. The minimum stress values for each MBR ranged from 0.44 to 0.46.

Fig. 5. Heatmap distribution of the most abundant core genera (present at a relative abundance > 0.5% in at least one of the 5 samples in each category) in the 5 full-scale MBRs. Core genera indicate shared membership (i.e. genera) across all samples in the same category (i.e. AS, early or mature). The color intensity in each cell shows the percentage of genus in the corresponding sample, referring to the color key at the top left. The numbers from 1 to 5 correspond to the 5 full-scale MBRs.

Table 1
Alpha diversity measures for the 15 pooled samples.

		Alpha diversity measures				
MBR plant	Sample description	Number of observed OTUs	Richness estimate (Chao 1)	Shannon diversity index (H)	Phylogenetic diversity (PD)	Good's coverage (%)
MBR 1	AS	2,836	4,705±58	8.22±0.01	147.85±0.53	96.35
	Early	2,531	4,377±101	8.20±0.01	138.04±0.58	95.77
	Mature	2,104	3,615±86	7.86±0.01	124.07±0.6	93.80
MBR 2	AS	3,158	5,573±95	7.65±0.01	174.55±0.55	96.12
	Early	1,002	2,231±95	7.88±0.01	80.50±0.57	84.20
	Mature	939	1,765±39	6.61±0.01	74.59±0.32	93.24
MBR 3	AS	4,927	6,024±187	6.41±0.01	173.65±1.88	97.29
	Early	3,852	8,113±3	7.65±0.01	226.01±0.036	93.92
	Mature	3,847	6,957±87	7.91±0.01	209.94±0.85	94.95
MBR 4	AS	5,820	7,814±142	7.99±0.01	212.08±2.43	96.51
	Early	6,943	7,216±312	8.03±0.01	204.87±2.64	97.48
	Mature	3,477	5,702±265	7.83±0.01	171.70±1.37	96.65
MBR 5	AS	2,563	4,874±124	7.79±0.01	156.78±0.64	94.33
	Early	4,650	5,212±148	7.99±0.01	170.28±1.69	97.91
	Mature	2,497	4,414±60	7.77±0.01	146.81±0.62	95.61

Table 2
Percentages of shared OTUs and their corresponding sequences between the different samples (AS, Early or Mature) from the 5 MBRs. For each MBR, AS: Early, AS: Mature and Early: Mature correspond to the shared OTUs and sequences within two types of samples.

	OTUs			Sequences		
Sample	Total	Shared	Shared (%)	Total	Shared	Shared (%)
AS (Combined)[a]	14,090	228	1.62	286,468	100,744	35.17
Early (Combined)[a]	14,323	138	0.96	310,757	86,829	27.94
Mature (Combined)[a]	9,518	114	1.20	146,745	37,864	25.80
MBR 1						
AS: Early	3,794	1,573	41.46	65,143	56,326	86.47
AS: Mature	4,016	1,121	27.91	53,442	45,826	85.75
Early: Mature	3,536	1,099	31.08	45,677	39,572	86.63
MBR 2						
AS: Early	3,443	717	20.82	43,523	3,1672	72.77
AS: Mature	3,497	600	17.15	47,243	3,1875	67.47
Early: Mature	1,577	364	23.08	11,734	6,103	52.01
MBR 3						
AS: Early	6,615	2,164	32.71	131,490	121,604	92.48
AS: Mature	6,726	2,048	30.44	134,487	123,135	91.56
Early: Mature	5,902	1,797	30.44	76,373	67,775	88.74
MBR 4						
AS: Early	9,338	3,425	36.67	231,592	217,864	94.07
AS: Mature	7,008	2,289	32.66	145,697	138,392	94.99
Early: Mature	7,841	2,579	32.89	194,763	184,864	94.92
MBR 5						
AS: Early	5,418	1,795	33.13	125,477	115,733	92.23
AS: Mature	3,684	1,376	37.35	52,344	49,042	93.69
Early: Mature	5,273	1,874	35.54	128,955	119,139	92.39

[a]Combined samples correspond to the five AS, five early biofilm, or five mature biofilm samples collected from the 5 MBRs.

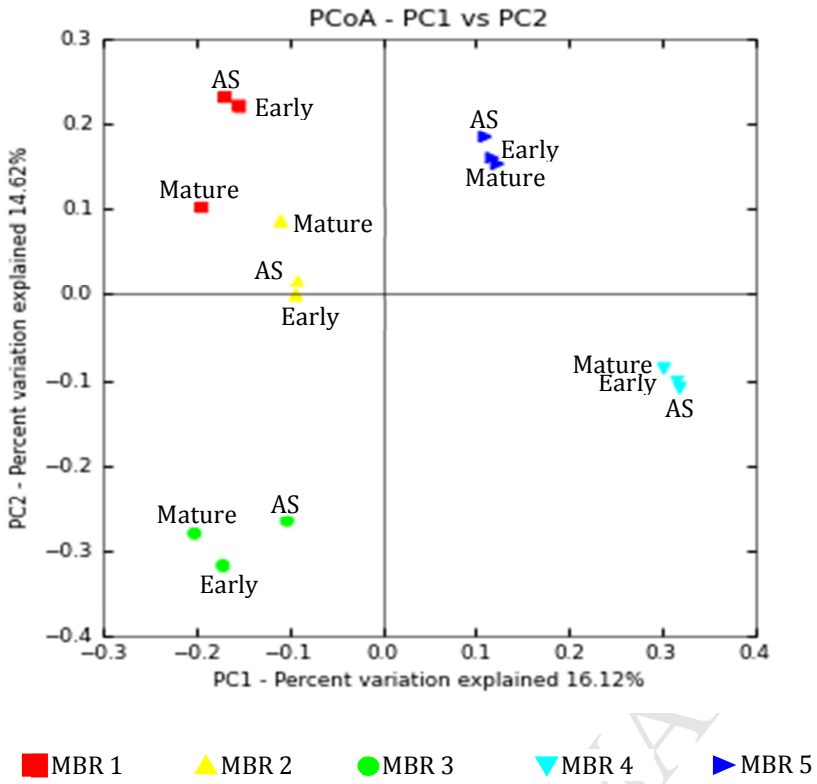


Fig. 1. Principal coordinate analysis (PCoA) of the 15 pooled samples based on unweighted UniFrac distance showing the relatedness of the bacterial community structure of AS and biofilms (Early and Mature). The numbers from 1 to 5 refer to the five different full-scale MBRs.

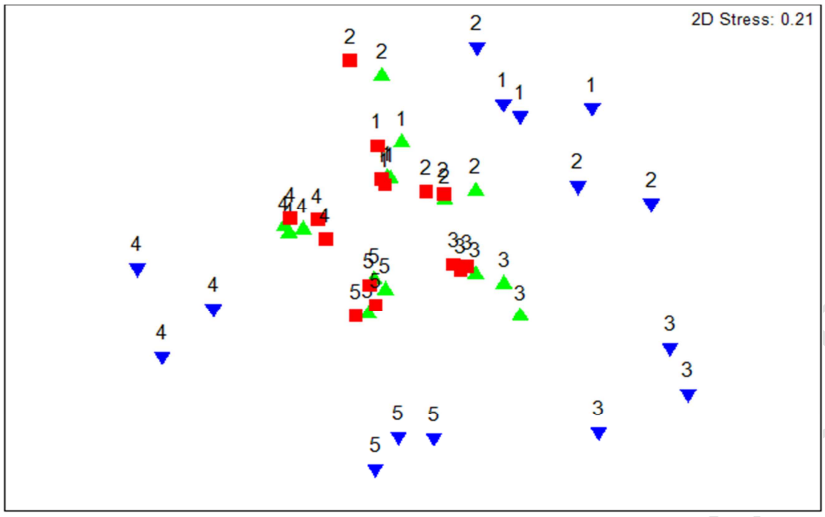


Fig. 2. Nonmetric multidimensional scaling (NMDS) plot of the 15 pooled samples based on unweighted UniFrac distance showing the total (green triangles), dominant (red squares) and rare taxa (blue triangles). The numbers from 1 to 5 correspond to the five different full-scale MBRs.

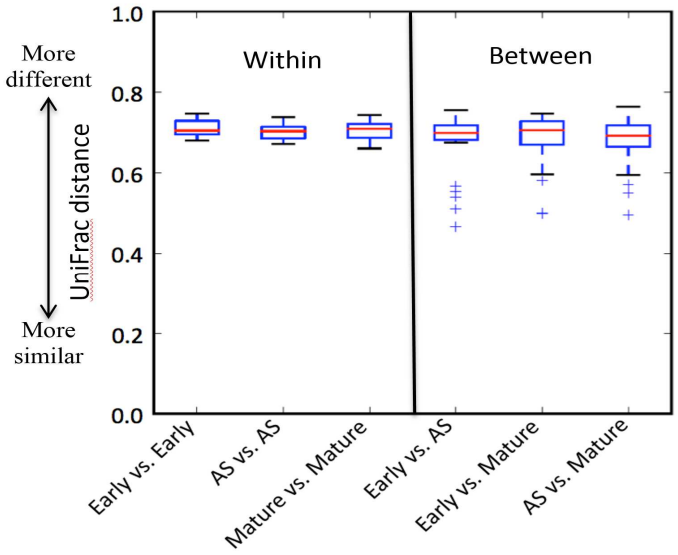


Fig. 3. Boxplot showing unweighted UniFrac distance within and between Early, Mature and AS communities in all five full-scale MBRs. The red lines within the box represent the median while the plus signs are for outliers.

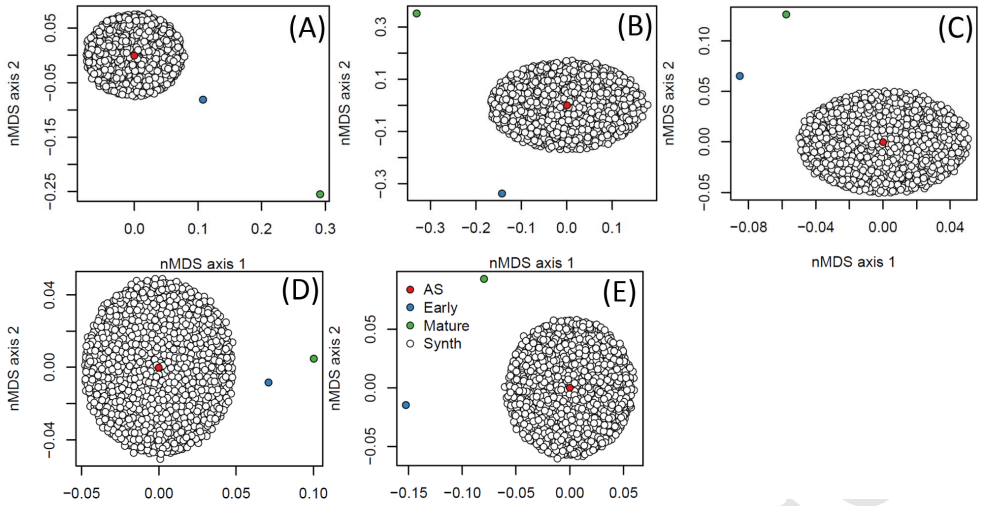


Fig. 4. Nonmetric multidimensional scaling (NMDS) analysis, visualizing the results of a random sampling procedure, to estimate the probability that the biofilm communities (Early and Mature) represented random samples of their respective AS communities. A total of 1,000 random subsamples of the AS communities were assembled for each MBR. Five examples A) MBR 1 B) MBR 2 C) MBR 3 D) MBR 4 and E) MBR 5 are shown, to illustrate the distribution of the randomly produced AS communities in relation to the biofilm community. White, red, blue, and green circles represent the random subsamples of the AS community, the AS community, the early biofilm community, and the mature biofilm community. NMDS was calculated based on the Horn Index. Plotted NMDS values were selected from ten independent random starting positions. The minimum stress values for each MBR ranged from 0.44 to 0.46.

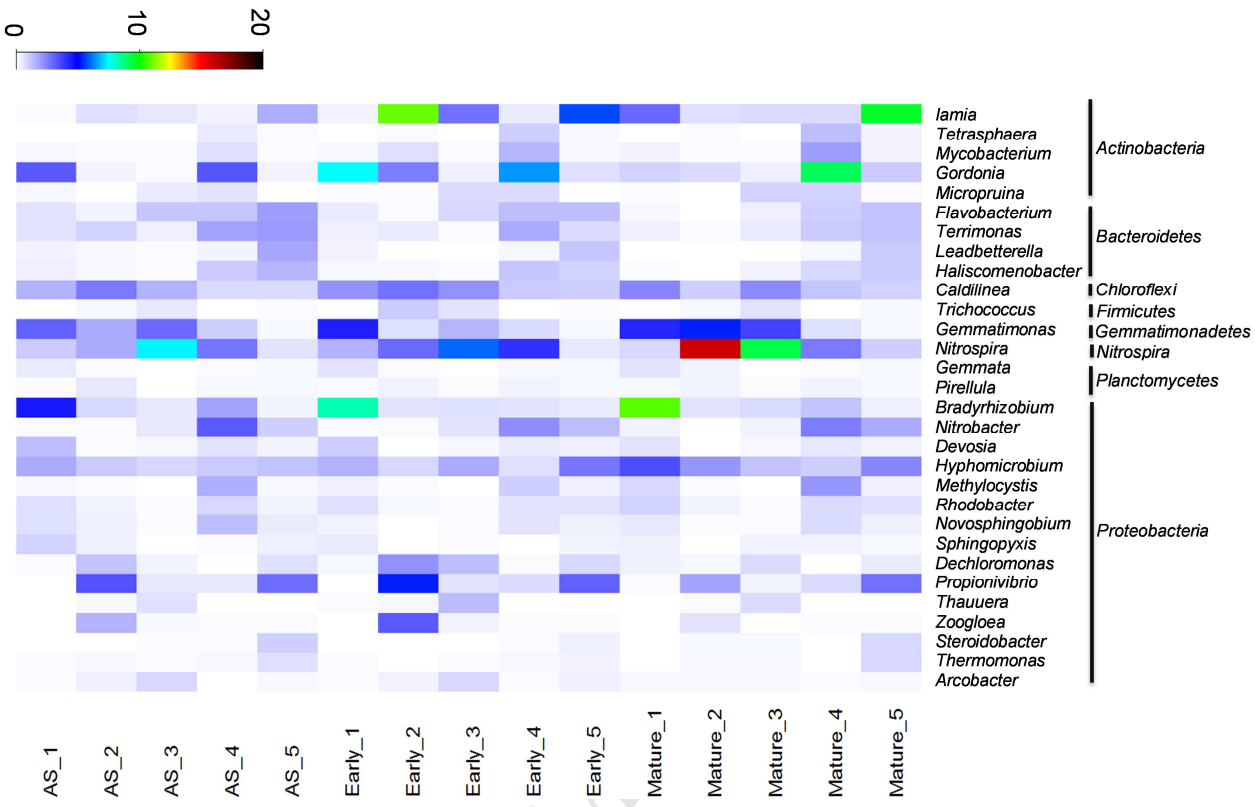


Fig. 5. Heatmap distribution of the most abundant core genera (present at a relative abundance > 0.5% in at least one of the 5 samples in each category) in the 5 full-scale MBRs. Core genera indicate shared membership (i.e. genera) across all samples in the same category (i.e. AS, early or mature). The color intensity in each cell shows the percentage of genus in the corresponding sample, referring to the color key at the top left. The numbers from 1 to 5 correspond to the 5 full-scale MBRs.

Highlights

- Membrane biofilm (early and mature) community analysis in five full-scale MBRs
- Clear difference in bacterial community diversity between AS and biofilm communities
- This difference was attributed to the presence of large number of unique but rare taxa in each sample
- Membrane biofilm (early and mature) communities are not randomly assembled from AS community
- A core membrane biofilm community exists in full-scale MBRs