Genome sequencing as a new window into the microbial community of membrane bioreactors – A critical review

Accepted version

Science of the Total Environment

October 2019

Luong N. Nguyen[a*], Audrey S. Commault[b], Tim Kahlke[b], Peter J. Ralph[b], Galilee U. Semblante[c], Md Abu Hasan Johir[a] and Long D. Nghiem [a, d]

[a] Centre for Technology in Water and Wastewater, School of Civil and Environmental Engineering, University of Technology Sydney, NSW 2007, Australia

[b] Climate Change Cluster (C3), University of Technology Sydney, NSW 2007, Australia

[c] Technical Services, Western Sydney University, Kingswood, NSW 2747, Australia

[d] NTT Institute of Hi-Technology, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam

[*]Corresponding author:

Luong N. Nguyen: Centre for Technology in Water and Wastewater, School of Civil and Environmental Engineering, University of Technology Sydney, NSW 2007, Australia

Phone: (+61) 468863865 E-mail: luongngoc.nguyen@uts.edu.au

Abstract

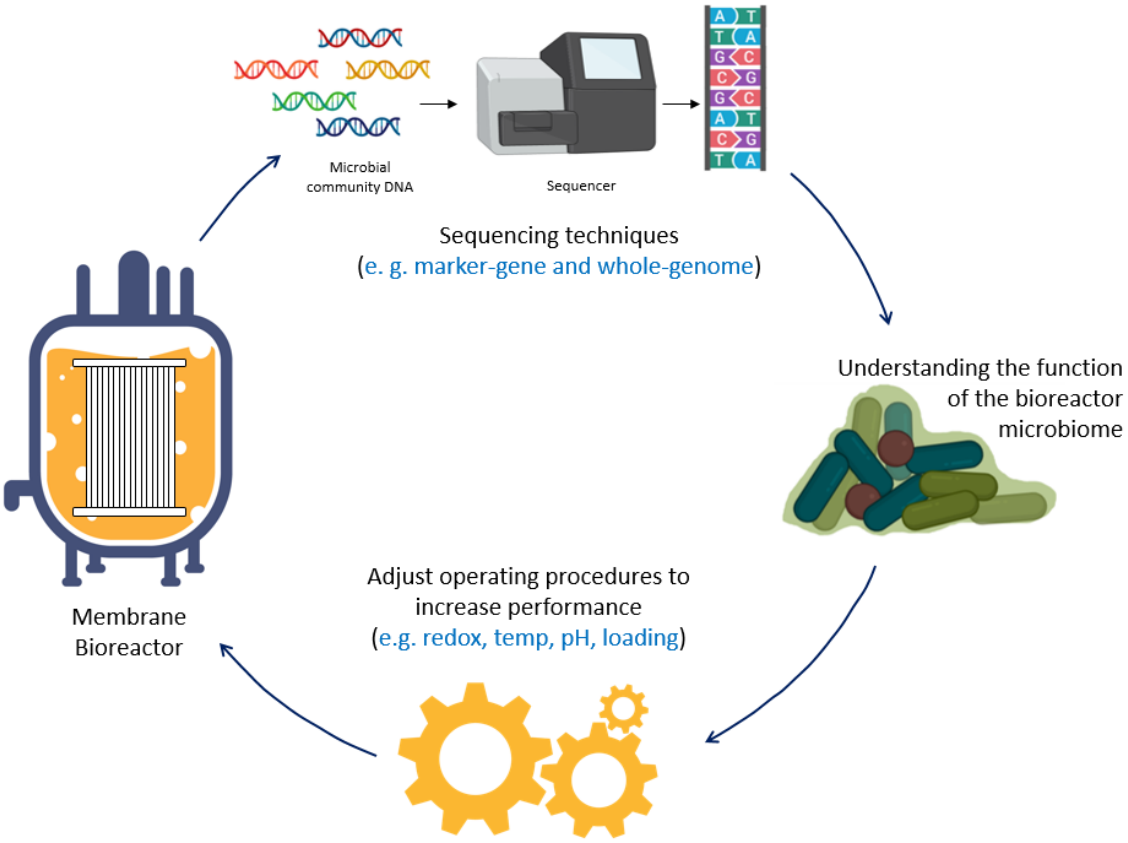
Recent developed sequencing techniques have resulted in a new and unprecedented way to study biological wastewater treatment, in which most organisms are uncultivable. This review provides (i) an insight on state-of-the-art sequencing techniques and their limitations; (ii) a critical assessment of the microbial community in biological reactor and biofouling layer in a membrane bioreactor (MBR). The data from high-throughput sequencing has been used to infer microbial growth conditions and metabolisms of microorganisms present in MBRs at the time of sampling. These data shed new insight to two fundamental questions about a microbial community in the MBR process namely the microbial composition (who are they?) and the functions of each specific microbial assemblage (what are their function?). The results to date also highlight the complexity of the microbial community growing on MBRs. Environmental conditions are dynamic and diverse, and can influence the diversity and structural dynamics of any given microbial community for wastewater treatment. The benefits of understanding the structure of microbial communities on three major aspects of the MBR process (i.e. nutrient removal, biofouling control, and micropollutant removal) were symmetrically delineated. This review also indicates that the deployment of microbial community analysis for a practical engineering context, in terms of process design and system optimization, can be further realized.

Keywords: Membrane bioreactor; Microbial community; Microbial ecology; Maker-gene sequencing; Whole-genome sequencing; wastewater treatment.

Highlight

- Molecular techniques can reveal microbial community composition and functionalities
- Insight to MBR performance is achieved through microbial community analysis
- Considerations (sample & data analysis) in microbial community studies are reviewed
- Possible directions (e.g. full-scale) to enhance engineering outcomes are suggested

52 Graphical abstract



53

1. Introduction

Since the inception of modern sanitation, large scale municipal wastewater treatment has relied almost exclusively on the aerobic activated sludge (AS) process (Sheik et al., 2014). In a simplistic representation of the AS process, microorganisms assimilate and convert dissolved organic matters and nutrients in wastewater to insoluble body cells (i.e. biomass or activated sludge) and simple gases such as CO_2 and N_2. The activated sludge can be removed from the treated wastewater by gravity in a conventional clarifier. In practice, several variations of the AS process with a combination of different biological treatment conditions including anaerobic, anoxic and aerobic can be applied to promote the growth and functions of different microbial communities to achieve overall performance (i.e. division of functionality to enhance efficiency).

A recent alternative to AS treatment is membrane bioreactors (MBR) which utilizes a membrane process for biomass separation instead the conventional clarifier (Xiao et al., 2019). Indeed, the MBR process is a hybrid of a biological and physical liquid-solid separation processes (Xia et al., 2010; Nguyen et al., 2012a; Wolff et al., 2018). Compared to the AS process, MBR has a much lower physical footprint and can produce higher and more reliable effluent quality (Xiao et al., 2019; Nguyen et al., 2012b). With the decrease in membrane cost, new and more stringent regulations on effluent quality, demand for water recycling, many MBR plants have recently been commissioned around the world especially for large scale water reuse applications (Xiao et al., 2019). In principle, MBR performance is governed by both physical and biological processes (Xia et al., 2010; Nguyen et al., 2012a). While the physical process can be readily controlled by regulating operational parameters and membrane selection, the performance of biological process relies on the microbial community in the bioreactor (Xiao et al., 2019). Microbial community is also subjected to changes in operation conditions (e. g. dissolved oxygen, sludge retention time and hydraulic retention time,

79 temperature). Thus, understanding the structure, functions and dynamics of microbial
80 communities involved in the biological process has been the objective of many studies recently
81 (Wolff et al., 2018; Wen et al., 2018; Inaba et al., 2018; Zhu et al., 2017).

82 Analysis of the microbial community structure, functions and dynamics in the biological
83 process has been possible since the emergence of high-throughput sequencing techniques.
84 Bypassing the reliance on cultivable microbes, high-throughput sequencing techniques provide
85 details of microbial assemblages in any given activated sludge samples. Sequencing techniques
86 can be used to target specific research questions. Marker-gene based approach provides
87 microbial profile (i.e. who are they?) while whole-genome approach create a functional profile
88 of a microbial community (i.e. what do they do?) based on the functional genes in the genomes
89 of the different microbes. Both approaches have recently been used to investigate the microbial
90 communities in biological treatment process. A review of the literature on the identity and
91 potential metabolic capabilities of microorganisms is imperative for a better design, control
92 and understanding of bioreactors. The broad range of microorganisms present in biological
93 reactor across different operational conditions has been revealed; however, the process design
94 and control have yet to be revised accordingly to this new knowledge.

95 This paper reviews the state-of-the-art knowledge on the microbial community of the
96 biological reactor obtained from modern molecular methods. Molecular methods including
97 both marker-gene and whole-genome sequencing are critically discussed with a focus on their
98 outcomes, considerations and limitations. Recent achievement of microbial community profile
99 and their functions are provided. Potential new strategies resulting from a better understanding
100 of the microbial community to improve MBR performance (i.e. three dominant performances:
101 nutrient removal, biofouling and micropollutant removal) is also discussed. This critical review
102 expects to guide future studies of the MBR microbiome.

103 **2. Contemporary Molecular Methods for Microbial Community Analysis**

2.1 Microbial community analysis

Two key aims of molecular biology analysis are to determine components of microbial assemblages (who are they?) and identify their functions (what do they do?) as outlined in Fig 1. These two aims are achieved through different molecular methodological approaches, namely marker-gene, whole-genome sequencing and other –omics methods (i.e., metatranscriptomics, metaproteomics). These approaches begin with microbial sampling then follow by either DNA, RNA or proteins extraction. These two initial steps are crucial as their results will largely influence subsequent steps. The extracted molecules can be subjected to different steps to target each of the key aims of molecular biology analysis (i.e. profiling the microbes or identifying their functions) (Fig 1). In the context of wastewater treatment, monitoring the biological activity of the microbial community is essential to ensure a high quality of treated water. Marker-gene sequencing allows for the identification of the microbial community present in the MBR, metagenomics is used to describe the putative functions of the microbial community, while metatranscriptomics and metaproteomics can highlight the active metabolic pathways and/or the active microorganisms at time of sampling (Fig. 1). Metatranscriptomics and metaproteomics are powerful methods to infer if the microbes are performing as expected. Metaproteomic studies of MBR communities were performed to understand fouling (Zhou et al., 2015) or the effect of substrate stress (Salerno et al., 2019). Metatranscriptomics have recently been used in combination with metagenomics to analysis MBR microbial community dynamics and interactions (Yang et al., 2019). However, these methods are still in their infancy and only a few studies have applied them in the context of MBR, making it hard to source sufficient literature for a thorough review. In addition, metatranscriptomics and metaproteomics are currently limited by a lack of quantitative approaches and incomplete coverage of reference databases impairing transcripts and proteins identification, meaning that metagenomics often need to be conducted alongside. We foresee

that these limitations will be overcome in the coming decades as tailored search databases are being constructed. For the above-mentioned reasons, this review focuses mostly on marker-gene and whole-genome sequencing for the study of microbial diversity and putative functions in MBR.

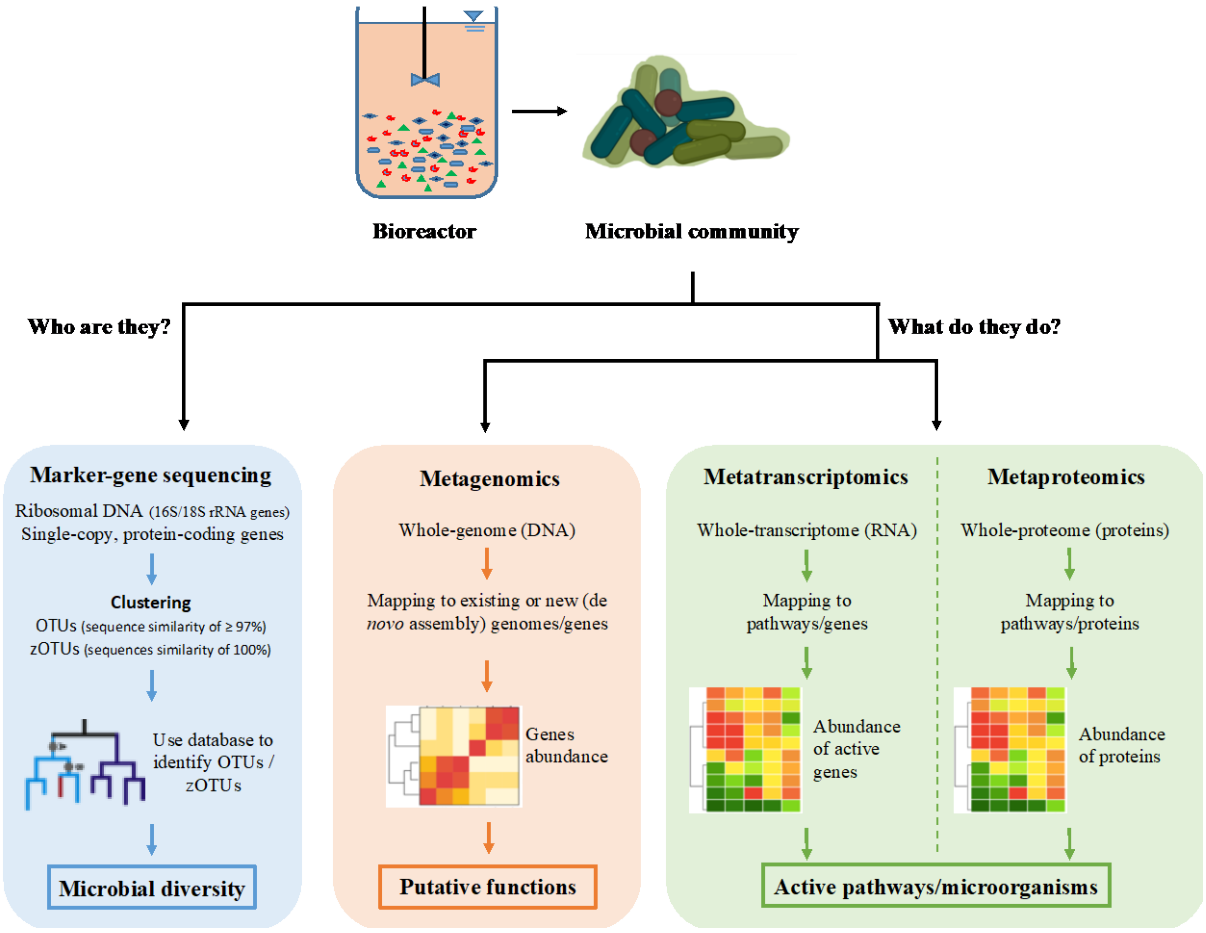


Figure 1. Different sequencing approaches to reveal bioreactor microbial community structure and functions.

2.1.1 Marker-gene based approach: Who are they?

Marker-gene based approach, also known as targeted and amplicon sequencing, utilizes universal marker genes to determine components of microbial assemblages (Fig 1). The most common marker genes are ribosomal RNA genes such as 16S for bacteria and archaea and 18S genes for eukaryotes (Brown et al., 2018). Depending on the sample and organisms of interest,

custom marker genes encoding proteins can also be used such as cytochrome b (cob) for dinoflagellates (Smith et al., 2017) and heat shock proteins such as hsp60 for specific bacterial taxa (Sakamoto & Ohkuma, 2010). In the context of wastewater treatment, markers targeting specific functional genes such as the ammonia monooxygenase (*amoA*) gene or the nitrogenase reductase (*nifH*) gene have been used to study the genetic diversity of ammonia-oxidizing bacteria (AOB) and species of the *Frankia* genus, respectively (Wang et al., 2014; Rodriguez et al., 2016).

Given the length limitations of short-read sequencing technologies, e.g., Illumina sequencing, often only a particular region of the chosen marker gene is targeted. The choice of marker gene and the gene region used are of crucial importance and can have a significant impact on the study outcome (Větrovský & Baldrian, 2013). Generally, marker genes have to be universal single-copy genes as copy-number variations inflate the apparent abundance of some taxa in comparison to others (Větrovský & Baldrian, 2013).

A key-concept of microbial community analysis is the *operational taxonomic unit* (OTU). An OTU represents a cluster of sequences that show certain sequence similarity and that are assumed to originate from the same taxonomic group of organisms (Fig 1). Konstantinidis and Tiedje (2005) showed intra-species 16S gene sequence similarity to be $\geq 97\%$ for most bacterial taxa. Based on these results a similarity cut-off of 97% was adopted by the community to cluster 16S sequences into OTUs (Konstantinidis & Tiedje, 2005). Although similarity clustering can minimize potential errors from sequencing errors and intra-species variation, recent studies have highlighted the importance of fine-scale community structure that can be obscured when merging sequences at a threshold of $97\% \leq$ similarity $< 100\%$ (Edgar, 2018; Callahan et al., 2017). Therefore, zero-radius/zero-difference OTUs (zOTUs), also called amplicon or exact sequence variants, are increasingly used as the lowest taxonomic rank in

microbial community analyses. An zOTU is defined as clusters of sequences with 100% similarity, i.e., each unique marker-gene sequence in a data set represents a separate zOTU.

Amplicon sequencing is a universal process (Fig. 2). For any marker gene and taxonomic ranking level, polymerase-chain-reaction (PCR) primers are used to bind to a specific region of the marker gene for PCR amplification. The amplified marker genes are then sequenced using either short-read sequencing technologies such as Illumina's *sequencing by synthesis* method or recently available long-read methods such as nanopore sequencing. Illumina short-reads are the de-facto standard in microbial community analysis due to the higher accuracy of the sequencing reads and more advanced analysis tools. Long-read techniques have also been increasingly used for community analysis. With future development and the potential advantages of such as the ability to sequence the complete marker gene in contrast to only a particular region, further application of long-read techniques for amplicon sequencing can be expected.

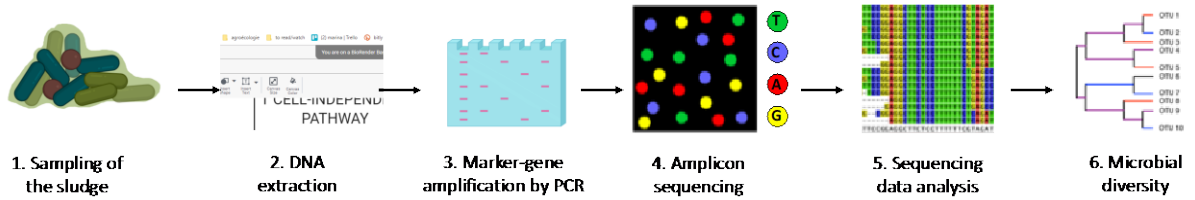


Figure 2. A basic flowchart of marker-gene sequencing approach

A large number of tools are available for amplicon sequence analysis. Two of the most common bioinformatics pipelines are Qiime (Caporaso et al., 2010) and Mothur (Schloss et al., 2009). Both Qiime and Mothur provide scripts for quality control and trimming of sequencing reads, OTU/zOTU picking, as well as methods for taxonomic classification in a user-friendly way. Taxonomic classification is still a challenge for many microbial community studies, especially when using custom marker genes or eukaryotic microbes such as protists and marine fungi (Nilsson et al., 2019; Andreakis et al., 2015) due to the lack of well-curated databases of

189 known sequences and taxonomies. There are several approaches for taxonomic classification
190 including machine learning algorithms such as Naïve Bayes and BLAST-based methods
191 (Bokulich et al., 2018). However, appropriate approach is often based on marker genes, the
192 sampled environments as well as user preferences.

193 The marker-gene based approach is mainly used to determine the microbial community
194 composition of a sample. However, in some cases, this approach has been used to infer
195 metabolic capabilities from a community profile, e.g., for bacterial assemblages (Aßhauer et
196 al., 2015; Langille et al., 2013), based on completely sequenced genomes of closely related
197 bacteria, on the assumption that closely related species share similar functional profile. Whole
198 genome sequencing is a better approach when determining the function of specific bacteria
199 within a community.

200 2.1.2 Whole Genome Sequencing: What are their functions?

201 The principal objectives of metagenomic approach in microbial ecology are to (i) determine
202 the metabolic and functional potential of the community of interest and (ii) to connect genes
203 and their metabolic functions with specific microbial taxa (Fig 1). In contrast to marker-gene
204 approaches, metagenome sequencing provides a snapshot of the complete genomic information
205 of a sample at a particular sampling time and not just a single gene.

206 The bioinformatic pipelines for metagenome data analysis are less standardized compare to
207 amplicom data. Commonly, metagenome analysis workflows including i) steps for quality-
208 control of raw reads; ii) assembly of reads into longer continuous DNA fragments (contigs)
209 using metagenome assemblers (Nurk et al., 2017; Liu et al., 2015); iii) a binning step to cluster
210 contigs that originate from the same organism into contigs bins and metagenome assembled
211 genomes (MAGs) (Lu et al., 2016; Kang et al., 2015); iv) subsequent gene prediction (Stanke
212 & Morgenstern, 2005; Hyatt et al., 2010) and taxonomic classification (Kahlke & Ralph, 2019;
213 Darling et al., 2014; Ounit et al., 2015). There are several pipelines to automate these steps

(Uritskiy et al., 2018; Tamames & Puente-Sánchez, 2019) for integrating state-of-the-art tools into single metagenome analysis pipelines. However, many bioinformaticians choose a combination of custom tools and software based on study design, research question and personal preferences. Downstream statistical and differential analysis can also be performed using specific software such as the R package DESeq and the metagenome analysis tool STAMP (Anders & Huber, 2010; Parks et al., 2014).

The strength of metagenome experiments lies in its ability to link functional profiles of samples with specific taxonomic groups. Some bioinformatic pipelines, however, estimate abundance-based community profiles similar to those provided by marker-gene based approaches (Ounit et al., 2015; Wood & Salzberg, 2014). Although these tools do not use the wealth of information in a metagenome sample, they provide valuable insight into the sample's microbial community and can achieve results comparable to those of amplicon sequencing studies.

2.2 Key considerations and limitations

In the context of wastewater treatment, whole genome sequencing has been carried out in an attempt to elucidate which species are involved in organic matter, ammonium, nitrogen and phosphorus removal (Siezen & Galardini, 2008; Ma et al., 2016; Nguyen et al., 2019a), as well as assessing the pathogenic potential of multi-drug resistant bacteria (Mahfouz et al., 2018). Amplicon sequencing, on the other hand, is commonly used to describe microbial community structure and monitor the abundance of key organisms. However, these methods have limitations and precautions have to be taken at every step (sampling, DNA extraction and data analysis) of the process to minimize mistakes when dealing with AS and wastewater samples. Some limitations encountered at each of these three steps are discussed below.

2.2.1 Sampling

Spatial scaling of microbial biodiversity needs to be taken into consideration when sampling wastewater for microbial community analysis as environmental heterogeneity can result in spatial patterns of microbial diversity (Green & Bohannan, 2006). In the context of wastewater treatment, microbes are subjected to diverse and transient environmental conditions (e.g., variation in dissolved oxygen and nutrient content). Thus, a sample taken at a specific time can only represent a snap shot of the microbial community (Hu et al., 2012). The need for sample replications is crucial, as the sampled environment is heterogeneous. For instance, the simple absence of mixing in a bioreactor can result in stratification of the microbial distribution (Nguyen et al., 2019b). To assess species diversity and make sure the species richness of a sampling site was adequately sampled, ecologists have developed tools, such as diversity indexes and rarefaction curves (de Vargas et al., 2015; Gotelli & Colwell, 2001). These methods should be used when sampling heterogeneous sites such as wastewater treatment bioreactors to ensure meaningful comparison of datasets and proper estimation of low-abundance species. When studying the impact of disturbances and unsteady environmental conditions on microbial diversity, time series sampling is important as it provides information on the dynamic of the community. In a recent study, Perez et al. (2019) performed 16S rRNA amplicon sequencing and metagenomics over a period of 3 years in a full-scale municipal activated sludge wastewater treatment plant (WWTP) to monitor the changes in bacterial populations overtime and understand the adaptive response of microbiomes to disturbances due to short-term plant shutdowns. These types of results contribute to the development of predictive models and help guide engineering and WWTP management practices.

2.2.2 DNA extraction

The DNA extraction step is also a source of potential errors. Indeed, it is well established that DNA extraction kits commonly used to extract genomic DNA from wastewater samples may be contaminated with bacterial DNA. Up to 181 contaminating microbes genera have been identified in common DNA extraction kits (Glassing et al., 2016). Many of these contaminating microbes are commonly found in the human gut and the environment, thus, it may not be possible to distinguish them from those in the sample. These contaminating microbes may affect the interpretation of low-abundant bacteria in the samples. Glassing et al. (2016) recommend careful scrutinisation of any unusual and unexpected results to distinguish between new findings and possible contamination. Extraction blanks (as no template) are recommended to be processed together with the samples and alongside proper controls to limit misinterpretations (Glassing et al., 2016) .

Sequencing experiments require high-quality DNA samples (or RNA in the case of transcriptomics) with very low to no nucleic acids degradation. The quality of DNA is often controlled using spectrophotometry (Nanodrop), fluorimetry (PicoGreen or Qubit) and gel electrophoretic methods (Bioanalyser). When using spectrophotometry, the A260:A280 and A260:A230 ratios should be higher than 1.8. Since DNA absorbs at 260 nm, ratios lower than 1.8 indicate contamination of the DNA sample with proteins (absorb at 280 nm) or chemicals (e.g., EDTA, phenol, carbohydrates absorb near 230 nm) used in the extraction procedure. DNA Integrity Number (DIN) or Genomic DNA Quality Score (GQS) can be calculated from the size distribution of the DNA sample using electrophoretic methods, they provide a robust method for DNA quality, with most experiments using samples with DIN>7.

2.2.3 Batch effects, PCR artefacts and sequencing errors

Another group of errors that can have significant effects on the outcomes of sequencing studies is those inherent to the technology used such as batch effects, PCR artefacts and sequencing

errors. Batch effects describe a broad group of factors that add variance to sequencing data that is not based on a true biological signal, such as DNA extracted on different days, different technicians performing the sampling or DNA extraction, different batches of chemicals as well as sequencing control and treatment samples on different days, machines or flow cells (Leek et al., 2010). This is especially problematic when comparing sequencing data from different studies, in time serious or longer temporal studies (Goh et al., 2017). Despite being known for more than a decade correcting for batch effects in microbial sequencing data is challenging and hard to distinguish from true biological signals. Therefore, care should be taken during experimental design to limit batch effects, e.g., via randomization of sample collection, DNA extraction and sequencing (Leek et al., 2010; Yang et al., 2008). Downstream data analysis approaches such as principal component analysis (PCA) and permutational multivariate analysis of variance (PERMANOVA) can help to identify batch effects (Holman et al., 2017). Combined with common-practice analyses such as Principal Component Analysis batch effects can be picked up. Additionally, recently developed bioinformatic approaches such as percentile-normalization can limit batch effects for analysis of pooled studies (Helbling et al., 2015).

Another systematic error of major concern is so-called *PCR chimeras* or *chimeric sequences*. These sequences originate from incomplete amplification of fragments during a PCR cycle that act as amplification primers in subsequent PCR cycles resulting in artificial sequences merged from more than one true biological parent sequence. It has been shown that PCR chimerase are generally very common, ranging anywhere from 8-80% of reads in a sample (Wang & Wang, 1996; Qiu et al., 2001). Although it has been shown that optimizing PCR conditions can reduce the formation of PCR chimeras (Smyth et al., 2010; Omelina et al., 2019) bioinformatic identification of chimeric sequences is crucial and implemented in all common data analysis pipelines. Similarly, sequencing errors, i.e., wrong bases introduced during the

sequencing process, can artificially inflate the number of unique sequences in a sample. This is especially problematic for zOTU approaches where sequences with as little as one nucleotide difference are assumed to originate from different organisms. One way of increasing the accuracy of sequences is to correct errors the forward reads of paired-end data with the overlapping part of its reverse mate. This is especially useful for short marker genes where the overlap of the two mates is large. Many common read-joining tools such as BBmerge (Bushnell et al., 2017) and FLASH (Magoč & Salzberg, 2011) already implement these strategies. However, current maximum read lengths of Illumina technology are <=350bp which is much shorter than most common marker genes. Additionally, more recent so called *denoising* algorithms such as UNOISE (Edgar & Flyvbjerg, 2015) achieve higher accuracy without the need of large read-pair overlaps.

2.2.4 Data analysis

The initial step in any genomic project is the quality control of the raw data, i.e., check for read length, quality, and removal of low quality bases and reads. Initial visualization of the raw sequencing data can be performed with tools like FastQC. Similarly, amplicon sequence analysis frameworks such as Qiime (Caporaso et al., 2010) and Mothur (Schloss et al., 2009) provide visualization and data statistics for raw read data as well as trimming and filtering functionality. For genomic and metagenome data a variety of tools for trimming and filtering is available such as Trimmomatic, PRINSEQ or the Fastx-toolkit to name just a few. Subsequent to the initial filtering error correction, batch effect adjustments and identification of chimeric sequences should be performed where applicable (see section 2.2.3 for details).

When performing amplicon sequencing, the length of the reads influences the taxonomic resolution, with longer reads allow distinguishing between related strains that will otherwise share the same amplified region. However, most of the sequencing platforms used today require short reads, typically 100 - 500 nucleotide or 16 - 33% of the total length of the marker-gene

(Callahan et al., 2019), thus limiting the resolution of taxonomic profiles. In recent years, new technologies that generate long sequencing reads (tens of thousands of nucleotides) have emerged (Goodwin et al., 2016). These technologies have the potential to drastically increase the resolution of microbial diversity, but the error rate in long-read sequencing is 20-times higher than in short-read sequencing (~10% against ~0.5%) (Callahan et al., 2019) and improvements are still needed before these methods can supplant the current sequencing platforms.

During data analysis, marker-gene sequences can be clustered either in OTUs (>97% similarity) or zOTUs (100% similarity) (Section 2.1.1). The zOTUs clustering enables resolution of closely related strains with potentially different phenotypes that would otherwise be lumped into the same cluster using conventional OTUs formed at 97% sequence similarity. The use of zOTUs is thought to maximize the phylogenetic resolution of the sequencing data, but with the risk that some species may be split over several zOTUs due to intra-species variations. Finding the balance between sensitivity and specificity is the key when choosing between zOTUs and OTUs. Jia et al. (2019) reported no clear advantage of the zOTU method over conventional OTU formation method with the zOTU method likely to discard some biologically relevant information, when using the UNOISE3 algorithm with default settings. Their analysis showed that the community taxonomic compositions from OTU and zOTU analyses were similar, though the zOTU method appeared to capture less phylogenetic diversity and produced a much larger proportion (31%) of phantom taxa than the OTU method (11%) (Jia et al., 2019). On the contrary, Callahan et al. 2017 argue that zOTUs (or amplicon sequence variants, ASVs) make marker-gene sequencing more precise, comprehensive, reusable across studies and reproducible in future data sets. They also suggest that unlike OTUs, zOTUs are not limited by incomplete reference databases (Callahan et al., 2017). Edgar et al. 2018 came to the same conclusion adding that zOTUs can be directly comparable between datasets without

360 re-clustering, providing that the same genetic locus (i.e., studies using the same primer set) be
361 compared (Edgar, 2017).
362 Data analysis can also be affected by horizontal transfer of 16S rRNA genes between different
363 species. Evidence of this genetic transfer mechanism has been documented (Schouls et al.,
364 2003) and would lead to misleading inferences with species identification based on 16S rRNA
365 genes. Unless using different genetic markers in parallel, no method exists to date to distinguish
366 native 16S rRNA genes from horizontally transferred genes. Fortunately, this exchange of
367 genetic material between species is considered rare, although impossible to quantify. Copy-
368 number variations of the small subunit 18S rRNA gene, most commonly used marker in
369 eukaryotes, is also a parameter to consider when analysing sequencing data (Wang et al., 2017;
370 Guo et al., 2016; Gong & Marchetti, 2019). For instance, some species of ciliates (Wang et al.,
371 2017) and dinoflagellates (Guo et al., 2016) have hundreds or even thousands of 18S copies,
372 which can lead to misinterpretation of the actual abundance of these organisms in a sample
373 (Gong & Marchetti, 2019). Alternative molecular markers can be used to mitigate the effect of
374 copy-number variation on organism's abundance. Guo et al. (2016) reported that *actin* gene
375 was a more appropriate molecular marker than 18S rDNA for the community analysis of
376 dinoflagellates (Guo et al., 2016). The bioinformatics pipeline described by Marchetti & Gong
377 (2019) can be used to correct for variations in 18S gene copy number and thus improve the
378 accuracy of eukaryotes abundance in microbial community profiles.
379 Most bacterial genomes contain multiple copies of the 16S rRNA gene with copy number
380 varying between species (Vos et al., 2012), which can impair the microbial diversity results
381 when based on 16S relative abundances. In addition, individual genomes might contain
382 different variants of the 16S rRNA gene (Pei et al., 2010). For these reasons, single-copy,
383 essential protein-encoding marker-gene such as *rpoB* can offer potential advantages over the
384 standard 16S rRNA gene-based approaches, as described by Vos et al. (2012) Moreover,

protein-encoding gene facilitates the elimination of sequencing errors if they disrupt the reading frame. Amplicon sequencing based on ribosomal DNA is a powerful method, but other approaches should also be considered as it can further improve the accuracy of taxonomic analysis.

3. What have we obtained to date?

3.1 Microbial community profile in the MBR process

Marker-gene based approach has identified almost 97% of microorganisms in the biological reactor of the MBR process. Of which, *Proteobacteria* is the dominant phylum (by at least 25%) of the sludge community (Fig. 3). The *Proteobacteria* phylum made up of at least eight classes including β-*proteobacteria,* α-*proteobacteria*, λ-*proteobacteria* and δ-*proteobacteria*. The *Proteobacteria* phylum contains bacterial groups that are responsible for nutrient removal, including ammonia-oxidizing bacteria (AOBs), nitrite-oxidizing bacteria (NOBs) and phosphorus accumulating organisms (PAOs) (Hu et al., 2012; Ye et al., 2011; Phan et al., 2014). It has been found that the population of *Proteobacteria* was correlated significantly with the functions and performance of biological reactors.

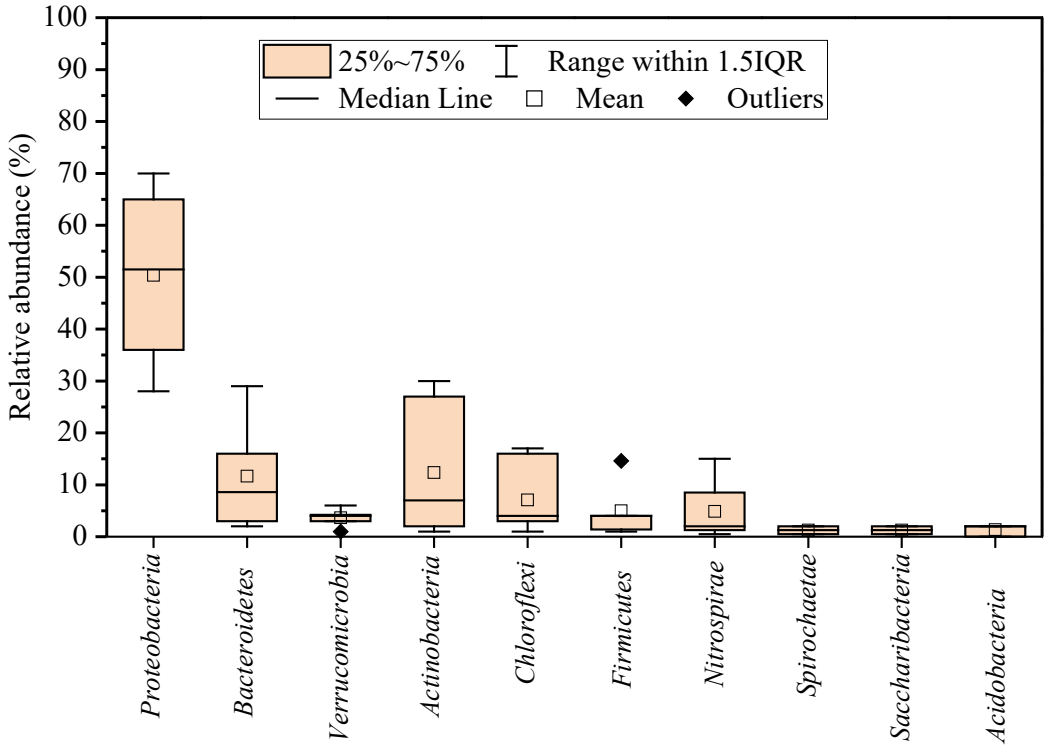


Figure 3. The relative abundance of major bacterial phylum in activated sludge of membrane bioreactor. Data were extracted from recent studies, which used high throughput sequencing technologies to detect their abundance (Phan et al., 2016; Ziegler et al., 2016).

Chloroflexi is another phylum that is frequently detected in the sludge community as well as in marine and freshwater sediments (Hug et al., 2013). The phenotype of *Chloroflexi* member includes carbon cycling, organohalide respiration, fermentation, CO_2 fixation and acetogenesis (i.e. production of volatile fatty acids and acetate) with ATP formation by substrate-level phosphorylation (Hug et al., 2013). Members of the *Chloroflexi* phylum have the ability to degrade a wide range of complex organic matters (Graber & Breznak, 2005). The abundance of *Chloroflexi* and their phenotype suggest they play a role in organic carbon removal in the MBR process.

The phylum *Saccharibacteria* was present at 0.5 to 2% of the total bacteria in the AS community. Members of *Saccharibacteria* can degrade various organic compounds in aerobic,

anoxic and anaerobic conditions (Ohashi et al., 2016). In the AS community, *Saccharibacteria* members could contribute to organic carbon removal and nitrate reduction in the AS process.

The phylum *Acidobacteria* was present in less than 2% of the total bacteria in the sludge community (Fig. 3). This phylum adapts to oligotrophic environments and contributes to carbon and nitrogen cycles (Eichorst et al., 2018). Bacteria of *Acidobacteria* phylum carry carbon metabolism-associated genes involved in the degradation of polysaccharides and aromatic compounds (Janssen et al., 2002; Hester et al., 2018). The phylum *Acidobacteria* is characterised as slow-growing microbes due to low energy generation in their metabolisms (Jones et al., 2009; Fierer et al., 2007). Their low growth rates could make it hard for them to compete with other phyla in the sludge community, explaining their low abundance.

3.2 Classification of functional microbes

3.2.1 Ammonia and nitrite-oxidising bacteria

The obtained results have unravelled the complexity of ammonia-oxidising bacteria (AOB) and nitrite-oxidising bacteria (NOB), involving in autotrophic nitrification processes in the MBR. AOB are mainly classified in the sub-class of *β-proteobacteria*, excepting *Nitrosococcus* that belongs to δ- *proteobacteria* (Table 1). NOBs are in the class of α-*proteobacteria* except *Nitrospira*.

Nitrosomonas sp. is the main functional groups of AOB in the MBR process (Phan et al., 2016). The relative abundance is much higher than the total abundance of all other AOB genera (Table 1). An abundance of *Nitrosomonas sp* suggests that the other AOB species play only a minor role in nitrification efficiency in the MBR process. Ecophysiological studies of isolated *Nitrosomonas sp* (e.g. *Nitrosomonas sp* Is79) suggested that *Nitrosomonas sp* may be resilient to fluctuating environmental conditions (e.g. presence of micropollutants, long sludge retention time) (Phan et al., 2016). *Nitrosomonas sp* Is79 is strictly aerobic, fixing carbon autotrophically from carbon dioxide and adapt to low ammonium levels (Bollmann et al.,

2013). Ammonium concentration in the wastewater varies significantly (e.g. dry *vs* wet weather, winter *vs* summer). The resilience of *Nitrosomonas sp* allows them to maintain their population in the MBR process.

Nitrospira sp. is the dominant group of NOBs (Table 1). Species of *Nitrospira* globally inhabit terrestrial and limnic environments, marine waters, deep-sea sediments, drinking water distribution systems, corroded iron pipes and WWTPs (Daims et al., 2001). The main ecological function of *Nitrospira* is nitrite oxidation. However, they also have versatile metabolism, including the utilisation of various organic compounds. Recently, it has been reported that *Nitrospira* species possess all the enzymes to catalyse the complete nitrification process (Daims et al., 2015). These species are referred to as 'comammox'. Phylogenetic analyses suggested that comammox *Nitrospira* are present in diverse environments (Daims et al., 2001; Fan et al., 2017). *Nitrospira* sp. are also present in the influent, contributing to their high abundance in the MBR process.

Heterotrophic nitrifiers including species from the genus of *Comamonas, Thauera, Accumulibacter* and *Dechloromonas* were present at 5 to 14% of total bacteria in the microbial community (Table 1). These species were previously found dominant in AS receiving ammonium-rich influent (Fan et al., 2017; Ma et al., 2015). Ma et al. (2015) observed more than 10% of heterotrophic nitrifiers (i.e. *Comamonas sp.* (6.6%)*, Thauera sp.* (4.0%) and *Azoarcus sp*. (7.8%) in six WWTPs receiving high ammonium-bearing wastewater (i.e. 300 mg/L). Species of *Accumulibacter sp*. and *Dechloromonas sp.* could also perform phosphorous removal (Section 3.4). The growth rate of heterotrophic nitrifiers is five-times faster than that of autotrophic nitrifiers. Therefore, published results often suggest that nitrogen removal is mainly due to the heterotrophic process in conventional WWTPs. In the MBR process, the addition of a membrane filter allows the operation of higher sludge retention times, promoting

the growth of autotrophic nitrifiers (Li et al., 2019). The presence of both autotrophic and heterotrophic nitrifiers could be the reason for better nitrogen removal in these MBR process.

Table 1. The relative abundance of AOB and NOB in recent MBR studies

Genera	Relative abundance (%)	MBR description	Reference
Functional group: AOB (autotrophic nitrification)			
Nitrosomonas (*Betaproteobacteria*)	4.8 – 16	Aerobic MBR receiving secondary effluent	(Cimbritz et al., 2019)
	21.3	Aerobic MBR receiving saline sewage wastewater	(Ye et al., 2011)
Nitrosomonadaceae (*Betaproteobacteria*)	0.2	Anoxic-aerobic MBR receiving synthetic wastewater	(Phan et al., 2016)
Nitrosospira (*Betaproteobacteria*)	11	Anoxic-oxic MBR receiving raw wastewater	(Sofia et al., 2004)
Nitrosovibrio (*Betaproteobacteria*)	0.2	Aerobic MBR receiving municipal wastewater	(Xia et al., 2016)
Nitrosococcus (*Deltaproteobacteria*)	0.1	Aerobic MBR receiving municipal wastewater	(Xia et al., 2016)
Functional group: NOB (autotrophic nitrification)			
Nitrospira	8.2 – 20	Aerobic MBR receiving secondary effluent	(Cimbritz et al., 2019)
	3	In the anoxic zone of anoxic-aerobic MBR receiving synthetic wastewater under infinite sludge retention time	(Phan et al., 2016)
	3.2	Aerobic MBR receiving saline sewage wastewater	(Ye et al., 2011)
	9.6	Oxic-anoxic-oxic MBR receiving municipal wastewater	(Li et al., 2019)
Functional group: AOB & NOB (heterotrophic nitrification)			
Comamonas	5.2 – 15	Anoxic-aerobic MBR receiving synthetic wastewater	(Phan et al., 2016)
Thauera	0.2 – 2		
Accumulibacter	0.2 – 1.3		
Dechloromonas	0.4 – 2.2		

The metabolism of nitrogen pathways by AOB and NOB is related to the abundances of genes coding for ammonia monooxygenase (*amo*), hydroxylamine oxidase (*hao*), nitrate reductase (*nar*), nitrite reductase (*nir*), nitric oxide reductase (*nor*) and nitrous oxide reductase

(*nos*). The presence of these functional genes indicates that the metagenomic approach can be used to investigate the functional genes of nitrifiers from the MBR process comprehensively. However, the correlation amongst the abundance of these genes, level of expression and nitrogen removal efficiency is still to be investigated.

3.2.2 Phosphate-accumulating organisms

PAOs have been identified in three main genera (Table 2). These microorganisms are ecologically significant as they remove phosphorus from wastewater. They can adapt for survival in both aerobic and anaerobic conditions. PAOs cycle molecules for energy generation or storage depending on the environment. To promote their activity as PAOs, the key is to induce appropriate conditions with the addition of anaerobic zones. It was also reported that the anaerobic micro-niches occurring in the non-enhanced biological phosphate removal MBR could promote the growth of PAOs (Silva et al., 2012; Saunders et al., 2013). The total abundance of PAO organisms was similar in enhanced biological phosphorus removal and non-enhanced biological phosphorus removal (i.e. $10 \pm 2\%$ *vs* $10 \pm 7\%$, respectively). It is suggested that the high removal of phosphorus in the enhanced biological phosphorus removal MBR is due to the high level of phosphorus accumulating in the PAOs. Therefore, the operation of MBR should favor the phosphorus accumulation process rather than promote the growth of PAOs community. PAOs (*β-proteobacteria* class) are capable of immobilising phosphorus from the mixed liquor using nitrate and oxygen as an electron acceptor in the anoxic and aeration zones of the bioreactor, respectively. By using nitrate as a final electron acceptor, the phosphorus accumulating organisms also contribute to denitrification, producing nitrogen gas.

Table 2. Relative abundance of PAOs in recent MBR studies

PAOs	Relative abundance (%)	MBR description	Reference
Candidatus Accumulibacter	0.54 – 5.54	Anoxic zone to aerobic zone to membrane zone (with aeration) Ferrous dosing at aerobic zone	(Ren et al., 2019)

	0.06 – 0.11	Anaerobic zone to anoxic zone to membrane zone (with aeration)	(Ziegler et al., 2016)
	1 – 6	Enhanced biological phosphate removal MBR	(Silva et al., 2012)
	1 – 11	Non enhanced biological phosphate removal MBR	(Silva et al., 2012)
	0.2 – 5.8	Anoxic zone to membrane zone (with aeration)	(Phan et al., 2016)
	2.8 - 15.3	Non enhanced biological phosphate removal MBR	(Saunders et al., 2013)
	3.6 – 10.1	Enhanced biological phosphate removal MBR	(Saunders et al., 2013)
Tetrasphaera	0.28 – 1.34	Anaerobic zone to anoxic zone to membrane zone (with aeration)	(Ziegler et al., 2016)
	1.1 – 19.2	Aerobic MBR	(Rodriguez-Sanchez et al., 2019)
	2 – 6	Enhanced biological phosphate removal MBR	(Silva et al., 2012)
	1 – 8	Non enhanced biological phosphate removal MBR	(Silva et al., 2012)
	0.1 – 13.7	Non enhanced biological phosphate removal MBR	(Saunders et al., 2013)
	6.7 – 9	Enhanced biological phosphate removal MBR	(Saunders et al., 2013)
Dechloromonas	1 – 6	Enhanced biological phosphate removal MBR	(Silva et al., 2012)
	1 – 9	Non enhanced biological phosphate removal MBR	(Silva et al., 2012)

The genomes of a few culturable PAOs have been sequenced and revealed the abundance of genes involved in the metabolisms of phosphorus and inorganic polyphosphate (Kawakoshi et al., 2012). These genes include polyphosphate kinase (*ppks* gene*)*, exopolyphosphatase (*ppx* gene), polyphosphate-glucose phosphotransferase (*ppgks* gene), phosphate transporters (*pits* gene) and phosphate ABC transporter (*phaABC* gene). The detection of molecular information of phosphorus accumulation process in the isolated/known PAOs paves the way for metagenomic, metatranscriptomics and metaproteomics analyses of a consortium performing phosphorus accumulation. In this way, techniques to use PAOs for phosphorus removal and recycle can be enhanced which will improve the overall effectiveness of AS using MBR.

3.3 Insight on microbial community and MBR performance relationships

503 3.3.1 Nutrient removal

504 A number of studies have demonstrated the benefit of microbial community information on
505 the operation of MBR process for nutrient removal. Ma et al. (2013) observed a reduction in
506 the population of AOB and NOB at high aeration intensity. This observation suggested that
507 MBR could be operated at low aeration rate to maintain the high abundance of NOB and AOB
508 for nitrogen removal, thus reducing operating cost used as aeration energy. Another example
509 is optimisation of ferrous dosage. Ferrous dosing has been used to enhance chemical removal
510 of phosphorus but could change the population of bacteria involved in aerobic denitrifying (e.g.
511 *Zoogloea*), anoxic denitrifying (e.g. *Dechloromonas*, *Hyphomicrobium* and *Thauera*), and
512 nitrifying bacteria (e.g. *Nitrospira*) as well as phosphorus accumulating (e.g. *Candidatus*
513 *Accumulibacter*) (Ren et al., 2019). Dosing of Fe/P 1:1 (molar ratio), there was no impact on
514 the bacterial community regulating nutrient removal. However, at a ratio of Fe/P 2:1, a sharp
515 decrease in the population of *Nitrospira*, *Dechloromonas* and *Candidatus Accumulibacter* was
516 observed (Ma et al., 2013; Ren et al., 2019). There were two main reasons for this observation,
517 i) excess ferrous dosage can outcompete the bacteria for the phosphorus, and ii) ferrous iron
518 can also induce the formation of reactive oxygen species (via *Fenton* reaction) that oxidise and
519 damage protein, lipids and DNA in cells. Optimisation of ferrous dosage should not interfere
520 with the biological removal of nitrogen and phosphorus. It is recommended that ferrous dosing
521 should be used as the post-treatment method for additional chemical removal of phosphorus
522 from the biological process.

523 *In-situ* ozonation of sludge in the MBR increased phosphorus removal due to the increased
524 abundance of PAOs (i.e. *Candidatus* and *Accumulibacter*) (Tang et al., 2019). It is likely that
525 species of *Candidatus* and *Accumulibacter* are resilient to ozonation due to their cell membrane
526 structure and morphology. On the other hand, anoxic denitrifying bacteria *Dechloromonas*
527 were inhibited by ozonation. Consequently, the efficiency of denitrification decreased in the

MBR with *in-situ* ozonation. The above-mentioned studies using the marker-gene approach have provided insights into the change of microbial community in the MBR induced by a selective pressure (i.e. a specific operating condition). However, the marker-gene approach is unable to evaluate the functional potential of the microbes. It is suggested marker-gene approach couple with metagenomic sequencing to provide insights into the microbial functional.

Flavobacterium, Thauera, Comamonas and *Dechloromonas* are key microbes for biological nutrient removal in the MBR process. Ma et al. (2016) observed the coherence between the percentages of protein-coding reads from metagenomic sequencing and taxonomic results from 16S rRNA amplicon sequencing. For example, *Flavobacterium* possess genes that encodes for proteins involved in the assimilation and dissimilation of nitrogen compounds by transporting nitrate/nitrite into cells, such as the ammonia monooxygenase gene (*amo*) known to increase the nitrification capacity of membrane bioreactors. Genes coding for denitrification enzymes are assigned to the syntrophic and heterotrophic communities in the activated sludge of the MBR, consisting of *Flavobacterium, Thauera, Dechloromonas* and *Niastella*. Identifying and tracking the prevalence of nitrogen and phosphorus removing bacteria along the operational regime may provide the information to operate MBR in a different way.

3.3.2 Biofouling

Analysis of biofilms attached to the membrane surfaces provided in depth information on microbial community structure responsible for biofouling (i.e. who colonizes the membrane surface). A study by Miura et al. (2007) revealed that bacteria in the class of *β-proteobacteria* dominated the biofilm on the membrane (61% of total bacteria) in the MBR system. Bacteria in the class of *β-proteobacteria* such as *Dechloromonas sp*. (was present at 50% abundance) express a putative lytic transglycosylase, cation transporter and pilin peptidase that are involved in host coloniosation (Salinero et al., 2009). This physiological character could promote the

formation of mature biofilm on the membrane in the MBR system. Huang et al. (2008) reported different microbial communities between suspended sludge and biofilm developed on the microfiltration membrane surface in the MBR system. Bacteria in the phylum of *Proteobacteria* dominated the community from the membrane surface, indicating that these phylotypes prefer to attach onto the membrane surface (Fig. 4) (Huang et al., 2008). Other phyla such as *Firmicutes* and *Nitrospira* were also found to colonise the membrane surface (Fig. 4) (Huang et al., 2008; Lim et al., 2012; Jo et al., 2016). *Nitrospira* bacteria are common NOB present in suspended sludge at 2-5% of total abundance. However, in a biofilm configuration, the abundance of *Nitrospira* can reach a total of 10%. The low amount of dissolved oxygen in the biofilm is likely to promote the growth of *Nitrospira* microbes.

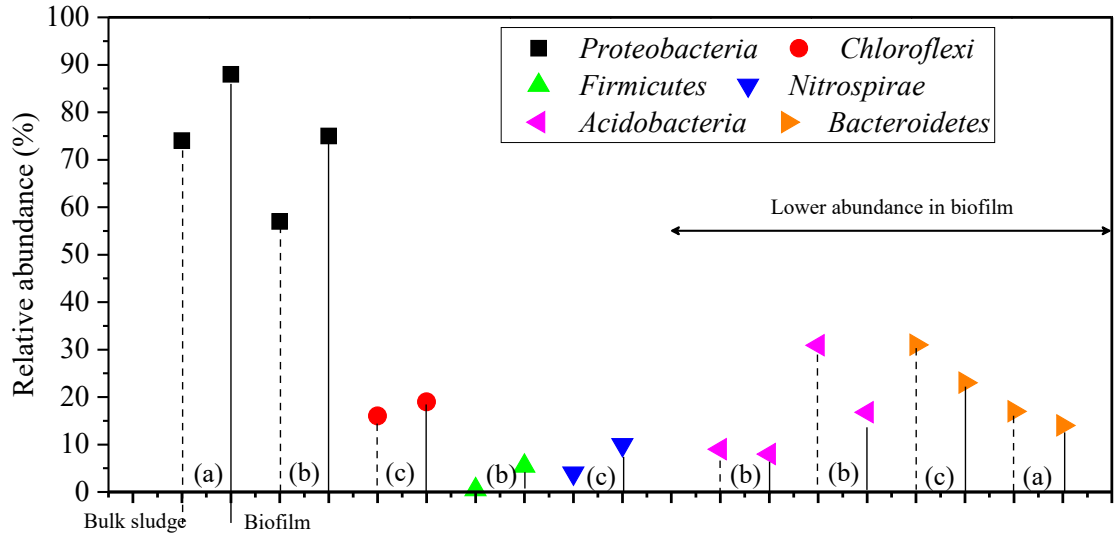


Figure 4. The relative abundance of phylum in bulk sludge (dotted line) and biofilm (plain line) during MBR operation from different studies (a) (Huang et al., 2008); (b) (Lim et al., 2012) and (c) (Jo et al., 2016).

Bacterial colonisation and biofilm development on the membrane surface in the MBR system is a complex process (Table 3). These phenomena cause unwanted biofouling problems in the MBR systems. The biofouling occurs sequentially by first cell attachment on the membrane, cell reproduction, exopolymeric substances (EPS) production and finally

membrane pore blockage. Ziegler et al. (2016) observed a high abundance of species in the genera of *Limnohabitans, Hydrogenophage* and *Malkia* on the initial stage of biofilm formation (i.e. week 1 to week 4) (Table 3). As the biofilm matured, these genera were replaced with bacteria of *Chloroflexi* phylum and *Gordonia* genus. *Dechloromonas* - a member of PAOs – many colonise the membrane surface due to their low potential of motility. A review paper by Meng et al. (2009) suggested that the genus of *Dechloromonas* causes irreversible membrane fouling during the MBR operation.

Table 3: List of biofilm forming bacteria and their phylotypes

Taxa	Relative abundance (%)	Phylotypes	Reference
Dechloromonas	3	- The low potential of motility - Cause irreversible membrane fouling	(Meng et al., 2009)
	0.58 – 9.75		(Jo et al., 2016)
Limnohabitans	3 – 13	- Pioneer species on the membrane surface - Provide initial adhesion and establishment of biofilm	(Ziegler et al., 2016)
Hydrogenophage	1.5 – 3.8		
Malkia	1.2 – 4.9		
Caldilinea	0.13 – 4.45	- Sludge bulking forming bacteria - Membrane attachment due to EPS production - Hydrophobicity cells membrane or flocs	(Jo et al., 2016)
Haliscomenobacter	0.13 -4.69		
Aeromonas	na	- Produce outer membrane proteins to aid in the colonisation	(Zhou et al., 2015)
Enterobacter			
Pseudomonas			
Thauera			
Ferruginibacter	7.46	- Non-motile bacteria	(Xiong et al., 2016)
Meiothermus	3.90	- Non-motile bacteria	(Xiong et al., 2016)
Betaproteobacteria	54.7	- Enriched in high loading MBR	(Xia et al., 2010)
Bacteroidetes	19.8		

[*] na = not available

Results from a few studies of microbial colonization, biofilm formation, and microbial community structures on the membrane surfaces imply that biofouling control strategy

583 development should focus on specific bacterial groups rather than the whole microbial
584 community presents in the sludge of MBR and aim at counteracting the mechanisms of cell
585 attachment and colonisation on membrane surfaces. The biofilm-forming bacteria are quite
586 diverse and mitigation method to target or inhibit these microorganisms is impossible to
587 develop as such a method cannot provide a selective inhibition mechanisms. One possible
588 method is the addition of carrier (i.e. activated carbon or sponge) to provide support to bacteria
589 that prefer to colonise hard surfaces. In this condition, the surface available for biofilm
590 formation will be larger, reducing potential fouling of the membrane. Addition of activated
591 carbon and sponge has demonstrated to be effective in membrane fouling control, mainly due
592 to the shear stress and scours effect (Nguyen et al., 2014). No one has tested the possible
593 hypothesis of membrane surface competition. This is probably a missing piece of the puzzle to
594 develop an effective biofouling control and mitigation process in the MBR process.

595 3.3.3 Micropollutant removal

596 Results to date have suggested the linkage between micropollutant removal and the MBR
597 microbial community (Wolff et al., 2018; Phan et al., 2016). The population of bacteria in the
598 phylum of *Proteobacteria* increased from 23% to 64% (i.e. significant difference), which
599 coincided with the observation of high removal of micropollutants (Phan et al., 2016). For
600 example, carbamazepine and gemfibrozil – two biologically recalcitrant compounds – were
601 well removed (i.e. above 50%) when there was more abundance of *Proteobacteria* in the MBR
602 systems (Phan et al., 2016). In the *Proteobacteria* phylum, Phan et al. (2016) suggested that
603 the family *Burkholderiales* may contribute to the degradation of carbamazepine and
604 gemfibrozil. Members of the *Burkholderiales* have been found to survive in limited nutrient
605 environments (Li et al., 2012) and be able to use chlorinated aliphatic compounds and aromatic
606 hydrocarbons as a source of carbon and energy (Abbai & Pillay, 2013; Boonnorat et al., 2014).
607 The presence of micro-pollutants (i.e. 22 compounds at 5 μg/L in Phan et al. (2016) and 8

608 compounds at 1000 μg/L (Boonnorat et al., 2014)) did not induce the proliferation of
609 *Burkholderiales*. The reason was thought to be the operation of MBR at infinite sludge
610 retention time. Therefore, MBR operating conditions influenced the development of different
611 microbial community structure that can effectively remove micropollutants. Amplicon and
612 metagenomic sequencing techniques provided essential clues to which microbes might be
613 beneficial to enrich in the MBR process.

614 Bacteria in the class of *α-proteobacteria*, *β-proteobacteria* and *γ-proteobacteria* are also
615 major contributors to micropollutant removal in the MBR process. Xia et al. (2012) achieved
616 high removal of antibiotics in the presence of bacteria in the class of *β-proteobacteria* and *γ-*
617 *proteobacteria*. The genus of *Rhodobacter* in the class of *α-proteobacteria* was significantly
618 enriched from 0.09% to 21% after the addition of three antibiotics in the influent of MBR (Wen
619 et al., 2018). *Rhodobacter* spp. is capable of cleaning up soil and water environments
620 contaminated with various organic and inorganic pollutants (e.g. aromatic hydrocarbons and
621 explosives) (Oberoi et al., 2015). *Rhodobacter* can generate an array of catalytic enzymes, such
622 as monooxygenase and dioxygenase (Oberoi et al., 2015) which are important for the
623 degradation of micropollutants.

624 The population of AOB and NOB has shown a positive correlation with micropollutant
625 removal efficiency in the MBR process. Species of *Nitrosomonas* sp. increased from 0.56 to
626 1.8% of total bacteria abundance in the MBR after micropollutants addition. The increment led
627 to enhanced removal of micropollutants bearing nitrogen elements (i.e. amines and amides). It
628 is suggested that *Nitrosomonas* sp. or AOB species oxidise micropollutants in similar pathways
629 to those utilised in the ammonia oxidation process. Tran et al. (2013) observed the dependence
630 of micropollutant biodegradation on the microbial community structure of the MBR process,
631 particularly species in the groups of AOBs and NOBs. In this aspect, high ammonia loading
632 stimulated the growth of AOBs and NOBs and provided a better removal of micropollutants.

633 Amplicon and metagenomic sequencing techniques provide insights to help isolating which
634 are the micropollutant-degrading microbes. It has been demonstrated that long-term exposure
635 of activated sludge microbiome to micropollutants can alter the microbial community and in
636 some cases, selectively enrich specific microbes with enhanced affinity for micropollutant
637 degradation. Nguyen et al. (2018) isolated a strain of *Bradyrhizobium* sp that can degrade
638 antibiotic ciprofloxacin from the activated sludge. The phylogenetic relatedness of newly
639 identified species to the previously cultured relatives allows follow-up ecophysiological and
640 isolation studies. Overall, there has been strong evidence on how MBR microbiome
641 composition can influence MBR ecosystems (i.e. micropollutant removal). There is a growing
642 interest in understanding and engineering of microbiomes for shaping microbiota that provides
643 ecosystems of interest.

644 **4. Future outlooks and challenges**

645 Modern molecular techniques have shed new light on understanding the microbial
646 community in the MBR process. However, there remain several challenges that need to be
647 addressed in the upcoming studies, allowing translation of microbial community knowledge to
648 process engineering and operate of MBR or other biological process efficiently.

649 As reviewed in Section 2, there are still several technical bottlenecks in the molecular
650 techniques that need to be overcome. First, the high financial cost of sequencing has led to the
651 lack of replicate measurements. Without adequate replication, the variation observed amongst
652 microbial communities may not be statistically different or may be due to artefacts of analytical
653 techniques. It is recommended that at the least, triplicate samples are required. Secondly, the
654 taxonomic classification of microbes in the community depends on comparison against a
655 reference database. The completeness of the reference database largely influences how the data
656 are analysed and explained. In the absence of a consensus methodology, the choice of the
657 database used to map the sequencing results (i.e., aligning short reads to a reference sequence)

is user-dependent. This leads to inconsistencies and limits comparisons amongst the reviewed literature. The observed differences could simply reflect the distinct molecular approaches and database used to characterize the microbial community. Future efforts to standardize a universal database (e.g. taxonomy for the organisms of wastewater treatment systems, functional proteins) will largely improve data analysis. However, this ambitious task requires contributed effort of scientists all around the world.

In this review, the data generated from sequencing techniques have been mainly used to describe and explain the two initial questions "who are there and what are their functions?" in the MBR process. However, these obtained data are often from an aftereffect. Thus, there is a long way to reach the ideal point where the MBR performance can be regulated online using a real-time microbial community analysis. This endeavor may be achievable subjecting to the development of sequencing technology and the readiness of data analysis. There is also a large research gap amongst lab-, pilot- and full-scale studies. The well-defined and control conditions in the lab- and pilot-scale studies could generate significantly different results with the full-scale studies. At the current stage, a possible recommendation is to focus on full-scale studies with a rigorous sampling plant over long time series. The microbial community data obtained from such studies can be integrated with the MBR operating conditions and performance into a network. Until then, the translation of "microbial community understanding" to the operation of better MBR plant may be achievable.

5. Conclusion

This paper reviews the state-of-the-art sequencing techniques as a new platform to unravel the complexity of microbial community in the MBR process. Two approaches including maker-gene and whole-genome sequencing have been analysed in terms of their benefits, considerations and limitations for future study. These included sample size, DNA extraction as well as bioinformatics analysis. These sequencing techniques have become increasingly more

683 powerful to provide details about microbes and their functions in the MBR process. The results
684 can be used to describe and explain the performance of the MBR process (i.e. nutrient removal,
685 biofouling and micropollutant removal). Key considerations to translate these findings to
686 practical outcomes are recommended. Results to date are significant but are still preliminary.
687 Further applications of sequencing techniques for the design and optimisation of the MBR
688 process are expected and can significantly enhance MBR performance.

689 **References**

690 [1] Sheik, A.R., Muller, E.E.L., Wilmes, P. 2014. A hundred years of activated sludge: time for a
691 rethink. Front. Microbiol., 5, 47-47. 10.3389/fmicb.2014.00047
692 [2] Xiao, K., Liang, S., Wang, X., Chen, C., Huang, X. 2019. Current state and challenges of full-scale
693 membrane bioreactor applications: A critical review. Bioresour. Technol., 271, 473-481.
694 https://doi.org/10.1016/j.biortech.2018.09.061
695 [3] Xia, S., Li, J., He, S., Xie, K., Wang, X., Zhang, Y., Duan, L., Zhang, Z. 2010. The effect of organic
696 loading on bacterial community composition of membrane biofilms in a submerged polyvinyl
697 chloride membrane bioreactor. Bioresour. Technol., 101(17), 6601-6609.
698 https://doi.org/10.1016/j.biortech.2010.03.082
699 [4] Nguyen, L.N., Hai, F.I., Kang, J., Price, W.E., Nghiem, L.D. 2012a. Removal of trace organic
700 contaminants by a membrane bioreactor–granular activated carbon (MBR–GAC) system.
701 Bioresour. Technol., 113, 169-173. https://doi.org/10.1016/j.biortech.2011.10.051
702 [5] Wolff, D., Krah, D., Dötsch, A., Ghattas, A.-K., Wick, A., Ternes, T.A. 2018. Insights into the
703 variability of microbial community composition and micropollutant degradation in diverse
704 biological wastewater treatment systems. Water Res., 143, 313-324.
705 https://doi.org/10.1016/j.watres.2018.06.033
706 [6] Nguyen, L.N., Hai, F.I., Kang, J., Price, W.E., Nghiem, L.D. 2012b. Removal of trace organic
707 contaminants by a membrane bioreactor–granular activated carbon (MBR–GAC) system.
708 Bioresour. Technol., 113, 169-173. http://dx.doi.org/10.1016/j.biortech.2011.10.051
709 [7] Wen, Q., Yang, L., Zhao, Y., Huang, L., Chen, Z. 2018. Insight into effects of antibiotics on reactor
710 performance and evolutions of antibiotic resistance genes and microbial community in a
711 membrane reactor. Chemosphere, 197, 420-429.
712 https://doi.org/10.1016/j.chemosphere.2018.01.067
713 [8] Inaba, T., Hori, T., Navarro, R.R., Ogata, A., Hanajima, D., Habe, H. 2018. Revealing sludge and
714 biofilm microbiomes in membrane bioreactor treating piggery wastewater by non-destructive

715 microscopy and 16S rRNA gene sequencing. Chem. Eng. J., 331, 75-83.
716 https://doi.org/10.1016/j.cej.2017.08.095
717 [9] Zhu, Y., Wang, Y., Jiang, X., Zhou, S., Wu, M., Pan, M., Chen, H. 2017. Microbial community
718 compositional analysis for membrane bioreactor treating antibiotics containing wastewater.
719 Chem. Eng. J., 325, 300-309. https://doi.org/10.1016/j.cej.2017.05.073
720 [10] Zhou, Z., Meng, F., He, X., Chae, S.-R., An, Y., Jia, X. 2015. Metaproteomic Analysis of Biocake
721 Proteins To Understand Membrane Fouling in a Submerged Membrane Bioreactor. Environ.
722 Sci. Technol., 49(2), 1068-1077. 10.1021/es504489r
723 [11] Salerno, C., Berardi, G., Laera, G., Pollice, A. 2019. Functional Response of MBR Microbial
724 Consortia to Substrate Stress as Revealed by Metaproteomics. Microb. Ecol. 10.1007/s00248-
725 019-01360-4
726 [12] Yang, P., Tan, G.-Y.A., Aslam, M., Kim, J., Lee, P.-H. 2019. Metatranscriptomic evidence for
727 classical and RuBisCO-mediated CO2 reduction to methane facilitated by direct interspecies
728 electron transfer in a methanogenic system. Sci. Rep., 9(1), 4116. 10.1038/s41598-019-40830-
729 0
730 [13] Brown, M.V., van de Kamp, J., Ostrowski, M., Seymour, J.R., Ingleton, T., Messer, L.F., Jeffries,
731 T., Siboni, N., Laverock, B., Bibiloni-Isaksson, J., Nelson, T.M., Coman, F., Davies, C.H.,
732 Frampton, D., Rayner, M., Goossen, K., Robert, S., Holmes, B., Abell, G.C.J., Craw, P.,
733 Kahlke, T., Sow, S.L.S., McAllister, K., Windsor, J., Skuza, M., Crossing, R., Patten, N.,
734 Malthouse, P., van Ruth, P.D., Paulsen, I., Fuhrman, J.A., Richardson, A., Koval, J., Bissett,
735 A., Fitzgerald, A., Moltmann, T., Bodrossy, L. 2018. Systematic, continental scale temporal
736 monitoring of marine pelagic microbiota by the Australian Marine Microbial Biodiversity
737 Initiative. Scientific Data, 5, 180130. 10.1038/sdata.2018.130
738 [14] Smith, K.F., Kohli, G.S., Murray, S.A., Rhodes, L.L. 2017. Assessment of the metabarcoding
739 approach for community analysis of benthic-epiphytic dinoflagellates using mock
740 communities. New Zealand Journal of Marine and Freshwater Research, 51(4), 555-576.
741 10.1080/00288330.2017.1298632
742 [15] Sakamoto, M., Ohkuma, M. 2010. Usefulness of the hsp60 gene for the identification and
743 classification of Gram-negative anaerobic rods. J. Med. Microbiol., 59(11), 1293-1302.
744 doi:10.1099/jmm.0.020420-0
745 [16] Wang, Z., Zhang, X.-X., Lu, X., Liu, B., Li, Y., Long, C., Li, A. 2014. Abundance and diversity
746 of bacterial nitrifiers and denitrifiers and their functional genes in tannery wastewater treatment
747 plants revealed by high-throughput sequencing. PLoS One, 9(11), e113603-e113603.
748 10.1371/journal.pone.0113603
749 [17] Rodriguez, D., Guerra, T.M., Forstner, M.R.J., Hahn, D. 2016. Diversity of Frankia in soil assessed
750 by Illumina sequencing of nifH gene fragments. Syst. Appl. Microbiol., 39(6), 391-397.
751 https://doi.org/10.1016/j.syapm.2016.06.007

[18] Větrovský, T., Baldrian, P. 2013. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. PLoS One, 8(2), e57923. 10.1371/journal.pone.0057923

[19] Konstantinidis, K.T., Tiedje, J.M. 2005. Towards a genome-based taxonomy for prokaryotes. J. Bacteriol., 187(18), 6258-6264. 10.1128/JB.187.18.6258-6264.2005

[20] Edgar, R.C. 2018. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. Bioinformatics, 34(14), 2371-2375. 10.1093/bioinformatics/bty113

[21] Callahan, B.J., McMurdie, P.J., Holmes, S.P. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME J, 11(12), 2639-2643. 10.1038/ismej.2017.119

[22] Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods, 7(5), 335-336. 10.1038/nmeth.f.303

[23] Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F. 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Appl. Environ. Microbiol., 75(23), 7537-7541. 10.1128/aem.01541-09

[24] Nilsson, R.H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P., Tedersoo, L. 2019. Mycobiome diversity: high-throughput sequencing and identification of fungi. Nature Reviews Microbiology, 17(2), 95-109. 10.1038/s41579-018-0116-y

[25] Andreakis, N., Høj, L., Kearns, P., Hall, M.R., Ericson, G., Cobb, R.E., Gordon, B.R., Evans-Illidge, E. 2015. Diversity of Marine-Derived Fungal Cultures Exposed by DNA Barcodes: The Algorithm Matters. PLoS One, 10(8), e0136130-e0136130. 10.1371/journal.pone.0136130

[26] Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G.A., Gregory Caporaso, J. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome, 6(1), 90. 10.1186/s40168-018-0470-z

[27] Aßhauer, K.P., Wemheuer, B., Daniel, R., Meinicke, P. 2015. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics (Oxford, England), 31(17), 2882-2884. 10.1093/bioinformatics/btv287

[28] Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G., Huttenhower, C. 2013.

Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology, 31, 814. 10.1038/nbt.2676
https://www.nature.com/articles/nbt.2676#supplementary-information
[29] Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome research, 27(5), 824-834. 10.1101/gr.213959.116
[30] Liu, C.-M., Li, D., Sadakane, K., Luo, R., Lam, T.-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics, 31(10), 1674-1676. 10.1093/bioinformatics/btv033
[31] Lu, Y.Y., Chen, T., Sun, F., Fuhrman, J.A. 2016. COCACOLA: binning metagenomic contigs using sequence COmposition, read CoverAge, CO-alignment and paired-end read LinkAge. Bioinformatics, 33(6), 791-798. 10.1093/bioinformatics/btw290
[32] Kang, D.D., Froula, J., Egan, R., Wang, Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ, 3, e1165. 10.7717/peerj.1165
[33] Stanke, M., Morgenstern, B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res., 33(Web Server issue), W465-W467. 10.1093/nar/gki458
[34] Hyatt, D., Chen, G.-L., Locascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC bioinformatics, 11, 119-119. 10.1186/1471-2105-11-119
[35] Kahlke, T., Ralph, P.J. 2019. BASTA – Taxonomic classification of sequences and sequence bins using last common ancestor estimations. Methods in Ecology and Evolution, 10(1), 100-103. 10.1111/2041-210x.13095
[36] Darling, A.E., Jospin, G., Lowe, E., Matsen, F.A.I.V., Bik, H.M., Eisen, J.A. 2014. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ, 2, e243. 10.7717/peerj.243
[37] Ounit, R., Wanamaker, S., Close, T.J., Lonardi, S. 2015. CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. BMC Genomics, 16(1), 236. 10.1186/s12864-015-1419-2
[38] Uritskiy, G.V., DiRuggiero, J., Taylor, J. 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome, 6(1), 158. 10.1186/s40168-018-0541-1
[39] Tamames, J., Puente-Sánchez, F. 2019. SqueezeMeta, A Highly Portable, Fully Automatic Metagenomic Analysis Pipeline. Front. Microbiol., 9(3349). 10.3389/fmicb.2018.03349
[40] Anders, S., Huber, W. 2010. Differential expression analysis for sequence count data. Genome Biology, 11(10), R106. 10.1186/gb-2010-11-10-r106
[41] Parks, D.H., Tyson, G.W., Hugenholtz, P., Beiko, R.G. 2014. STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics (Oxford, England), 30(21), 3123-3124. 10.1093/bioinformatics/btu494

825 [42] Wood, D.E., Salzberg, S.L. 2014. Kraken: ultrafast metagenomic sequence classification using
826 exact alignments. Genome Biology, 15(3), R46. 10.1186/gb-2014-15-3-r46
827 [43] Siezen, R.J., Galardini, M. 2008. Genomics of biological wastewater treatment. Microbial
828 Biotechnology, 1(5), 333-340. 10.1111/j.1751-7915.2008.00059.x
829 [44] Ma, J., Wang, Z., Li, H., Park, H.-D., Wu, Z. 2016. Metagenomes reveal microbial structures,
830 functional potentials, and biofouling-related genes in a membrane bioreactor. Appl. Microbiol.
831 Biotechnol., 100(11), 5109-5121. 10.1007/s00253-016-7312-3
832 [45] Nguyen, L.N., Nghiem, L.D., Pramanik, B.K., Oh, S. 2019a. Cometabolic biotransformation and
833 impacts of the anti-inflammatory drug diclofenac on activated sludge microbial communities.
834 Sci. Total Environ., 657, 739-745. 10.1016/j.scitotenv.2018.12.094
835 [46] Mahfouz, N., Caucci, S., Achatz, E., Semmler, T., Guenther, S., Berendonk, T.U., Schroeder, M.
836 2018. High genomic diversity of multi-drug resistant wastewater Escherichia coli. Scientific
837 Reports, 8(1), 8928. 10.1038/s41598-018-27292-6
838 [47] Green, J., Bohannan, B.J.M. 2006. Spatial scaling of microbial biodiversity. Trends in Ecology &
839 Evolution, 21(9), 501-507. https://doi.org/10.1016/j.tree.2006.06.012
840 [48] Hu, M., Wang, X., Wen, X., Xia, Y. 2012. Microbial community structures in different wastewater
841 treatment plants as revealed by 454-pyrosequencing analysis. Bioresour. Technol., 117, 72-79.
842 https://doi.org/10.1016/j.biortech.2012.04.061
843 [49] Nguyen, L.N., Johir, M.A.H., Commault, A., Bustamante, H., Aurisch, R., Lowrie, R., Nghiem,
844 L.D. 2019b. Impacts of mixing on foaming, methane production, stratification and microbial
845 community in full-scale anaerobic co-digestion process. Bioresource Technology, 281, 226-
846 233. https://doi.org/10.1016/j.biortech.2019.02.077
847 [50] de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le
848 Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner,
849 L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horák, A., Jaillon, O.,
850 Lima-Mendez, G., Lukeš, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent,
851 F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Acinas, S.G., Bork,
852 P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant,
853 S., Raes, J., Sieracki, M.E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker,
854 P., Karsenti, E. 2015. Eukaryotic plankton diversity in the sunlit ocean. Science, 348(6237),
855 1261605. 10.1126/science.1261605
856 [51] Gotelli, N.J., Colwell, R.K. 2001. Quantifying biodiversity: procedures and pitfalls in the
857 measurement and comparison of species richness. Ecology Letters, 4(4), 379-391.
858 10.1046/j.1461-0248.2001.00230.x
859 [52] Pérez, M.V., Guerrero, L.D., Orellana, E., Figuerola, E.L., Erijman, L. 2019. Time series genome-
860 centric analysis unveils bacterial response to operational disturbance in activated sludge.
861 mSystems, 4(4), e00169-19. 10.1128/mSystems.00169-19

862 [53] Glassing, A., Dowd, S.E., Galandiuk, S., Davis, B., Chiodini, R.J. 2016. Inherent bacterial DNA
863 contamination of extraction and sequencing reagents may affect interpretation of microbiota in
864 low bacterial biomass samples. Gut Pathog, 8, 24-24. 10.1186/s13099-016-0103-7
865 [54] Leek, J.T., Scharpf, R.B., Bravo, H.C., Simcha, D., Langmead, B., Johnson, W.E., Geman, D.,
866 Baggerly, K., Irizarry, R.A. 2010. Tackling the widespread and critical impact of batch effects
867 in high-throughput data. Nature Reviews Genetics, 11(10), 733-739. 10.1038/nrg2825
868 [55] Goh, W.W.B., Wang, W., Wong, L. 2017. Why Batch Effects Matter in Omics Data, and How to
869 Avoid Them. Trends in Biotechnology, 35(6), 498-507.
870 https://doi.org/10.1016/j.tibtech.2017.02.012
871 [56] Yang, H., Harrington, C.A., Vartanian, K., Coldren, C.D., Hall, R., Churchill, G.A. 2008.
872 Randomization in Laboratory Procedure Is Key to Obtaining Reproducible Microarray Results.
873 PLoS One, 3(11), e3724. 10.1371/journal.pone.0003724
874 [57] Holman, D.B., Brunelle, B.W., Trachsel, J., Allen, H.K. 2017. Meta-analysis To Define a Core
875 Microbiota in the Swine Gut. mSystems, 2(3), e00004-17. 10.1128/mSystems.00004-17
876 [58] Helbling, D.E., Johnson, D.R., Lee, T.K., Scheidegger, A., Fenner, K. 2015. A framework for
877 establishing predictive relationships between specific bacterial 16S rRNA sequence
878 abundances and biotransformation rates. Water Res., 70, 471-484.
879 https://doi.org/10.1016/j.watres.2014.12.013
880 [59] Wang, G.C.Y., Wang, Y. 1996. The frequency of chimeric molecules as a consequence of PCR
881 co-amplification of 16S rRNA genes from different bacterial species. Microbiology, 142(5),
882 1107-1114. https://doi.org/10.1099/13500872-142-5-1107
883 [60] Qiu, X., Wu, L., Huang, H., McDonel, P.E., Palumbo, A.V., Tiedje, J.M., Zhou, J. 2001. Evaluation
884 of PCR-Generated Chimeras, Mutations, and Heteroduplexes with 16S rRNA Gene-Based
885 Cloning. Appl. Environ. Microbiol., 67(2), 880-887. 10.1128/aem.67.2.880-887.2001
886 [61] Smyth, R.P., Schlub, T.E., Grimm, A., Venturi, V., Chopra, A., Mallal, S., Davenport, M.P., Mak,
887 J. 2010. Reducing chimera formation during PCR amplification to ensure accurate genotyping.
888 Gene, 469(1), 45-51. https://doi.org/10.1016/j.gene.2010.08.009
889 [62] Omelina, E.S., Ivankin, A.V., Letiagina, A.E., Pindyurin, A.V. 2019. Optimized PCR conditions
890 minimizing the formation of chimeric DNA molecules from MPRA plasmid libraries. BMC
891 Genomics, 20(7), 536. 10.1186/s12864-019-5847-2
892 [63] Bushnell, B., Rood, J., Singer, E. 2017. BBMerge – Accurate paired shotgun read merging via
893 overlap. PLoS One, 12(10), e0185056. 10.1371/journal.pone.0185056
894 [64] Magoč, T., Salzberg, S.L. 2011. FLASH: fast length adjustment of short reads to improve genome
895 assemblies. Bioinformatics, 27(21), 2957-2963. 10.1093/bioinformatics/btr507
896 [65] Edgar, R.C., Flyvbjerg, H. 2015. Error filtering, pair assembly and error correction for next-
897 generation sequencing reads. Bioinformatics, 31(21), 3476-3482.
898 10.1093/bioinformatics/btv401

[66] Callahan, B.J., Wong, J., Heiner, C., Oh, S., Theriot, C.M., Gulati, A.S., McGill, S.K., Dougherty, M.K. 2019. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. bioRxiv, 392332. 10.1101/392332

[67] Goodwin, S., McPherson, J.D., McCombie, W.R. 2016. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics, 17, 333. 10.1038/nrg.2016.49

[68] Jia, Y., Leung, M.H.Y., Tong, X., Wilkins, D., Lee, P.K.H. 2019. Rare Taxa Exhibit Disproportionate Cell-Level Metabolic Activity in Enriched Anaerobic Digestion Microbial Communities. mSystems, 4(1), e00208-18. 10.1128/mSystems.00208-18

[69] Edgar, R.C. 2017. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. bioRxiv, 192211. 10.1101/192211

[70] Schouls, L.M., Schot, C.S., Jacobs, J.A. 2003. Horizontal transfer of segments of the 16S rRNA genes between species of the Streptococcus anginosus group. Journal of bacteriology, 185(24), 7241-6.

[71] Wang, C., Zhang, T., Wang, Y., Katz Laura, A., Gao, F., Song, W. 2017. Disentangling sources of variation in SSU rDNA sequences from single cell analyses of ciliates: impact of copy number variation and experimental error. Proceedings of the Royal Society B: Biological Sciences, 284(1859), 20170425. 10.1098/rspb.2017.0425

[72] Guo, L., Sui, Z., Liu, Y. 2016. Quantitative analysis of dinoflagellates and diatoms community via Miseq sequencing of actin gene and v9 region of 18S rDNA. Scientific reports, 6, 34709-34709. 10.1038/srep34709

[73] Gong, W., Marchetti, A. 2019. Estimation of 18S Gene Copy Number in Marine Eukaryotic Plankton Using a Next-Generation Sequencing Approach. Frontiers in Marine Science, 6(219). 10.3389/fmars.2019.00219

[74] Vos, M., Quince, C., Pijl, A.S., de Hollander, M., Kowalchuk, G.A. 2012. A Comparison of rpoB and 16S rRNA as markers in Pyrosequencing studies of bacterial diversity. PLoS One, 7(2), e30600. 10.1371/journal.pone.0030600

[75] Pei, A.Y., Oberdorf, W.E., Nossa, C.W., Agarwal, A., Chokshi, P., Gerz, E.A., Jin, Z., Lee, P., Yang, L., Poles, M., Brown, S.M., Sotero, S., DeSantis, T., Brodie, E., Nelson, K., Pei, Z. 2010. Diversity of 16S rRNA genes within individual Prokaryotic genomes. Appl. Environ. Microbiol., 76(12), 3886. 10.1128/AEM.02953-09

[76] Ye, L., Shao, M.-F., Zhang, T., Tong, A.H.Y., Lok, S. 2011. Analysis of the bacterial community in a laboratory-scale nitrification reactor and a wastewater treatment plant by 454-pyrosequencing. Water Res., 45(15), 4390-4398. https://doi.org/10.1016/j.watres.2011.05.028

[77] Phan, H.V., Hai, F.I., Kang, J., Dam, H.K., Zhang, R., Price, W.E., Broeckmann, A., Nghiem, L.D. 2014. Simultaneous nitrification/denitrification and trace organic contaminant (TrOC) removal by an anoxic–aerobic membrane bioreactor (MBR). Bioresour. Technol., 165, 96-104. https://doi.org/10.1016/j.biortech.2014.03.094

936 [78] Hug, L.A., Castelle, C.J., Wrighton, K.C., Thomas, B.C., Sharon, I., Frischkorn, K.R., Williams,
937 K.H., Tringe, S.G., Banfield, J.F. 2013. Community genomic analyses constrain the distribution
938 of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling.
939 Microbiome, 1(1), 22. 10.1186/2049-2618-1-22
940 [79] Graber, J.R., Breznak, J.A. 2005. Folate cross-feeding supports symbiotic homoacetogenic
941 spirochetes. Appl. Environ. Microbiol., 71(4), 1883-1889. 10.1128/AEM.71.4.1883-1889.2005
942 [80] Ohashi, A., Ozaki, N., Uehara, R., Yamaoka, S., Kindaichi, T., Albertsen, M., Nielsen, P.H.,
943 Nielsen, J.L. 2016. Phylogenetic diversity and ecophysiology of Candidate phylum
944 Saccharibacteria in activated sludge. FEMS Microbiol. Ecol., 92(6). 10.1093/femsec/fiw078
945 [81] Eichorst, S.A., Trojan, D., Roux, S., Herbold, C., Rattei, T., Woebken, D. 2018. Genomic insights
946 into the Acidobacteria reveal strategies for their success in terrestrial environments. Environ.
947 Microbiol., 20(3), 1041-1063. 10.1111/1462-2920.14043
948 [82] Janssen, P.H., Yates, P.S., Grinton, B.E., Taylor, P.M., Sait, M. 2002. Improved culturability of
949 soil bacteria and isolation in pure culture of novel members of the divisions Acidobacteria,
950 Actinobacteria, Proteobacteria, and Verrucomicrobia. Appl. Environ. Microbiol., 68(5), 2391-
951 2396. 10.1128/AEM.68.5.2391-2396.2002
952 [83] Hester, E.R., Harpenslager, S.F., van Diggelen, J.M.H., Lamers, L.L., Jetten, M.S.M., Lüke, C.,
953 Lücker, S., Welte, C.U. 2018. Linking nitrogen load to the structure and function of wetland
954 soil and rhizosphere microbial communities. mSystems, 3(1), e00214-17.
955 10.1128/mSystems.00214-17
956 [84] Jones, R.T., Robeson, M.S., Lauber, C.L., Hamady, M., Knight, R., Fierer, N. 2009. A
957 comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library
958 analyses. The ISME J, 3(4), 442-453. 10.1038/ismej.2008.127
959 [85] Fierer, N., Bradford, M.A., Jackson, R.B. 2007. Toward an ecological classification of soil bacteria.
960 Ecology, 88(6), 1354-1364. doi:10.1890/05-1839
961 [86] Phan, H.V., Hai, F.I., Zhang, R., Kang, J., Price, W.E., Nghiem, L.D. 2016. Bacterial community
962 dynamics in an anoxic-aerobic membrane bioreactor – Impact on nutrient and trace organic
963 contaminant removal. Int. Biodeterior. Biodegrad., 109, 61-72.
964 https://doi.org/10.1016/j.ibiod.2016.01.002
965 [87] Bollmann, A., Sedlacek, C.J., Norton, J., Laanbroek, H.J., Suwa, Y., Stein, L.Y., Klotz, M.G., Arp,
966 D., Sayavedra-Soto, L., Lu, M., Bruce, D., Detter, C., Tapia, R., Han, J., Woyke, T., Lucas,
967 S.M., Pitluck, S., Pennacchio, L., Nolan, M., Land, M.L., Huntemann, M., Deshpande, S., Han,
968 C., Chen, A., Kyrpides, N., Mavromatis, K., Markowitz, V., Szeto, E., Ivanova, N., Mikhailova,
969 N., Pagani, I., Pati, A., Peters, L., Ovchinnikova, G., Goodwin, L.A. 2013. Complete genome
970 sequence of Nitrosomonas sp. Is79, an ammonia oxidizing bacterium adapted to low
971 ammonium concentrations. Standards in Genomic Sciences, 7(3), 469-482.
972 10.4056/sigs.3517166

973 [88] Daims, H., Nielsen, J.L., Nielsen, P.H., Schleifer, K.-H., Wagner, M. 2001. In Situ Characterization
974 of Nitrospira-Like Nitrite-Oxidizing Bacteria Active in Wastewater Treatment Plants. Appl.
975 Environ. Microbiol., 67(11), 5273. 10.1128/AEM.67.11.5273-5284.2001
976 [89] Daims, H., Lebedeva, E.V., Pjevac, P., Han, P., Herbold, C., Albertsen, M., Jehmlich, N.,
977 Palatinszky, M., Vierheilig, J., Bulaev, A., Kirkegaard, R.H., von Bergen, M., Rattei, T.,
978 Bendinger, B., Nielsen, P.H., Wagner, M. 2015. Complete nitrification by Nitrospira bacteria.
979 Nature, 528, 504. 10.1038/nature16461
980 https://www.nature.com/articles/nature16461#supplementary-information
981 [90] Fan, X.-Y., Gao, J.-F., Pan, K.-L., Li, D.-C., Dai, H.-H. 2017. Temporal dynamics of bacterial
982 communities and predicted nitrogen metabolism genes in a full-scale wastewater treatment
983 plant. RSC Advances, 7(89), 56317-56327. 10.1039/C7RA10704H
984 [91] Ma, Q., Qu, Y., Shen, W., Zhang, Z., Wang, J., Liu, Z., Li, D., Li, H., Zhou, J. 2015. Bacterial
985 community compositions of coking wastewater treatment plants in steel industry revealed by
986 Illumina high-throughput sequencing. Bioresour. Technol., 179, 436-443.
987 https://doi.org/10.1016/j.biortech.2014.12.041
988 [92] Li, N., Zeng, W., Miao, Z., Wang, B., Li, Z., Peng, Y. 2019. Enhanced nitrogen removal and in
989 situ microbial community in a two-step feed oxic/anoxic/oxic-membrane bioreactor (O/A/O-
990 MBR) process. J Chem Technol Biotech., 94(4), 1315-1322. doi:10.1002/jctb.5889
991 [93] Silva, A.F., Carvalho, G., Oehmen, A., Lousada-Ferreira, M., van Nieuwenhuijzen, A., Reis,
992 M.A.M., Crespo, M.T.B. 2012. Microbial population analysis of nutrient removal-related
993 organisms in membrane bioreactors. Appl. Microbiol. Biotechnol., 93(5), 2171-2180.
994 10.1007/s00253-011-3499-5
995 [94] Saunders, A.M., Larsen, P., Nielsen, P.H. 2013. Comparison of nutrient-removing microbial
996 communities in activated sludge from full-scale MBRs and conventional plants. Water Sci.
997 Technol., 68(2), 366-371. 10.2166/wst.2013.183
998 [95] Kawakoshi, A., Nakazawa, H., Fukada, J., Sasagawa, M., Katano, Y., Nakamura, S., Hosoyama,
999 A., Sasaki, H., Ichikawa, N., Hanada, S., Kamagata, Y., Nakamura, K., Yamazaki, S., Fujita,
1000 N. 2012. Deciphering the genome of polyphosphate accumulating actinobacterium
1001 Microlunatus phosphovorus. DNA research : an international journal for rapid publication of
1002 reports on genes and genomes, 19(5), 383-394. 10.1093/dnares/dss020
1003 [96] Ma, J., Wang, Z., Yang, Y., Mei, X., Wu, Z. 2013. Correlating microbial community structure and
1004 composition with aeration intensity in submerged membrane bioreactors by 454 high-
1005 throughput pyrosequencing. Water Res., 47(2), 859-869.
1006 https://doi.org/10.1016/j.watres.2012.11.013
1007 [97] Ren, B., Li, C., Zhang, X., Zhang, Z. 2019. Fe(II)-dosed ceramic membrane bioreactor for
1008 wastewater treatment: Nutrient removal, microbial community and membrane fouling analysis.
1009 Sci. Total Environ., 664, 116-126. https://doi.org/10.1016/j.scitotenv.2019.02.019

[98] Tang, S., Li, J., Zhang, Z., Ren, B., Zhang, X. 2019. Comparison of long-term ceramic membrane bioreactors without and with in-situ ozonation in wastewater treatment: Membrane fouling, effluent quality and microbial community. Sci. Total Environ., 652, 788-799. https://doi.org/10.1016/j.scitotenv.2018.10.284

[99] Miura, Y., Watanabe, Y., Okabe, S. 2007. Membrane Biofouling in Pilot-Scale Membrane Bioreactors (MBRs) Treating Municipal Wastewater: Impact of Biofilm Formation. Environ. Sci. Technol., 41(2), 632-638. 10.1021/es0615371

[100] Salinero, K.K., Keller, K., Feil, W.S., Feil, H., Trong, S., Di Bartolo, G., Lapidus, A. 2009. Metabolic analysis of the soil microbe Dechloromonas aromatica str. RCB: indications of a surprisingly complex life-style and cryptic anaerobic pathways for aromatic degradation. BMC genomics, 10, 351-351. 10.1186/1471-2164-10-351

[101] Huang, L.-N., De Wever, H., Diels, L. 2008. Diverse and distinct bacterial communities induced biofilm fouling in membrane bioreactors operated under different conditions. Environ. Sci. Technol., 42(22), 8360-8366. 10.1021/es801283q

[102] Lim, S., Kim, S., Yeon, K.-M., Sang, B.-I., Chun, J., Lee, C.-H. 2012. Correlation between microbial community structure and biofouling in a laboratory scale membrane bioreactor with synthetic wastewater. Desalination, 287, 209-215. https://doi.org/10.1016/j.desal.2011.09.030

[103] Jo, S.J., Kwon, H., Jeong, S.-Y., Lee, C.-H., Kim, T.G. 2016. Comparison of microbial communities of activated sludge and membrane biofilm in 10 full-scale membrane bioreactors. Water Res., 101, 214-225. https://doi.org/10.1016/j.watres.2016.05.042

[104] Ziegler, A.S., McIlroy, S.J., Larsen, P., Albertsen, M., Hansen, A.A., Heinen, N., Nielsen, P.H. 2016. Dynamics of the Fouling Layer Microbial Community in a Membrane Bioreactor. PLoS One, 11(7), e0158811. 10.1371/journal.pone.0158811

[105] Meng, F., Chae, S.-R., Drews, A., Kraume, M., Shin, H.-S., Yang, F. 2009. Recent advances in membrane bioreactors (MBRs): Membrane fouling and membrane material. Water Res., 43(6), 1489-1512. https://doi.org/10.1016/j.watres.2008.12.044

[106] Nguyen, L.N., Hai, F.I., Nghiem, L.D., Kang, J., Price, W.E., Park, C., Yamamoto, K. 2014. Enhancement of removal of trace organic contaminants by powdered activated carbon dosing into membrane bioreactors. J. Taiwan Inst. Chem. Eng., 45(2), 571-578. https://doi.org/10.1016/j.jtice.2013.05.021

[107] Li, R., Khafipour, E., Krause, D.O., Entz, M.H., de Kievit, T.R., Fernando, W.G.D. 2012. Pyrosequencing reveals the influence of organic and conventional farming systems on bacterial communities. PLoS One, 7(12), e51897. 10.1371/journal.pone.0051897

[108] Abbai, N.S., Pillay, B. 2013. Analysis of Hydrocarbon-Contaminated Groundwater Metagenomes as Revealed by High-Throughput Sequencing. Molecular Biotechnology, 54(3), 900-912. 10.1007/s12033-012-9639-z

[109] Boonnorat, J., Chiemchaisri, C., Chiemchaisri, W., Yamamoto, K. 2014. Microbial adaptation to biodegrade toxic organic micro-pollutants in membrane bioreactor using different sludge sources. Bioresour. Technol., 165, 50-59. https://doi.org/10.1016/j.biortech.2014.04.024

[110] Xia, S., Jia, R., Feng, F., Xie, K., Li, H., Jing, D., Xu, X. 2012. Effect of solids retention time on antibiotics removal performance and microbial communities in an A/O-MBR process. Bioresour. Technol., 106, 36-43. https://doi.org/10.1016/j.biortech.2011.11.112

[111] Oberoi, A.S., Philip, L., Bhallamudi, S.M. 2015. Biodegradation of Various Aromatic Compounds by Enriched Bacterial Cultures: Part A–Monocyclic and Polycyclic Aromatic Hydrocarbons. Appl. Biochem. Biotechnol., 176(7), 1870-1888. 10.1007/s12010-015-1684-1

[112] Tran, N.H., Urase, T., Ngo, H.H., Hu, J., Ong, S.L. 2013. Insight into metabolic and cometabolic activities of autotrophic and heterotrophic microorganisms in the biodegradation of emerging trace organic contaminants. Bioresour. Technol., 146, 721-731. https://doi.org/10.1016/j.biortech.2013.07.083

[113] Nguyen, L.N., Nghiem, L.D., Oh, S. 2018. Aerobic biotransformation of the antibiotic ciprofloxacin by Bradyrhizobium sp. isolated from activated sludge. Chemosphere, 211, 600-607. https://doi.org/10.1016/j.chemosphere.2018.08.004