Enzyme and Microbial Technology 102 (2017) 16–25

Contents lists available at ScienceDirect

Enzyme and Microbial Technology

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

Unraveling microbial structure and diversity of activated sludge in a full-scale simultaneous nitrogen and phosphorus removal plant using metagenomic sequencing

Jianhua Guo[a,b,*], Bing-Jie Ni[a], Xiaoyu Han[c], Xueming Chen[a], Philip Bond[a], Yongzhen Peng[b], Zhiguo Yuan[a]

[a] Advanced Water Management Centre (AWMC), The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia
[b] Key Laboratory of Beijing for Water Quality Science and Water Environmental Recovery Engineering, Engineering Research Center of Beijing, Beijing University of Technology, Beijing 100124, PR China
[c] Beijing Drainage Group Co., Ltd, Beijing 100022, PR China

ARTICLE INFO

Keywords:
Activated sludge
Metagenomic sequencing
Microbial community structure
Biological nitrogen removal
Bacterial diversity

ABSTRACT

Activated sludge contains highly complex microbial communities, which play crucial roles in pollutant removal performance in wastewater treatment plants (WWTPs). Metagenomic sequencing was applied to characterize microbial community and functional profiles within activated sludge from a full-scale municipal WWTP carrying out simultaneous nitrogen and phosphorous removal (SNPR). We applied the assembled contigs (N90 of 591 bp) and predicted genes to conduct taxonomic and function annotations, respectively. Results revealed the extraordinary microbial diversity of activated sludge, which included detection of minority populations that are difficult to be explored by traditional molecular methods. Taxonomic analysis indicated that the dominant bacterial phyla were *Proteobacteria*, *Nitrospirae*, *Bacteroidetes*, *Actinobacteria* and *Firmicutes*. The abundance of the key organisms involved in nitrogen and phosphorous removal were qualified. Aerobic ammonia-oxidizing bacteria distinctly dominate over ammonia-oxidizing archaea and anaerobic ammonium oxidation bacteria. Various key enzymes involved in the global nitrogen cycle were annotated in the activated sludge. High abundance of the known polyphosphate accumulating organisms was detected (approximately 4.89% of the overall population reads), supporting good phosphorous removal performance. This study provides a comprehensive insight into the community structure and diversity of the SNPR system, and will provide foundation for optimal operation of nutrient removal systems.

1. Introduction

Activated sludge technology is widely used to treat domestic and industrial wastewater in wastewater treatment plants (WWTPs) [1]. Simultaneous nitrogen and phosphorous from wastewater is often a mandatory requirement to protect receiving waters from eutrophication. Although many aspects of the engineering and design of full-scale WWTPs performing biological nutrient (nitrogen and phosphorous) removal (BNR) have been improved, a number of operational problems are still encountered and the nutrient removal efficiency can at times deteriorate. For example, sub-optimal performance has been associated with the competition between beneficial and detrimental bacteria, such as polyphosphate accumulating organisms (PAOs) and the glycogen-accumulating organisms (GAOs) [2,3], as well as floc-forming bacteria *versus* filamentous bacteria in WWTPs [4,5].

Activated sludge contains highly complex microbial communities [6,7], which will determine the overall operational performance of the WWTP, including nutrient removal efficiency and sludge settleability. Sludge population optimisation is proposed for performance control of biological wastewater treatment systems [8]. However, major barriers of sludge population optimisation are attributed to inadequate microbiological data, specifically community structure, function and growth kinetics. Consequently, more comprehensive understanding of the microbial communities and manipulation of those can potentially lead to improved operation of WWTPs. In particular, more fundamental knowledge of microbial interactions is essential to elucidate the biological mechanisms behind operational problems such as poor nutrient removal or serious sludge bulking.

Recently, modern molecular techniques for microbial community analysis have developed rapidly [7]. Molecular approaches, such as

* Corresponding author at: Advanced Water Management Centre (AWMC), The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia.
E-mail address: j.guo@awmc.uq.edu.au (J. Guo).

http://dx.doi.org/10.1016/j.enzmictec.2017.03.009
Received 21 November 2016; Received in revised form 19 March 2017; Accepted 20 March 2017
Available online 22 March 2017
0141-0229/

fluorescent *in situ* hybridization (FISH), denaturing gradient gel electrophoresis (DGGE), and 16S rRNA gene amplification and clone-based sequencing are commonly used to reveal microbial communities [9]. However, activated sludge can contain complex diversity and approaches are required for deep analysis and relatively complete detection of populations important to the system [2,10,11]. This is being realized through high-throughput sequencing of PCR amplified genes, however, due to the bias of amplification this provides qualitative rather than quantitative information [12,13]. Very recently, metagenomic sequencing (*e.g.* 454 pyrosequencing and Illumina sequencing technologies) are adopted to reveal microbial communities as well as the functional potential of the complex activated sludge [14]. This approach makes it possible to obtain valuable information on low abundant members and their possible roles in the activated sludge treatment system.

So far, a few metagenomic studies have been conducted on full-scale WWTPs [5,15,16]. Using 454 pyrosequencing, the pathogenic bacteria [17] and the bacterial diversity of activated sludge samples [15,18,19], as well as the microbial community structure of anaerobic digesters [20] have been revealed. With increased capacity, less expensive and more reliable for quantitative assessment of genetic diversity, Illumina sequencing is likely more appropriate for metagenomic analysis of natural or engineered systems [21]. This has been performed on human gut samples [22], full-scale anaerobic digesters [23,24], and drinking water [25] to reveal the microbial composition and microbial diversity of activated sludge full-scale WWTPs [2,5,26,27]. As expected, metagenomic sequencing has been demonstrated to be a very powerful tool for unraveling community structure in engineered ecosystems with extremely high diversity. Compared to numerous metagenomic studies focusing on community structure in activated sludge WWTPs with organic pollutants removal [15,26], there are limited metagenomic studies on full-scale WWTPs performing simultaneous biological nitrogen and phosphorous removal so far [2].

The aim of this study was to reveal microbial community structure, diversity and functional traits related to both nitrogen and phosphorus removal in a full-scale BNR activated sludge system. To achieve this, we extracted DNA from activated sludge collected from a full-scale WWTP carrying out simultaneous nitrogen and phosphorus removal (SNPR), and conducted high-throughout metagenomic sequencing on the Illumina HiSeq2000 platform. The microbial community structures, functional profiles, and key metabolic pathways of the activated sludge community were comprehensively analysed. In particular, the abundance and diversity of the key organisms involved in both N and P removals were qualified. This included analysis of the ammonia-oxidizing bacteria (AOB), nitrite-oxidizing bacteria (NOB), ammonia-oxidizing archaea (AOA), anaerobic ammonia oxidation bacteria (Anammox), denitrifying bacteria, and the PAOs and GAOs. Sequences associated with both N and P metabolisms were analysed in detail from the metagenomic data.

2. Materials and methods

2.1. Sampling, DNA extraction, library construction and sequencing

A 50 mL sample of activated sludge was taken using a plastic dipper from an aeration tank of a full-scale WWTP, with a mean influent flow of 1×10^6 m^3/day. This WWTP can simultaneously remove organic carbon, nitrogen and phosphate. The detailed operational conditions and operational performance are shown in Table 1 in Ref. [28].

Total DNA was extracted within 24 h after sampling from biomass sample using The FastDNA SPIN Kit for Soil (QBIOgene Inc., Carlsbad, CA, USA). The metagenomic sequencing was performed using Illumina HiSeq 2000 platform. Around 4.5 Gb reads were generated for the metagenomic dataset. After DNA trimming, a total of 4.0 Gb high quality DNA reads were used for the following bioinformatic analyses. The detailed DNA extraction, library construction and sequencing can be found in Ref. [28]. The sequence data were deposited in MG-RAST server (project ID: mgm4735473.3).

2.2. Bioinformatic analyses

2.2.1. DNA assembly and gene prediction from contigs

After quality filtering, clean reads were assembled into contigs using SOAPdenovo assembler (v 1.05) [29]. SOAP2 [30] was further adopted to align all reads to the contigs [31]. The detailed parameters used for DNA assembly can be found in the previous study [32]. Based on assembled contigs, open reading frames (ORFs) were predicted using MetaGeneMark (version 2.10, default setting) [33].

2.2.2. Taxonomic classification

The study performed the taxonomic classification by using two different analysis methods, including mapping reads against to database or mapping contigs to database. Considering most of existing studies used reads for taxonomic classification [27,34], DNA datasets were firstly annotated against the Metagenomics Rapid Annotation (MG-RAST) server (v3.1) [35]. The taxonomic profile was generated through Best Hit classification, with an E-value cutoff of 10^{-5} and a minimum alignment length of 50 bp, based on all the annotation source databases used by MG-RAST. It is recommended that merging short reads into contigs could increase annotation quality in taxonomic [23]. Thus, taxonomic classification was further preformed by search the contigs against the NCBI NT database using SOAP2 (v2.21, with the default settings) [31]. The detailed taxonomic analysis can be found in the previous study [32].

2.2.3. Global functional annotation

To get whole metabolic pathways information, protein sequences of the predicted genes were mapped against SEED Subsystems, Clusters of Orthologous Groups of proteins (COG), and Kyoto Encyclopedia of Genes and Genomes (KEGG, version 59) [36,37] databases. For these annotations, the minimum alignment length cutoff and the maximum E-value cutoff and were also set as 17 amino acids (aa) and 10^{-5}, respectively [23,38].

2.2.4. Nitrogen and phosphorus metabolic pathway analyses

Detailed analysis of the DNA datasets was conducted by quantifying sequence hits corresponding to enzyme subunits in nitrogen and phosphorus metabolisms. The module 'KEGGviewer' in MEGAN was used to analyze pathways, in particular for nitrogen metabolic pathways [39,40]. Genes or gene categories not represented in KEGG, were manually analysed through keyword searches based on NCBI-NR annotations of BLASTX outputs. The manual searches focused on key enzymes related to nitrogen fixation; ammonification; nitrification; denitrification; Anammox and dissimilatory nitrate reduction to ammonium (DNRA) [32]. This included genes for nitrogen fixation (*nif*); ammonia monooxygenase (*amo*); hydroxylamine oxidoreductase (*hao*); nitrate reductase (*nar*); nitrite reductase (*nir*); nitric oxide reductase (*nor*); nitrous oxide reductase (*nos*); hydrazine dehydrogenase (*hdh*); hydrazine synthase (*hzs*) and dissimilatory nitrite reductase (*nrf*). Furthermore; polyphosphate kinase gene (*ppk*) and exopolyPase (*ppx*); which play an important role in enhanced biological phosphorus removal (EBPR); were also analysed according to the metagenomic sequencing data.

3. Results

3.1. Operational performance of the WWTP

The WWTP is operated as an anaerobic-anoxic-oxic configuration. The influent and effluent wastewater characteristics for 6 months before sampling time show that the average chemical oxygen demand (COD) and suspended solids (SS) removal efficiencies were around 90%