Bioresource Technology 129 (2013) 209–218

Contents lists available at SciVerse ScienceDirect

Bioresource Technology

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

Metagenomic analysis reveals the prevalence of biodegradation genes for organic pollutants in activated sludge

Hua Fang [a,b], Lin Cai [a], Yunlong Yu [b], Tong Zhang [a,*]

[a] Environmental Biotechnology Laboratory, Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, SAR
[b] Department of Plant Protection, College of Agriculture & Biotechnology, Zhejiang University, Hangzhou 310058, China

HIGHLIGHTS

- High prevalence of biodegradation genes was reported in activated sludge.
- The *p450* gene was determined as the most abundant biodegradation gene.
- The *pmo* (2-monoxygenase) gene was the most dominant phenol degradation gene.
- Approximately 11.5–40.0% of biodegradation genes were carried on plasmids.
- *Mycobacterium* was the dominant population for degrading organic pollutants.

ARTICLE INFO

Article history:
Received 13 August 2012
Received in revised form 7 November 2012
Accepted 11 November 2012
Available online 23 November 2012

Keywords:
Biodegradation genes
High-throughput sequencing
Metagenomic analysis
Activated sludge
Organic pollutants

ABSTRACT

The abundance, diversity, and distribution of biodegradation genes (BDGs) and phenol degradation genes (PDGs) in activated sludge (AS) from two wastewater treatment plants (WWTPs) at different sampling times were assessed by metagenomic analysis using a total of 15 datasets derived from Illumina high-throughput sequencing and BLAST comparisons to BDGs and PDGs databases. The results showed that the abundance (0.015–0.030%) and diversity of BDGs in AS varied with the WWTP and the sampling times. The *p450* and *pmo* genes were the most abundant genes in the BDGs and PDGs subgroups, respectively. MG-RAST analysis revealed that 87 detected bacterial genera potentially capable of degrading pollutants were mostly affiliated with *Proteobacteria* (59.8%), *Bacteroidetes* (17.2%), and *Actinobacteria* (9.2%). *Mycobacterium*, belonging to *Actinobacteria*, was found to be the most abundant genus (23.4%). This method could be used to monitor an AS's biodegradation ability for organic pollutants and to evaluate its wastewater treatment efficiency.

1. Introduction

Wastewater and sludge are considered to be the most common environmental sink for organic pollutants from domestic and industrial sources, including 114 organic compounds which have been designated as priority pollutants by United States Environmental Protection Agency. Some of priority pollutants have been frequently and persistently detected in wastewater and sludge systems and pose a potential risk to natural environments and human health (Negreanu et al., 2012). A number of studies have reported that they could be removed by wastewater treatment processes (Sales et al., 2011). It is worth noting that complex microbial consortia containing various biodegradation genes (BDGs) in activated sludge (AS) play a key role in removing or detoxifying those priority chemical pollutants (Suenaga et al., 2007). Therefore, there is an increasing interest in revealing the abundance, diversity, and distribution of BDGs and the corresponding species present in AS.

The priority pollutant biodegradation capability of microorganisms and related BDGs in AS are usually studied using traditional cultivation-dependent, conventional PCR or quantitative polymerase chain reaction (qPCR) techniques (Felfoldi et al., 2010). Microbial degradation involves many important chemical processes, including dehalogenation, dealkylation, hydrolysis, oxidation, reduction, ring cleavage, conjugation, and methylation, which are catalysed correspondingly by various enzymes (e.g., monooxygenase, laccase, dioxygenase, cytochrome P450, dehydrogenase, lignin peroxidase, esterase, and dehalogenase) encoded by BDGs of microorganisms in biological wastewater treatment systems (Jechalke et al., 2011). Bunge et al. (2003) reported that microbial populations could dechlorinate recalcitrant chlorinated benzenes and dioxins, utilising them as carbon sources for growth by dehalogenation. van Beilen and Funhoff (2007) reported that cytochrome P450 (CYP) enzyme systems were mainly involved in the

* Corresponding author. Tel.: +852 2857 8551; fax: +852 2859 8987.
E-mail address: zhangt@hku.hk (T. Zhang).

0960-8524/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.biortech.2012.11.054

biodegradation of oil, chlorinated hydrocarbons, fuel additives, and many other compounds. Li et al. (2011) reported a dioxygenase produced by *Fusarium* sp. HJ01 and used for 4-chlorophenol degradation. In addition, Wu et al. (2011) isolated an *n*-alkane-degrading strain of *Burkholderia cepacia* GS3C and considered monooxygenation to be the first step of its aerobic biodegradation of alkane. Although culture- and PCR-based approaches have revealed some BDGs in AS, more than 99% of microorganisms are thought to be uncultivable or difficult to culture in a laboratory and primers used for specific PCR are also limited, which may constrain our comprehensive understanding of the diversity of BDGs in AS (Davis et al., 2011). Recently, advanced metagenomic approaches based on high-throughput sequencing, which can be used to directly sequence genomic DNA, have been widely used in studies to explore the structures and functions of microbial communities because this methodology may overcome the biases of both culture-dependent approaches and PCR-based methods. This promising approach will provide a great opportunity and new insights to reveal the composition of microbial communities, the diversity of BDGs, novel BDGs in the environment. Debroas et al. (2009) studied the taxonomic and metabolic functional diversity of bacterial communities in a lacustrine ecosystem using a metagenomic approach. Yergeau et al. (2012), using a metagenomic approach, found that *Pseudomonas* sp. expressing hydrocarbon-degrading genes was the most abundant microorganism in diesel-contaminated Canadian high Arctic soils. However, little information is currently available on the application of metagenomic analysis to explore the prevalence of BDGs in AS.

In the present study, a comprehensive BDGs database and a specific PDGs database were first constructed by retrieving annotated sequences and related information. Illumina Hiseq 2000 high-throughput sequencing and metagenomic analyses were then performed to reveal the relative abundance and diversity of BDGs and PDGs in 9 AS samples collected over 4 years. The objective of this study were (1) to reveal temporal variation in the abundance and diversity of BDGs and PDGs in AS samples at different sampling times; (2) to determine the abundance and diversity of BDGs and PDGs present on plasmids; (3) to annotate bacterial genera with biodegradation potential in the AS samples. To the best of our knowledge, this study describes the first effort to reveal the abundance and diversity of BDGs in AS at different sampling times using metagenomic analysis, which could be used to monitor an AS's biodegradation ability for organic pollutants and to evaluate wastewater treatment efficiency.

2. Methods

2.1. Activated sludge sampling

Eight AS samples were collected at different time from Shatin Wastewater Treatment Plant (WWTP, Hong Kong) and designated ST01, ST02, ST03, ST04, ST05, ST06, ST07, and ST08. One AS sample (ID: SL15) was collected from the Stanley WWTP (Hong Kong). The Shatin WWTP has treated saline wastewater with a capacity of 205,000 m^3/d for 30 years, and its current aeration tank retention time is 11 h. The Stanley WWTP contains 3 large caverns, which are each 120 m long, 15 m wide and 17 m high, together with over 450 m of access roads and ventilation tunnels and shafts, and has treated freshwater wastewater with a capacity of 11,600 m^3/d for more than 17 years. The effluents of both WWTPs flow into the South China Sea. Sampling information and characteristics of influents and effluents are summarised in Table S1. On site, the AS samples were concentrated by brief settling and then fixed by mixing with 100% ethanol at a volume ratio of 1:1. The fixed samples were kept on ice and immediately transported to the laboratory for DNA extraction.

2.2. DNA extraction

For each AS sample, 10 ml was centrifuged for 10 min at 4000*g* at 4 °C. Pellets (approximately 200 mg) were collected for DNA extraction using the FastDNA SPIN Kit for soil (MP Biomedicals, Solon, OH, USA). The individual DNA extracts were visualised using 0.8% gel electrophoresis, and the DNA concentrations and purities of the extracts were determined by microspectrophotometry (NanoDrop-1000, Thermo Scientific, Willmington, DE, USA). Finally, the nine quality-checked DNA samples were used for shotgun paired-end library construction and subsequent high-throughput sequencing.

2.3. High-throughput sequencing

High-throughput sequencing was performed by BGI (Shenzhen, China) using an Illumina Hiseq 2000 (Illumina Inc., San Diego, CA, USA). The sequencing strategy was an index PE101+8+101 cycle (paired-end sequencing, 101-bp reads with an 8-bp index sequence). Approximately 5 Gb of metagenomic data were generated from each sample. Additionally, 30 Gb was generated for the ST03 sample using the same sequencing library for a deep sequence analysis. To evaluate the reproducibility of the Illumina deep sequencing and to fairly compare all samples of the same data size, the ST03 deep sequencing dataset was randomly divided into 6 subsamplings (IDs: ST09, ST10, ST11, ST12, ST13, and ST14, 5 Gb each). Technical duplicate reads and the raw reads containing three or more "N" or contaminated with adaptors were removed from the dataset and the remaining clean reads (>98.9%) were normalised and used for further analysis.

2.4. Database construction

The BDGs protein database contained 15 sub-databases (designed using the gene names *alkb*, *benA*, *bph*, *bphA1*, *bphA2*, *carA*, *dbfA1*, *dxnA*, *dxnA-dbfA1*, *glx*, *mmox*, *npah*, *p450*, *ppah*, and *ppo*) provided by six contributors and could be downloaded from the public Fungene protein database (http://fungene.cme.msu.edu/index.spr) (Bateman et al., 2004). Proteins from each subgroup were merged and replicated proteins were filtered out using a self-written Python script. The final version of the BDGs database contained 45,319 non-redundant protein sequences capable of degrading different kinds of organic pollutants (Table 1). The PDGs protein database included four sub-databases [phenol 2-monooxygenase (*pmo*), 4-hydroxybenzoate decarboxylase (*hbd*), 4-hydroxybenzoate-CoA ligase (*hbcl*), and 4-hydroxybenzoyl-CoA reductase (*hbcr*)] that were retrieved from the Biocatalysis/Biodegradation Database (The University of Minnesota, http://umbbd.ethz.ch/) (Gao et al., 2010). These four subgroups were also merged and de-duplicated to finally obtained 167 non-redundant protein sequences (Table 1). The BDGs and PDGs databases used in the present study are available for free download at http://web.hku.hk/~zhangt/ZhangT.htm.

2.5. Bioinformatic analysis

A flow chart of the metagenomic analysis of BDGs and PDGs in the AS samples is shown in Fig. 1. All clean reads were assembled using Velvet with Kmer of 51-subunit Kmers. Based on the contigs obtained, MetaGeneMark was used to predict the open reading frames (ORFs). To explore the abundance of BDGs and PDGs, the clean reads, contigs, and ORFs of each AS sample were aligned using BLASTx, BLASTx, and BLASTp (Altschul et al., 1997), respectively, and compared with the two established BDGs and PDGs databases using an *E*-value cut-off of 10^{-5} (Kristiansson et al., 2011), and then, the best hit results were filtered using cut-offs at identity