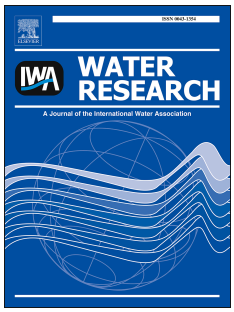
Accepted Manuscript



Biotransformation of nitrogen- and sulfur-containing pollutants during coking wastewater treatment: Correspondence of performance to microbial community functional structure

Dev Raj Joshi, Yu Zhang, Yinxin Gao, Yuan Liu, Min Yang

PII: S0043-1354(17)30409-8

DOI: 10.1016/j.watres.2017.05.045

Reference: WR 12924

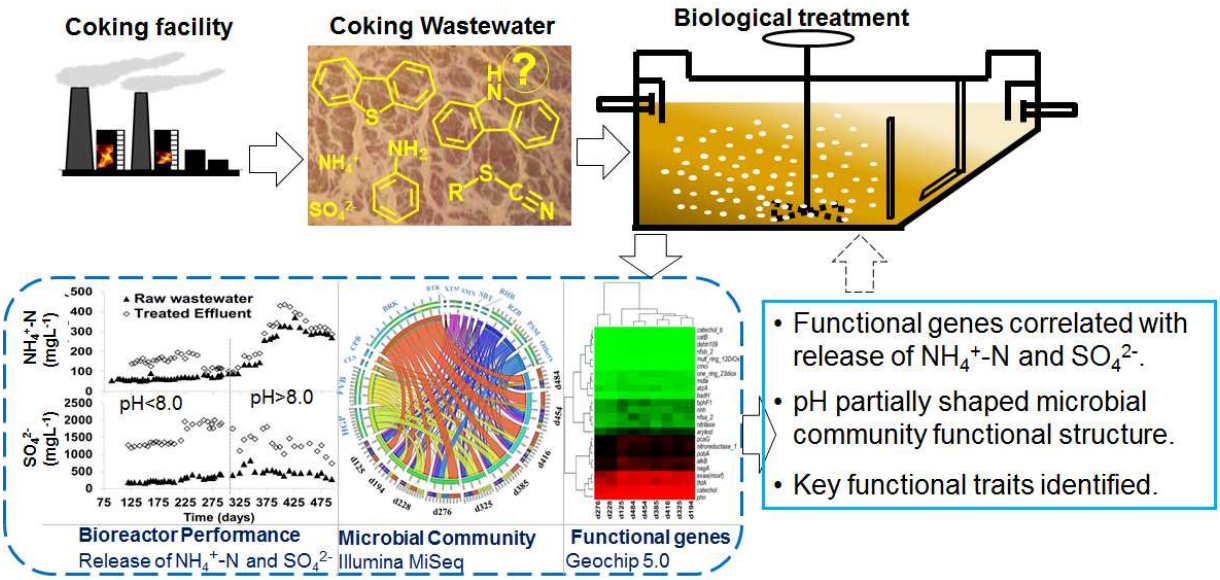
To appear in: *Water Research*

Received Date: 13 March 2017

Revised Date: 18 May 2017

Accepted Date: 21 May 2017

Please cite this article as: Joshi, D.R., Zhang, Y., Gao, Y., Liu, Y., Yang, M., Biotransformation of nitrogen- and sulfur-containing pollutants during coking wastewater treatment: Correspondence of performance to microbial community functional structure, *Water Research* (2017), doi: 10.1016/j.watres.2017.05.045.



Biotransformation of nitrogen- and sulfur-containing pollutants during coking wastewater treatment: Correspondence of performance to microbial community functional structure

Dev Raj Joshi [a,b], Yu Zhang [a,b,*], Yinxin Gao [a,b], Yuan Liu [a], Min Yang [a,b]

[a] State Key Laboratory of Environmental Aquatic Chemistry, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

[b] University of Chinese Academy of Sciences, Beijing 100049, China

***Corresponding author**

Telephone: +86 10 62919883; Fax: +86 10 62923541

Email: zhangyu@rcees.ac.cn (Y. Zhang)

Abstract

Although coking wastewater is generally considered to contain high concentration of nitrogen- and sulfur-containing pollutants, the biotransformation processes of these compounds have not been well understood. Herein, a high throughput functional gene array (GeoChip 5.0) in combination with Illumina MiSeq sequencing of the 16S rRNA gene were used to identify microbial functional traits and their role in biotransformation of nitrogen- and sulfur-containing compounds in a bench-scale aerobic coking wastewater treatment system operated for 488 days. Biotransformation of nitrogen and sulfur-containing pollutants deteriorated when pH of the bioreactor was increased to >8.0, and the microbial community functional structure was significantly associated with pH (Mantels test, $P < 0.05$). The release of ammonia nitrogen and sulfate was correlated with both the taxonomic and functional microbial community structure ($P < 0.05$). Considering the abundance and correlation with the release of ammonia nitrogen and sulfate, aromatic dioxygenases (e.g. *xylXY*, *nagG*), nitrilases (e.g. *nhh*, *nitrilase*), dibenzothiophene oxidase (*DbtAc*), and thiocyanate hydrolase (*scnABC*) were important functional genes for biotransformation of nitrogen- and sulfur-containing pollutants. Functional characterization of taxa and network analysis suggested that *Burkholderiales*, *Actinomycetales*, *Rhizobiales*, *Pseudomonadales*, and *Hydrogenophiliales* (*Thiobacillus)* were key functional taxa. Variance partitioning analysis showed that pH and influent ammonia nitrogen jointly explained 25.9% and 35.5% of variation in organic pollutant degrading genes and microbial community structure, respectively. This study revealed a linkage between microbial community functional structure and the likely biotransformation of nitrogen- and sulfur-containing pollutants, along with a suitable range of pH (7.0–7.5) for stability of the biological system treating coking wastewater.

Key words: Coking wastewater, Nitrogen and Sulfur containing organic compound, functional gene, Taxa-function relationship, microbial network

1. Introduction

Coking wastewater is liquid waste from coke production laden with numerous pollutants, including phenols, polyaromatic hydrocarbons, nitrogen-, sulfur- and oxygen-containing heterocyclics, and acyclic compounds (Liu et al. 2017, Sharma and Philip 2016, Zhang et al. 1998), and can induce toxic and carcinogenic impacts (Dehua et al. 2016, Zhao et al. 2014) on the environment. Increasing environmental awareness coupled with more stringent standards has triggered various industries to challenge themselves in seeking appropriate wastewater treatment technologies (Teh et al. 2016).

Biological treatment of coking wastewater has long been of interest for environmental engineering studies. Most of the identified compounds, including phenols, thiocyanates, cyanides, and polyaromatic hydrocarbons, can be biologically removed (Bai et al. 2011, Jeong and Chung 2006, Li et al. 2003). Recent pyrosequencing analysis has shown that microbial genera, such as *Thiobacillus*, *Comamonas*, *Pseudomonas*, *Thaurea*, *Burkholderia*, and *Trichosporon*, might play an important role in the degradation of phenols, thiocyanates, and cyanides in coking wastewater (Joshi et al. 2016, Ma et al. 2015a, Zhu et al. 2016). However, system performance instability or sudden failures of full-scale applications treating coking wastewater have also been reported (Kim et al. 2009, Vazquez et al. 2006b). The presence of toxic compounds like phenols and cyanides is often speculated as the main reason for the deterioration of treatment performance (Amor et al. 2005, Sharma and Philip 2014). Judging from the high biodegradability of these pollutants (Feng et al. 2015, Marrot et al. 2006, Papadimitriou et al. 2009), however, it is

assumed that some as yet unknown compounds might be more sensitive to environmental conditions and thus cause process failure.

On the other hand, excessive ammonium and sulfate can be generated during biological treatment of coking wastewater (Joshi et al. 2016, Staib and Lant 2007, Vazquez et al. 2006a), suggesting possible biodegradation of nitrogen- and sulfur-containing organic and inorganic compounds. Based on the low chemical oxygen demand to total organic carbon (COD/TOC) ratio (Lim et al. 2003), refractory organic compounds including nitrogen- and sulfur-containing compounds (Huang et al. 2016, Zhang et al. 2013a), may represent a substantial fraction as only nitrogenous compounds constitute approximately 20 - 40% of the organic component in coking wastewater (Li et al. 2003, Meng et al. 2016). Indeed, nitrogen and sulfur-containing compounds are of great environmental importance due to their high toxicity and persistence (Dehua et al. 2016, Jensen et al. 2003) and hence might be critical for the treatment of coking wastewater. So understanding the functional ecology of biotransformation of these pollutants during the treatment process has great practical significance. Biodegradation of nitrogen and sulfur heterocycles via deamination and desulfurization pathways by several bacterial isolates including *Pseudomonas, Burkholderia, Rhodococcus, Sphingomonas, Comamonas* (Gai et al. 2007, Jiang et al. 2016, Tao et al. 2011) and thiocyanates via carbonyl pathway by *Thiobacillus thioparus* (Kim and Katayama 2000, Watts and Moreau 2016) are well studied. However, the knowledge on key microbial taxa and associated functional genes involved in the biotransformation of nitrogen and sulfur-containing pollutants in coking wastewater treatment system is still very limited. In addition, how environmental variables affect the biotransformation of these pollutants and influence the microbial community functional structure are not clearly understood. As an important environmental factor, pH has great impact on the biotransformation (Shen et al. 2015).

Considering high concentration of ammonia in coking wastewater (Zhang et al. 2009), unionized free ammonia could be inhibitory to microbes. Since, pH equilibrates free ammonia and ammonium ion; it may be key factor to maintain the stability in a treatment system by controlling the equilibrium between free and ionized ammonia (Lay-Son and Drakides 2008). However, effect of small shift of pH on microbial community functional structure and consequently, on biotransformation of nitrogen and sulfur pollutants have not yet been evaluated for coking wastewater treatment system.

In this study, a bench-scale activated sludge reactor was used to treat anaerobically pretreated coking wastewater over a period of 488 days, with a focus on the deamination and desulfurization processes. Except for the removal of COD and total phenols, parameters describing the release of ammonia and sulfate with respect to influent COD were used for the evaluation of wastewater treatment performance. Sludge samples from different temporal points during the operation period were taken for phylogenic and functional gene community analysis using Illumina MiSeq sequencing of 16S rRNA genes and functional gene microarray (GeoChip 5.0), respectively. GeoChip 5.0 contains 167,044 distinct probes, covering 395,894 coding sequences from 1593 functional gene families involved in microbial biogeochemical cycling and organic remediation (http://ieg.ou.edu/), and has been extensively employed to analyze the functional gene structure of microbial communities in different environments (Chan et al. 2013, Zhang et al. 2013b). The functional traits of abundant microbial taxa were identified by assigning taxa to functional gene categories involved in the degradation of organic pollutants, as described in GeoChip 5.0 (Chan et al. 2013). Potential bacterial hosts and functional genes associated with the biotransformation of organic pollutants, particularly those containing nitrogen or sulfur, were further explored by network analysis based on the GeoChip and MiSeq data. Lastly, the

contributions of wastewater variables to the microbial community and functional structures were analyzed by variation partitioning analysis (VPA). This study could advance our understanding of biological treatment processes of coking wastewater, and improve the optimization of system operation.

2. Materials and methods

2.1 Coking wastewater treatment and sludge sample collection

Coking wastewater was obtained from a coking facility in Tangshan City, Hebei Province, China, and was treated using a bench-scale bioreactor consisting of anaerobic pretreatment and aerobic treatment (Joshi et al. 2016) for 488 days. The anaerobic pretreatment is described in the supplementary information (Experimental section 1). The aerobic bioreactor was operated with a constant hydraulic retention time (HRT) of 72 h, dissolved oxygen (DO) of 2–4 mg L^{-1}, and temperature of 20–25 °C. The influent wastewater characteristics are given in supplementary information (Table S1). After 300 days, the pH (7.2±0.3) of the aerobic bioreactor was gradually increased up to 8.0 to 9.0 by addition of 0.1 M NaOH solution (Chao et al. 2006).

Composite sludge samples were taken from the aerobic bioreactor at nine temporal points (125, 194, 228, 276, 325, 285, 416, 454, and 484 days) and stored at -80 °C until DNA extraction. In parallel, grab samples of influent and effluent wastewater were collected. COD, TOC, total phenol, total nitrogen, ammonia nitrogen, and sulfate were measured as described previously (Joshi et al. 2016). Sample processing and analytical methods are given in the supplementary information (Experimental section 2). The release of ammonia nitrogen ($d\mathrm{NH_4^+}$–N) and sulfate ($d\mathrm{SO_4^{2-}}$) with respect to influent COD was calculated as follows (1):

$$d\mathrm{NH_4^+}-\mathrm{N/COD\ or,}\ d\mathrm{SO_4^{2-}}/\mathrm{COD} = \frac{[\text{Effluent concentration (mg L}^{-1}) - \text{Influent concentration (mg L}^{-1})\ \text{of ammonia nitrogen } (\mathrm{NH_4^+}-\mathrm{N}) \text{ or sulfate } (\mathrm{SO_4^{2-}})]}{\text{Influent concentration of COD (mg L}^{-1})} \quad (1)$$

Free ammonia was calculated as described previously (Anthonisen et al. 1976) using the following equation:

$$\text{Free ammonia, } NH_3(\text{mg } L^{-1}) = \frac{17}{14} \times \frac{\text{Total ammonia nitrogen (mg } L^{-1}) \times 10^{pH}}{e^{(6{,}344/273+^{\circ}C)} + 10^{pH}} \quad (2)$$

2.2 GeoChip 5.0 analysis

Total community DNA for GeoChip analysis was extracted using a PowerSoil® DNA Isolation Kit (Mo Bio Laboratories, USA; Catalog no. 12888-100) (Supplementary Information experimental section 3). Each DNA sample was prepared by pooling independent extracts from three replicate sludge samples collected at different time (8.00 AM, 1:00 PM, 5:00 PM) of same day. DNA sample (2.0 μg) was labeled using cyanine (Cy3) dye with random primers and the Klenow fragment of DNA polymerase I (IMER Inc., USA), then purified (Qiagen QIAquick Kit, Germany) and dried using a SpeedVac at Vacuum Level 5.1 (ThermoSavent, USA) for 2 h at 45 °C (Nostrand et al. 2016). The labeled DNA was re-suspended in hybridization buffer (Oligo aCGH Hybridization Kit, large, catalog number 5188-5380, Agilent Technologies Inc., USA) and denatured at 95 °C for 3 min, with the array then hybridized at 67 °C for 24 h at a rotation speed of 20 rpm in the chamber. The GeoChip 5.0 (Agilent Technologies Inc., USA) 180 k array was applied for microarray hybridization. After hybridization, arrays were scanned with a SureScan Microarray Scanner (Agilent Technologies Inc., USA) in red and green channels (lasers with excitation wavelengths at 640 and 532 nm, respectively), with 3 μm resolution, 20-bit Tiff dynamic range (>10 5), and 100% photomultiplier tube sensitivity for both channels. The raw data were extracted from the scanned images using the Feature Extraction program (Agilent Technologies Inc., CA, USA). GeoChip data normalization and quality filtering were performed as previously described (Nostrand et al. 2016), using the microarray data manager from the

Institute for Environmental Genomics, University of Oklahoma (USA) (http://ieg.ou.edu/entrance.html). Before statistical analysis, logarithmic transformation (log_{10}) was performed, and the signals of all spots were transferred into relative abundances. All microarray hybridization data are available at the Institute for Environmental Genomics, University of Oklahoma (http://ieg.ou.edu).

2.3 Illumina MiSeq sequencing

Total DNA for Illumina MiSeq sequencing was extracted using the FastDNA® SPIN Kit for soil (Qbiogene, Solon, OH, USA) as described in supplementary information (experimental section 3). DNA extracts of replicate sludge samples were pooled together as mentioned in GeoChip analysis. The hyper-variable V4 region of the bacterial 16S rRNA gene was amplified using forward primer 515 F (5′-GTGCCAGCMGCCGCGGTAA-3′) and reverse primer 806 R (5′-GGACTACHVGGGTTCTAAT-3′) containing a variable 12 bp barcode sequence (Caporaso et al. 2012). Polymerase chain reaction (PCR) amplification was performed with 25 μL of PCR mixture, constituting 0.1 μL of AccuPrime High Fidelity Taq Polymerase, 1 μL of each primer (10 μM), 2.5 μL of 10×AccuPrime PCR buffer II (Invitrogen, USA), and 1 μL of template DNA. A Veriti96-Well Thermal Cycler (Applied Biosystems, USA) was applied for amplification using the following thermal cycling conditions: initial pre-denaturation at 94 °C for 1 min, 35 denaturation cycles at 94 °C for 20 s, annealing at 53 °C for 25 s, elongation at 68 °C for 45 s, and a final extension at 68 °C for 10 min. Each sample was amplified in triplicate, and PCR products were pooled and purified using a QIAquick Gel Extraction Kit (Qiagen, Germany). Purified PCR products were quantified with PicoGreen. The purified mixture was diluted and denatured to obtain a sample DNA library, as described in the MiSeq Reagent Kit Preparation Guide (Illumina, USA), and mixed with an equi-volume of 8 *p*M PhiX (Illumina, San Diego, CA,

USA). The DNA sample mixture was loaded with read 1, read 2, and index sequencing primers on a 300-cycle (2×150 paired ends) kit, and run on a MiSeq.

As raw sequences were obtained, primers and spacers were trimmed out. The paired-end reads were overlapped to assemble the V4 tag sequences using FLASH (Magoc and Salzberg 2011). Low quality fragments and sequences shorter than 240 bp were removed. The chimeras were checked and filtered using UCHIME (Edgar et al. 2011). The OTUs were classified using UCLUST (Edgar 2010) at a 97% similarity level. Taxonomic assignment was performed using the RDP classifier (Cole et al. 2009) (COLE) (http://rdp.cme.msu.edu). The raw sequencing data were submitted to the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra/) under accession numbers SRR4253894 to SRR4253902.

2.4 Network analysis

The possible co-occurrence between microbial taxa and organic pollutant degrading genes was examined by random correlation matrix-based microbial network analysis (Deng et al. 2012, Tian et al. 2016). The bacterial OTUs obtained by Illumina MiSeq sequencing were combined with functional genes obtained by GeoChip 5.0 across all nine samples. For constructing the network, 269 items (175 OTUs and 94 organic pollutant degrading genes) that existed in at least five samples were combined and a correlation matrix was created. A similarity matrix was then obtained by taking the absolute values of the correlation matrix. The most suitable lowest threshold (Deng et al. 2012) was selected to obtain the Poisson distribution of the calculated eigenvalues. Online analysis pipeline (http://ieg2.ou.edu/MENA) was used for network analysis, and Cytoscape 3.3.0 software (http://cytoscape.org/) was applied to visualize the network graph.

197 *2.5 Statistical analysis*

198 Prior to statistical analysis, wastewater quality (Table S1) and bioreactor performance
199 (Table S2) variables were standardized by dividing the difference between the sample values and
200 the mean value of all samples by the standard deviation. Microbial community diversity indices,
201 that is, Shannon-Weiner (H) and Simpson were calculated using R 3.2.5 (http://www.r-
202 project.org/) with the vegan package. The relationships among microbial OTU and functional
203 gene abundances with wastewater variables (influent wastewater, bioreactor performance, and
204 pH) were examined by Mantel tests.

205 Canonical correspondence analysis (CCA) was carried out to discern possible
206 associations among the microbial (phylogenetic and functional) community and wastewater
207 variables, and partial CCA-based VPA was used to analyze the contributions of the wastewater
208 variables in microbial community structures. Mantel tests, CCA, and partial CCA were
209 performed using R 3.2.5 (http://www.r-project.org/) with the vegan and stats packages.
210 Significance tests were conducted by Monte Carlo permutation (999 times). P values < 0.05 were
211 regarded as significant.

212

213 **3. Results**

214 *3.1 Bioreactor performances*

215 Anaerobically pretreated coking wastewater with an average COD of 1978.8 mg L^{-1} (min.
216 1423.5 to max. 2817.0) (Table S1) was treated for 488 days by an aerobic bioreactor under
217 constant operational conditions, except that the sludge pH was shifted from 7.2 ± 0.3 to > 8.0 (up
218 to 9.0) from 300 days onwards. As pH was increased, the average COD and TOC removal rates

219 decreased from 72.6 ± 2.5 to 68.5 ± 6.7% and 77.6 ± 3.1 to 72.6 ± 5.3%, respectively; however,
220 total phenol removal was consistently high (99.8 ± 0.1%) across the operation period (Fig. 1a
221 and Table S2). Overall average ammonia nitrogen and sulfate concentrations increased after
222 aerobic treatment; however, the increase rate reduced from 235.6 ± 36.7 to 128.8 ± 30.1% and
223 624.5 ± 47.5 to 311.5 ± 37.2%, respectively when pH was increased to >8.0 (Table S2). Our
224 results indicated that the system was comparatively steady when pH was in a range of 7.1 to 7.6
225 (Fig. 1 and Table S2). The release of ammonia nitrogen ($d NH_4^+$–N/COD, range 0.01–0.06) and
226 sulfate ($d SO_4^{2-}$/COD, range 0.25–0.93) with respect to influent COD showed a characteristic
227 decreasing pattern with increasing pH and free ammonia (NH_3), as fitted by following
228 polynomial equations (Fig. 1b):

229 For $d NH_4^+$–N/COD ($R^2 = 0.50$); $y = - 0.009 x^2 + 0.124 x - 0.396$ (3)

230 For $d SO_4^{2-}$/COD ($R^2 = 0.43$); $y = - 0.209 x^2 + 3.149 x - 11.13$ (4)

231 For NH_3 ($R^2 = 0.81$); $y = 47.51 x^2 + 714.9 x + 2685.8$ (5)

232 3.2 *Microbial phylogenetic community structure revealed by MiSeq sequencing*

233 Illumina MiSeq sequencing of the 16S rRNA gene revealed 10,582–19,610 sequence
234 reads with a total of 298 bacterial OTUs. Alpha diversity (Shannon-Weiner index) and
235 percentage of unique bacterial OTUs ranged from 2.74 to 3.42 and 0.64 to 4.89, respectively,
236 among the samples (Table S3). In total, 16 phyla were obtained including *Proteobacteria*
237 (58.95–87.75%), *Bacteroidetes* (3.87–23.88%), and *Actinobacteria* (0.98–3.70%). At the genera
238 level, unclassified genera belonging to order *Burkholderiales* (mostly *Comamonadaceae*) and
239 *Thiobacillus* represented the major core community, accounting for 24.8–53.1% of total bacterial
240 OTUs (Fig. S1).

The microbial community structure was correlated with wastewater quality and operational conditions (Table 1), revealing significant associations with influent ammonia nitrogen (r = 0.571, P = 0.006), total nitrogen (r = 0.621, P = 0.001), and pH (r = 0.534, P = 0.005). To discern the possible key functions of the bacterial community, correlation between bioreactor performance and phylogenetic community structure was also analyzed. The results revealed that the microbial community was significantly correlated in combined with the release of ammonia nitrogen ($d\mathrm{NH_4^+}$–N/COD) and sulfate ($d\mathrm{SO_4^{2-}}$/COD) (r = 0.535, P = 0.002). However, no significant correlation was observed with COD or phenol removal efficiency.

3.3 Microbial community functional structure and key functional genes revealed by GeoChip analysis

The GeoChip-based microarray detected a total of 67,395 functional genes (1047 gene categories) from the nine samples, with alpha diversity (Shannon-Weiner index) ranging from 10.97 to 11.02 (Table S4). On average, 91.6% of the functional genes were derived from bacteria, 2.4% from archaea, 5.2% from eukaryote, and 0.9% from viruses. Bacterial functional genes were derived mainly from *Proteobacteria* (57.7%), *Actinobacteria* (19.9%), *Firmicutes* (8.2%), *Bacteroidetes* (2.9%), and *Cyanobacteria* (2.1%).

In this study, functional genes involved in different biological processes, including nutrient carbon cycling (15.8%), nitrogen cycling (4.3%), sulfur cycling (2.8%), organic pollutant removal (8.8%), metal homeostasis (26.6%), stress response (15.5%), virulence (15.5%), phosphorus cycling (2.1%), and secondary metabolism (2.8%), were detected (Fig. S2). Table S5 demonstrates the frequency distribution of functional genes involved in all bioprocesses. The nutrient cycling genes involved in carbon (10781 genes), nitrogen (2910 genes), sulfur (1950 genes), and phosphorus (1438 genes) cycling, including those involved in degradation of

264 complex carbon compounds (e.g., amylase (*amyA*), *chitinase*, *acetylglucosaminidase*, *cellobiase*,
265 arabinofuranosidase (*ara*), *xylanase*), ammonification (glutamate dehydrogenase (*gdh)* and
266 urease (*ureC*)), and sulfur/sulfide oxidation (sulfide-quinone reductase (*sqr*), flavocytochrome
267 sulfide dehydrogenase (*fccAB*), and sulfur oxidase (*soxABCYV*)), were detected.

268 A total of 5867 functional genes belonging to 99 gene families involved in organic
269 pollutant removal (Table S5), degradation of aromatics (average relative abundance 6.3% of all
270 genes), xenobiotic (herbicide related) compounds (1.1%), chlorinated solvents (1.1%), and other
271 hydrocarbons (0.43%) were detected. Most abundant aromatic degrading genes included
272 intradiol ring-cleavage dioxygenase genes, *catechol* (0.58%), *one_ring_12diox* (0.22%), and
273 *mult_ring_12DiOx* (0.19%); xenobiotic related compound degrading gene, *phn* (0.41%);
274 chlorinated aromatic (containing amine) degrading gene, *tfdA* (0.36%); aromatic carboxylic acid
275 degrading gene, *nagG* (0.36%); BTEX (benzene, toluene, ethylbenzene, and xylene) compound
276 degrading gene, *catB* (0.2%); nitro-aromatics degradation genes, *nitroreductase* (0.45%), *nhh*,
277 and *nsfA* (each 0.25%); and alkane monooxygenase, *alkB* (0.32%) (Fig. 2).

278 The Mantel test revealed that pH of the sludge was significantly correlated with the
279 whole functional community structure (r = 0.31, P = 0.038) and the organic pollutant degrading
280 functional community structure (r = 0.319, P = 0.040) (Table 1). However, no significant
281 correlation was observed between the whole functional community structures and influent
282 wastewater quality (P > 0.05). Table S6 demonstrates correlation analysis between the
283 abundances of individual gene families and wastewater variables and operational pH (of sludge).
284 Majority of genes (18 families) correlated significantly with pH of the sludge. Apart from whole
285 functional gene structure, 9 and 5 genes independently correlated with influent COD and influent
286 ammonia nitrogen, respectively.

The possible association between microbial community functional genes and treatment performance was also evaluated by correlation analysis using Mantel tests. The organic pollutant degrading gene structure was strongly correlated with $d\mathrm{NH_4^+}$–N/COD and $d\mathrm{SO_4^{2-}}$/COD (r = 0.489, P = 0.002) (Table 1), but was not significantly (P > 0.05) related to either COD or total phenol removal efficiencies. Furthermore, correlation analysis between the relative abundances of individual gene families (organic pollutant removal) and $d\mathrm{NH_4^+}$–N/COD and $d\mathrm{SO_4^{2-}}$/COD revealed that 28 gene families (Table 2), which could cleave aromatic rings and degrade nitro- and heterocyclic-aromatics showed significant correlation. Strong correlation was demonstrated by toluate 1,2-dioxygenase, *xyl*XY (r = 0.810, P = 0.001); methylamine dehydrogenase, *mauAB* (r = 0.649, P = 0.001); salicylate hydroxylase, *nagG* (r = 0.469, P = 0.004); cytochrome P450 monooxygenase, *p450aro*, (r = 0.495, P = 0.008); nitrile hydratase, *nhh* (r = 0.527, P = 0.006); and cyanuric acid amidohydrolase, *atzD* (r = 0.539, P = 0.004) with $d\mathrm{NH_4^+}$–N/COD and $d\mathrm{SO_4^{2-}}$/COD. Importantly, nitrogen-containing organics degrading genes nitroreductases (*nfsA/B*, *nitroreductase*) and aniline dioxygenases (*tfdA*, *tdnB*), and sulfur-containing heterocyclics degrading gene dibenzothiophene oxidase (*DbtAc*) were also significantly correlated (P < 0.05) with $d\mathrm{NH_4^+}$–N/COD and $d\mathrm{SO_4^{2-}}$/COD. Although thiocyanate hydrolase (*scnABC*) was not directly correlated with $d\mathrm{NH_4^+}$–N/COD and $d\mathrm{SO_4^{2-}}$/COD, GeoChip analysis detected 23 of *scnABC* genes (Fig. S3) which were mostly derived from *Thiobacillus thioparus*. Table S7 lists the most abundant hosts of the important organic compound degrading genes suggesting that bacteria belonging to genera *Burkholderia*, including others were important functional trait.

3.4 Organic pollutant degrading microbial taxa revealed by taxa-function and network analyses

Functional traits of abundant microbial taxa were identified by assigning taxa to the functional gene categories involved in the degradation of organic pollutants (Chan et al. 2013).

Results revealed the presence of genes indicating organic pollutant removal in 25 phyla, mostly bacterial ones. Among them, 18, 14, 13, 6, and 16 phyla contained genes for degrading nitro-aromatic compounds, aromatic carboxylic acids, BTEX related compounds, chlorinated aromatics, and other hydrocarbons, respectively. Since most functional genes derived from *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, *Actinobacteria*, and unclassified bacteria demonstrated the highest signal intensities for the degradation of organic pollutants (Fig. S4), we assigned taxa-function relationships at the order level (Fig. 3). We found that aromatic carboxylic acids and nitro-aromatic compound degrading genes were derived from *Burkholderiales* (relative abundance of genes 3.2% and 1.8%, respectively) and *Actinomycetales* (2.9% and 1.9%, respectively), whereas xenobiotic compound degrading genes were abundant in *Rhizobiales* (2.3%) and *Rhodobacteriales* (2.0%). Other aromatic compound degrading genes were mostly contained in *Actinomycetales* (6.7%), *Burkholderiales* (5.7%), *Rhizobiales* (3.9%), and *Pseudomonadales* (3.0%). Other hydrocarbon (mostly aliphatic) degrading functional genes were found in unclassified *Alphaproteobacteria* (2.5%), unclassified *Gammaproteobacteria* (2.8%), and *Aeromonadales* (1.0%).

The concurrences between taxa and functional genes, based on the hypothesis that the abundance pattern of genes is similar to that of the host taxa, were further verified using a RMT-based network approach (Deng et al. 2012). Fig. 4 visualizes the first-ranked cluster of positive edges (links) connected among nodes of functional genes (GeoChip) and bacterial genera (Illumina MiSeq). The network analysis clearly showed that *Proteobacteria* (22 nodes) and *Actinobacteria* (7 nodes) were dominantly linked with functional genes (Table S8). *Burkholderia* (3 nodes) co-occurred with sulfur heterocyclic compound degrading gene, *DbtAc* (Dibenzothiophene oxidase), aromatic ring hydroxylating genes, *pchcf* (cresol hydroxylase),

xlnD, and *mult_ring_12Diox*, and nitrogen containing organics degrading gene, *nitrilase* (Fig. 4). Genera belonging to *Alphaprobacteria viz. Rhizobium* (1 node), *Sphingobium* (2 nodes), and *Paracoccus* (2 nodes) were also linked with functional genes involved in the degradation of similar compounds.

3.5 Contribution of wastewater variables in shaping microbial community and functional gene structure

Canonical correspondence analysis (CCA) was applied to demonstrate the links between environmental variables and microbial community functional and phylogenetic structure. Based on automatic forward selection and variance inflation factors with 999 Monte Carlo permutations, four wastewater quality parameters (influent COD, total nitrogen, NH_4^+–N, and sulfate) and an operational parameter (pH) were included in the CCA bi-plot ($P < 0.05$) (Fig. S5). The first axis was negatively correlated with COD, pH, and NH_4^+–N and the second axis was positively correlated with influent COD. Both axes combined explained 34.4% of organic pollutant degrading gene diversity (Fig. S5a). Contributions of COD (C), NH_4^+–N (N), and pH (P) on organic pollutant degrading functional structure and phylogenetic microbial community structure were estimated by VPA. A total of 43.9% and 56.9% of the variation in organic pollutant degrading genes and phylogenetic microbial community diversity, respectively (Fig. 5), were explained by three environmental variables ($p < 0.05$). Influent COD, NH_4^+–N, and pH independently explained 16.96, 10.30, and 10.69% of total variations observed in organic pollutant degrading genes, and 16.5, 10.34, and 10.17% of total variations in the phylogenetic microbial community, respectively. Notably, combined NH_4^+–N and pH explained 25.9% and 35.5% of variation of organic pollutant degrading genes and the phylogenetic microbial

community, respectively (Fig. 5). Similar results were obtained by VPA of the variation in the whole microbial functional gene structure by the same variables (Fig. S6).

4. Discussion

Coking wastewater has variety of organic and inorganic pollutants, most of which are very toxic and harmful to human health and environment (Dehua et al. 2016). For instance, PAHs including heterocyclic compounds are known carcinogens (Zhang et al. 2013a), while many other pollutants including phenols and cyanides could inhibit microbial processes like nitrification (Li et al. 2010). Treatment technology and process operation have been studied for quite a long time, but limited advancement were achieved (Pal and Kumar 2014). Because of the significant presence and environmental consequences of the nitrogen- and sulfur-containing pollutants, understanding their biotransformation during treatment process has important practical implications for harmless disposal of coking wastewater. Herein, we applied high throughput molecular methods to investigate a linkage between possible biotransformation of nitrogen- and sulfur-containing pollutants with microbial community functional structure in coking wastewater treatment system.

In this study, gradual increase in pH from 8.0 to 9.0, reduced the release of ammonia nitrogen ($d\mathrm{NH_4^+}$–N/COD) and sulfate ($d\mathrm{SO_4^{2-}}$/COD) (Fig. 1), suggesting that the transformation of nitrogen- and sulfur-containing pollutants is one of the key factors in the biological treatment of coking wastewater. Normally, pH range of 6.5 to 8.5 is acceptable for biological treatment of wastewaters (Eckenfelder 2000); however, in this study we found that $d\mathrm{NH_4^+}$–N/COD and $d\mathrm{SO_4^{2-}}$/COD decreased when pH increased to > 8.0. In agreement to this result, Shen et al. also found that pH range of 7.0 to 8.0 was the most suitable for the degradation of petroleum hydrocarbons (Shen et al. 2015). As pH controls the equilibrium between free ammonia and

ammonium ions, the toxicity caused by un-ionized ammonia (Lay-Son and Drakides 2008) might be an important reason for deterioration of biotransformation of nitrogen and sulfur containing pollutants. Of note, the free ammonia concentration, in the bioreactor was drastically increased when pH was increased to >8.0 as 'free ammonia' is the function of pH and temperature (Anthonisen et al. 1976). Besides, increased free ammonia (>2.0 mgL^{-1}) inhibits the biodegradation of thiocyanate (Lay-Son and Drakides 2008). This result indicated that free ammonia might have exerted toxic effect on microbial community and their functions during the treatment process. Consequently, we found that pH was significantly correlated with both the whole microbial community (r = 0.442, P = 0.005) and organic pollutant degrading gene structure (r = 0.319, P = 0.04) (Table 1, Fig. S6). In addition, pH in combination with influent ammonia nitrogen was the main factor in shaping the microbial community and functional gene structure (35.5 and 25.9% of variations, respectively) (Fig. 5). Our results clearly suggested that maintenance of the pH between 7.0–8.0 and reducing free ammonia below toxic level could be crucial for removing the nitrogen- and sulfur-containing pollutants in coking wastewater treatment system. However, in contrast, the removal of total phenol was consistently high throughout the operational period (Fig. 1). This might be attributed to certain phenol degrading bacteria which may tolerate slight pH shift towards alkaline condition (Gallizia et al. 2003).

We observed a significant correlation of combined $d{NH_4}^+$–N/COD and $d{SO_4}^{2-}$/COD with taxonomic microbial community (r = 0.535, P = 0.002), organic pollutant degrading genes (r =0.489, P =0.002) (Table 1) and individual gene families having potential to cleave broad spectrum aromatic rings including heterocyclic aromatics (Table 2). BTEX-related compound degrading dioxygenase genes, for instance, *xyl*XY, *nagG*, *p450aro*, *catechol*, *catB*, and *one_ring_12diox*, (Table 2) might have expressed to the enzymes related to the peripheral or

central pathways for degradation of aromatic pollutants, including PAHs (Sierra-Garcia et al. 2014, Suenaga et al. 2014). Since single aromatic ring hydroxylating dioxygenase enzyme may have a wide range of substrate specificity (Fuchs et al. 2011, Suenaga et al. 2009), these genes allow the degradation of multiple aromatic pollutants. Some functional genes including *nhh*, *atzD, tfdA*, and *tdnB*, which are responsible for degrading nitrogen heterocycles (Fetzner 1998, Suenaga et al. 2009), and dibenzothiophene oxidase gene (*DbtAc*) for oxidation of dibenzothiophene (Andreolli et al. 2011) were also correlated with $d\mathrm{NH_4^+}$–N/COD and $d\mathrm{SO_4^{2-}}$/COD. Additionally, detection of plenty of thiocyanate hydrolase (*scnABC*) genes (Fig. S4) was indicative of their role in conversion of thiocyanate compounds into ammonia and sulfate (Kim and Katayama 2000, Watts and Moreau 2016) as previously we found 98.2% removal of thiocyanates during coking wastewater treatment (Joshi et al. 2016). However, there was no statistical correlation between abundances of *ScnABC* genes and $d\mathrm{NH_4^+}$–N/COD and/or $d\mathrm{SO_4^{2-}}$/COD. In overall, our results are suggestive that the above functional genes in the aerobic sludge were linked to the bioreactor performance, particularly to the biotransformation of nitrogen- and sulfur-containing pollutants possibly via deamination and desulfurization pathways.

Taxonomic diversity obtained from 16S rRNA gene sequencing (Fig. S1) of this study corroborated previous molecular surveys of coking wastewater treatment plants (Joshi et al. 2016, Ma et al. 2015a, Zhu et al. 2016), showing a unique differences from those commonly occurring in municipal wastewater treatment plants (Wang et al. 2012). Most abundant bacteria *Comamonas* and *Thiobacillus* along with *Burkholderia* are characteristics of phenolic wastewaters particularly, coking wastewater treatment sludge and are commonly appreciated for biodegradation of phenol, thiocyanate and various nitrogen- and sulfur-containing PAHs like carbazole, dibenzothiophene (Felföldi et al. 2010, Jiang et al. 2016, Ma et al. 2015b). However,

direct relations between these taxa and their functions for removal of organic pollutants have not yet been clearly demonstrated in coking wastewater treatment systems.

In this study, taxa-function analysis revealed that majority of aromatic pollutants, including nitroaromatics, aromatic carboxylic acids, BTEX and xenobiotic related compounds could be degraded mainly by *Burkholderiales, Actinomycetales, Rhizobiales, Pseudomonadales* (Fig. 3). Liang et al. found that *Actinobacteria* (*Rodococcus sp.*, *Mycobacterium sp.*, *Nocardioides sp.*, etc), *Burkholderia sp.*, and *Pseudomonas sp.* most abundant members of PAHs degrading community in oil contaminated soil by GeoChip analysis (Liang et al. 2011). Previously, different isolated strains belonging to these taxa have been documented for their potential to degrade many of nitrogen-, sulfur- and oxygen-containing heterocyclic pollutants (Seo et al. 2009, Xu et al. 2006). This finding was further supported by network analysis based on combined data of 16s rRNA gene sequencing by Illumina Miseq and functional genes by GeoChip5.0, which showed a correlation based co-occurrence of mainly *Proteobacteria* and *Actiobacteria* with aromatic pollutant removal genes (Fig. 4). The positive edge between *Burkholderia* and *DbtAC* gene was specifically notable because this gene is actually derived from *Burkholderia sp.* DBT1 (Andreolli et al. 2011). While *Thiobacillus* (order *Hydrogenophiliales*) was not regarded as a key taxa in taxa-function analysis, its taxonomic abundance (Fig. S1) and high signal intensities of *scnABC* gene (Fig. S4) should not be overlooked. So, *Thiobacillus* was also considered as important functional taxa in this study. The linkage between organic pollutant degrading genes with their phylogenetic identity revealed how these unique microbial communities assemble due to functional adaptation in coking wastewater sludge. This is particularly important because of existence of complex and toxic compounds,

446 which may challenge the optimal growth of functionally important microorganisms and inhibit
447 removal efficiency of pollutants during the treatment of coking wastewater.

448 Understanding the core functional taxa is valuable in upgrading a sound process and
449 operational strategy for effective biological treatment of coking wastewater. So adoption of new
450 strategies would be useful for the removal of pollutants with a focus on key functional taxa and
451 their functional genes. Since organic pollutant degrading genes of *Burkholderia, Actinomycetes,*
452 *Pseudomaonas*, *Thiobacillus* etc. were correlated with the likely transformation of nitrogen- and
453 sulfur-containing compounds, it may be possible to fortify their role in the treatment process. At
454 the same time, realizing the crucial impact of pH in the transformation of nitrogen- and sulfur-
455 ontaining pollutants, precise maintenance of pH may greatly contribute to achieve optimum
456 bioreactor performance and the system stability during aerobic treatment of coking wastewater.

457 Given the high functional potential analyzed at the DNA level, it should be noted that the
458 biotransformation potential might have been overestimated in this study. To validate the results
459 from this study, additional in-depth analyses applying metatranscriptomic and metaproteomic
460 tool are needed. Besides, pH and ammonia nitrogen in combination only explained 25.9% of
461 functional gene variations, and therefore other important environmental variables should be
462 further investigated.

463 **5. Conclusion**

464 The biotransformation of nitrogen- and sulfur-containing pollutants was linked with
465 microbial community functional structure in a long run aerobic coking wastewater treatment
466 bioreactor. Following specific conclusions were made from this study:

- The likely biotransformation of nitrogen- and sulfur-containing pollutants was decreased characteristically when pH of the bioreactor was increased to >8.0.
- The microbial community, functional structure and organic pollutant degrading genes were correlated (Mantel test, $P < 0.05$) with pH.
- Considering their significant presence and correlation with the release of ammonia nitrogen and sulfate, aromatic dioxygenases (e.g. *xylXY*, *nagG*), nitrilases (e.g. *nhh*, *nitrilase*), dibenzothiophene oxidase (*DbtAc*), and thiocyanate hydrolase (*scnABC*) were important functional genes for biotransformation of nitrogen- and sulfur-containing pollutants.
- Functional characterization revealed that *Burkholderiales*, *Actinomycetales*, *Rhizobiales*, *Pseudomonadales*, and *Hydrogenophiliales* (*Thiobacillus*) were key functional taxa for degradation of pollutants.
- The microbial community functional structure was significantly associated (Mantel test, $P < 0.05$) with pH. Two parameters, pH (7.0–9.0) and ammonia nitrogen jointly explained 25.9 and 35.5% of variations in organic pollutant degrading genes and microbial community structure, respectively.

Appendix A. Supplementary data

Additional experimental details and data are presented in the Supplementary Information sections.

Acknowledgements

This study was supported by the National Natural Scientific Foundation of China (21437005) and the State Hitech Research and Development Project of the Ministry of Science and Technology, Peoples Republic of China (2012AA063401). This project was also supported

by a special fund of the State Key Joint Laboratory of Environmental Simulation and Pollution Control (15L03ESPC).

References

Amor, L., Eiroa, M., Kennes, C. and Veiga, M.C. (2005) Phenol biodegradation and its effect on the nitrification process. Water Res. 39(13), 2915-2920. Doi: http://dx.doi.org/10.1016/j.watres.2005.05.019

Andreolli, M., Lampis, S., Zenaro, E., Salkinoja-Salonen, M. and Vallini, G. (2011) *Burkholderia fungorum* DBT1: a promising bacterial strain for bioremediation of PAHs-contaminated soils. FEMS Microbiol. Lett. 319(1), 11-18. Doi: http://dx.doi.org/10.1111/j.1574-6968.2011.02259.x

Anthonisen, A.C., Loehr, R.C., Prakasam, T.B.S. and Srinath, E.G. (1976) Inhibition of nitrification by ammonia and nitrous acid. J. Water Pollut. Control Fed. 48(5), 835-852. Doi: http://dx.doi.org/10.2307/25038971

Bai, Y., Sun, Q., Sun, R., Wen, D. and Tang, X. (2011) Bioaugmentation and adsorption treatment of coking wastewater containing pyridine and quinoline using zeolite-biological aerated filters. Environ. Sci. Tech. 45(5), 1940-1948. Doi: http://dx.doi.org/10.1021/es103150v

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G. and Knight, R. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6, 1621-1624.Doi: http://dx.doi.org/10.1038/ismej.2012.8

Chan, Y., Nostrand, J.D.V., Zhou, J., Pointing, S.B. and Farrell, R.L. (2013) Functional ecology of an Antarctic dry valley. PNAS 110(22), 8990-8995. Doi: http://dx.doi.org/10.1073/pnas.1300643110

515 Chao, Y.M., Tseng, I.C. and Chang, J.S. (2006) Mechanism for sludge acidification in aerobic
516 treatment of coking wastewater. J. Hazard. Mater. B137, 1781-1787. Doi:
517 http://dx.doi.org/10.1016/j.jhazmat.2006.05.024

518 Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S.,
519 McGarrell, D.M., Marsh, T., Garrity, G.M. and Tiedje, J.M. (2009) The ribosomal
520 database project: improved alignments and new tools for rRNA analysis. Nucleic Acids
521 Res. 37, D141-D145. Doi: http://dx.doi.org/10.1093/nar/gkn879

522 Dehua, M., Cong, L., Xiaobiao, Z., Rui, L. and Lujun, C. (2016) Acute toxicity and chemical
523 evaluation of coking wastewater under biological and advanced physicochemical
524 treatment processes. Environ. Sci. Pollut. Res. Int. 23(18), 18343-18352. Doi:
525 http://dx.doi.org/10.1007/s11356-016-6882-z

526 Deng, Y., Jiang, Y.H., Yang, Y., He, Z., Luo, F. and Zhou, J. (2012) Molecular ecological
527 network analyses. BMC Bioinformatics 13, 113. Doi: DOI:
528 http://dx.doi.org/10.1186/1471-2105-13-113

529 Eckenfelder, W.W. (2000) Industrial water pollution control, 3rd ed. McGraw-Hill
530 Science/Engineering/Math, USA. Pp246-247. ISBN-13: 978-0070393646

531 Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST.
532 Bioinformatics 26(19), 2460-2461. Doi: http://dx.doi.org/10.1093/bioinformatics/btq461

533 Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. and Knight, R. (2011) UCHIME improves
534 sensitivity and speed of chimera detection. Bioinformatics 27(16), 2194-2200. Doi:
535 http://dx.doi.org/10.1093/bioinformatics/btr381

536 Felföldi, T., Székely, A.J., Gorál, R., Barkács, K., Scheirich, G., András, J., Rácz, A. and
537 Márialigeti, K. (2010) Polyphasic bacterial community analysis of an aerobic activated
538 sludge removing phenols and thiocyanate from coke plant effluent. Bioresour. Technol.
539 101(10), 3406-3414. Doi: http://dx.doi.org/10.1016/j.biortech.2009.12.053

Feng, C., Huang, L., Yu, H., Yi, X. and Wei, C. (2015) Simultaneous phenol removal, nitrification and denitrification using microbial fuel cell technology. Water Res. 76, 160-170. Doi: http://dx.doi.org/10.1016/j.watres.2015.03.001

Fetzner, S. (1998) Bacterial degradation of pyridine, indole, quinoline and their derivatives under different redox conditions. Appl. Microbiol. Biotech. 49, 237-250. http://dx.doi.org/10.1007/s002530051164

Fuchs, G., Boll, M. and Heider, J. (2011) Microbial degradation of aromatic compounds — from one strategy to four. Nat. Rev. Microbiol.9, 803-816. Doi: http://dx.doi.org/10.1038/nrmicro2652

Gai, Z., Yu, B., Li, L., Wang, Y., Ma, C., Feng, J., Deng, Z. and Xu, P. (2007) Cometabolic degradation of dibenzofuran and dibenzothiophene by a newly isolated carbazole-degrading *Sphingomonas* sp. strain. Appl. Environ. Microbiol. 73(9), 2832-2838. Doi: http:// dx.doi.org/10.1128/AEM.02704-06

Gallizia, I., McClean, S. and Banat, I.M. (2003) Bacterial biodegradation of phenol and 2,4-dichlorophenol. J. Chem. Technol. Biotechnol., 78, 959-963. Doi: http://dx.doi.org/10.1002/jctb.890

Huang, Y., Hou, X., Liu, S. and Ni, J. (2016) Correspondence analysis of bio-refractory compounds degradation and microbiological community distribution in anaerobic filter for coking wastewater treatment. Chem. Eng. J. 304, 864-872. Doi: http://dx.doi.org/10.1016/j.cej.2016.05.142

Jensen, A.M., Finster, K.W. and Karlson, U. (2003) Degradation of carbazole, dibenzothiophene, and dibenzofuran at low temperature by *Pseudomonas* sp. strain C3211. Environ. Toxicol. Chem. 22(4), 730-735. Doi: http:// dx.doi.org/10.1002 /etc.5620220408

Jeong, Y.S. and Chung, J.S. (2006) Simultaneous removal of COD, thiocyanate, cyanide and nitrogen from coal process wastewater using fluidized biofilm process. Process Biochem. 41, 1141-1147. Doi: http:// dx.doi.org/10.1016/j.procbio.2005.12.010

566 Jiang, B., Tan, L., Ning, S. and Shi, S. (2016) A novel integration system of magnetically
567 immobilized cells and a pair of graphite plate-stainless iron mesh electrodes for the
568 bioremediation of coking wastewater. Bioresour. Technol. 216, 684-690. Doi:
569 http://dx.doi.org/10.1016/j.biortech.2016.06.009

570 Joshi, D.R., Zhang, Y., Tian, Z., Gao, Y. and Yang, M. (2016) Performance and microbial
571 community composition in a long-term sequential anaerobic-aerobic bioreactor operation
572 treating coking wastewater. Appl. Microbiol. Biotechnol. 100 (18), 8191-202.Doi:
573 http://dx.doi.org/10.1007/s00253-016-7591-8

574 Kim, S.J. and Katayama, Y. (2000) Effect of growth conditions on thiocyanate degradation and
575 emission of carbonyl sulfide by *Thiobacillus thioparus* THI115. Water Res. 34(11),
576 2887-2894.Doi: http://dx.doi.org10.1016/S0043-1354(00)00046-4

577 Kim, Y.M., Park, D., Lee, D.S., Jung, K.A. and Park, J.M. (2009) Sudden failure of biological
578 nitrogen and carbon removal in the full-scale pre-denitrification process treating cokes
579 wastewater. Bioresour. Technol. 100(19), 4340-4347.
580 Doi:http://dx.doi.org/10.1016/j.biortech.2009.04.014

581 Lay-Son, M. and Drakides, C. (2008) New approach to optimize operational conditions for the
582 biological treatment of a high-strength thiocyanate and ammonium waste: pH as key
583 factor. Water Res. 42(3), 774-780. Doi: http://dx.doi.org/10.1016/j.watres.2007.08.009

584 Li, H.B., Cao, H.B., Li, Y.P., Zhang, Y. and Liu, H.R. (2010) Effect of organic compounds on
585 nitrite accumulation during the nitrification process for coking wastewater. Water Sci.
586 Technol. 62(9), 2096-2105. DOI: http://dx.doi.org/10.2166/wst.2010.371

587 Li, Y.M., Gu, G.W., Zhao, J.F., Yu, H.Q., Qiu, Y.L. and Peng, Y.Z. (2003) Treatment of coke-
588 plant wastewater by biofilm systems for removal of organic compounds and nitrogen.
589 Chemosphere 52, 997-1005. Doi: http://dx.doi.org/10.1016/S0045-6535(03)00287-X

590 Liang, Y., Van Nostrand, J.D., Deng, Y., He, Z., Wu, L., Zhang, X., Li, G. and Zhou, J. (2011)
591 Functional gene diversity of soil microbial communities from five oil-contaminated
592 fields in China. ISME J. 5(3), 403-413. Doi: http://dx.doi.org/10.1038/ismej.2010.142

593 Lim, B.R., Y., H.H. and K, F. (2003) Biological degradation and chemical oxidation
594 characteristics of coke-oven wastewater. Water Air Soil Pollut. 146(1), 23-33. Doi:
595 http://dx.doi.org/10.1023/A:1023923409422

596 Liu, Y., Liu, J., Zhang, A. and Liu, Z. (2017) Treatment effects and genotoxicity relevance of the
597 toxic organic pollutants in semi-coking wastewater by combined treatment process.
598 Environ. Pollut. 220(A), 13-19. Doi: http://dx.doi.org/10.1016/j.envpol.2016.04.095

599 Ma, Q., Qu, Y., Shen, W., Zhang, Z., Wang, J., Liu, Z., Li, D., Li, H. and Zhou, J. (2015a)
600 Bacterial community compositions of coking wastewater treatment plants in steel
601 industry revealed by Illumina high-throughput sequencing. Bioresour. Technol. 179, 8.
602 Doi: http://dx.doi.org/10.1016/j.biortech.2014.12.041

603 Ma, Q., Qu, Y., Zhang, X., Liu, Z., Li, H., Zhang, Z., Wang, J., Shen, W. and Zhou, J. (2015b)
604 Systematic investigation and microbial community profile of indole degradation
605 processes in two aerobic activated sludge systems. Sci. Rep. 5, 17674. Doi:
606 http://dx.doi.org/10.1038/srep17674

607 Magoc, T. and Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve
608 genome assemblies. Bioinformatics 27(21), 2957-2963. Doi:
609 http://dx.doi.org/10.1093/bioinformatics/btr507

610 Marrot, B., Barrios-Martinez, A., Moulin, P. and Roche, N. (2006) Biodegradation of high
611 phenol concentration by activated sludge in an immersed membrane bioreactor. Biochem.
612 Eng. J. (2006) 30, 174-183. Doi: http://dx.doi.org/10.1016/j.bej.2006.03.006

613 Nostrand, J.D.V., Yin, H., Wu, L., Yuan, T. and Zhou, J. (2016) Hybridization of environmental
614 microbial community nucleic acids by GeoChip. Martin, F. and Uroz, S. (eds), pp. 183-
615 196, Springer Science+Business Media, Newyork. Doi: http://dx.doi.org/10.1007/978-1-
616 4939-3369-3_11

617 Pal, P. and Kumar, R. (2014) Treatment of coke wastewater: A critical review for developing
618 sustainable management strategies. Sep. Purif. Rev. 43(2), 89-123. Doi:
619 http://dx.doi.org/10.1080/15422119.2012.717161

620 Papadimitriou, C.A., Samaras, P. and Sakellaropoulos, G.P. (2009) Comparative study of phenol
621 and cyanide containing wastewater in CSTR and SBR activated sludge reactors.
622 Bioresour. Technol. 100, 31-37. Doi: http://dx.doi.org/10.1016/j.biortech.2008.06.004

623 Seo, J.S., Keum, Y.S. and Li, Q.X. (2009) Bacterial degradation of aromatic compounds. Int. J.
624 Environ. Res. Public Health 6(1), 278-309. Doi: http://dx.doi.org/10.3390/ijerph6010278

625 Sharma, N.K. and Philip, L. (2014) Effect of cyanide on phenolics and aromatic hydrocarbons
626 biodegradation under anaerobic and anoxic conditions. Chem. Eng. J. 256, 255-267. Doi:
627 http://dx.doi.org/10.1016/j.cej.2014.06.070

628 Sharma, N.K. and Philip, L. (2016) Combined biological and photocatalytic treatment of real
629 coke oven wastewater. Chem. Eng. J. 295, 20-28. Doi:
630 http://dx.doi.org/10.1016/j.cej.2016.03.031

631 Shen, T., Pi, Y., Bao, M., Xu, N., Li, Y. and Lu, J. (2015) Biodegradation of different petroleum
632 hydrocarbons by free and immobilized microbial consortia. Environ. Sci. Processes
633 Impacts 17(12), 2022-2033. Doi: http://dx.doi.org/10.1039/C5EM00318K

634 Sierra-Garcia, I.N., Correa Alvarez, J., de Vasconcellos, S.P., Pereira de Souza, A., dos Santos
635 Neto, E.V. and de Oliveira, V.M. (2014) New hydrocarbon degradation pathways in the
636 microbial metagenome from Brazilian petroleum reservoirs. PLoS One 9(2), e90087.
637 Doi: http://dx.doi.org/10.1371/journal.pone.0090087

638 Staib, C. and Lant, P. (2007) Thiocyanate degradation during activated sludge treatment of coke-
639 ovens wastewater. Biochem. Eng. J. 34, 122-130. Doi:
640 http://dx.doi.org/10.1016/j.bej.2006.11.029

641 Suenaga, H., Koyama, Y., Miyakoshi, M., Miyazaki, R., Yano, H., Sota, M., Ohtsubo, Y., Tsuda,
642 M. and Miyazaki, K. (2009) Novel organization of aromatic degradation pathway genes
643 in a microbial community as revealed by metagenomic analysis. ISME J. 3(12), 1335-
644 1348. Doi: http://dx.doi.org/10.1038/ismej.2009.76

645 Suenaga, H., Mizuta, S., Miyazaki, K. and Yaoi, K. (2014) Diversity of extradiol dioxygenases
646 in aromatic-degrading microbial community explored using both culture-dependent and

culture-independent approaches. FEMS Microbiol. Ecol. 90, 367-379. Doi: http://dx.doi.org/10.1111/1574-6941.12390

Tao, F., Zhao, P., Li, Q., Su, F., Yu, B., Ma, C., Tang, H., Tai, C., Wu, G. and Xu, P. (2011) Genome sequence of *Rhodococcus erythropolis* XP, a biodesulfurizing bacterium with industrial potential. J. Bacteriol. 193(22), 6422-6423. Doi: http://dx.doi.org/10.1128/JB.06154-11

Teh, C.Y., Budiman, P.M., Shak, K.P.Y. and Wu, T.Y. (2016) Recent advancement of coagulation–flocculation and its application in wastewater treatment. Ind. Eng. Chem. Res. 55 (16), 4363–4389. Doi: http://dx.doi.org/10.1021/acs.iecr.5b04703

Tian, Z., Zhang, Y., Yu, B. and Yang, M. (2016) Changes of resistome, mobilome and potential hosts of antibiotic resistance genes during the transformation of anaerobic digestion from mesophilic to thermophilic. Water Res. 98, 261-269. Doi: http://dx.doi.org/10.1016/j.watres.2016.04.031

Vazquez, I., Rodriguez, J., Maranon, E., Castrillon, L. and Fernandez, Y. (2006a) Simultaneous removal of phenol, ammonium and thiocyanate from coke wastewater by aerobic biodegradation. J. Hazard. Mater. 137(3), 1773-1780. Doi: http://dx.doi.org/10.1016/j.jhazmat.2006.05.018

Vazquez, I., Rodriguez, J., Maranon, E., Castrillon, L. and Fernandez, Y. (2006b) Study of the aerobic biodegradation of coke wastewater in a two and three-step activated sludge process. J. Hazard. Mater. 137(3), 1681-1688. Doi: http://dx.doi.org/10.1016/j.jhazmat.2006.05.007

Wang, X., Hu, M., Xia, Y., Wen, X. and Ding, K. (2012) Pyrosequencing analysis of bacterial diversity in 14 wastewater treatment systems in China. Appl. Environ. Microbiol. 78(19), 7042-7047. Doi: http://dx.doi.org/10.1128/AEM.01617-12

Watts, M.P. and Moreau, J.W. (2016) New insights into the genetic and metabolic diversity of thiocyanate-degrading microbial consortia. Appl. Microbiol. Biotechnol. 100(3), 1101-1108. Doi: http://dx.doi.org/10.1007/s00253-015-7161-5

674 Xu, P., Yu, B., Li, F.L., Cai, X.F. and Ma, C.Q. (2006) Microbial degradation of sulfur, nitrogen
675 and oxygen heterocycles. Trends Microbiol. 14(9), 398-405. Doi:
676 http://dx.doi.org/10.1016/j.tim.2006.07.002

677 Zhang, M., Tay, J.H., Qian, Y. and Gu, X.S. (1998) Coke plant wastewater treatment by fixed
678 biofilm system for COD and NH_3-N removal. Water Res. 32(2), 519-527. Doi:
679 http://dx.doi.org/10.1016/S0043-1354(97)00231-5

680 Zhang, T., Ding, L., Ren, H. and Xiong, X. (2009) Ammonium nitrogen removal from coking
681 wastewater by chemical precipitation recycle technology. Water Res. 43(20), 5209-5215.
682 Doi: https://doi.org/10.1016/j.watres.2009.08.054

683 Zhang, W., Wei, C., Yan, B., Feng, C., Zhao, G., Lin, C., Yuan, M., Wu, C., Ren, Y. and Hu, Y.
684 (2013a) Identification and removal of polycyclic aromatic hydrocarbons in wastewater
685 treatment processes from coke production plants. Environ. Sci. Pollut. Res. Int. 20(9),
686 6418-6432. Doi: http://dx.doi.org/10.1007/s11356-013-1697-7

687 Zhang, Y., Xie, J., Liu, M., Tian, Z., He, Z., van Nostrand, J.D., Ren, L., Zhou, J. and Yang, M.
688 (2013b) Microbial community functional structure in response to antibiotics in
689 pharmaceutical wastewater treatment systems. Water Res. 47(16), 6298-6308. Doi:
690 http://dx.doi.org/10.1016/j.watres.2013.08.003

691 Zhao, J.L., Jiang, Y.X., Yan, B., Wei, C., Zhang, L.J. and Ying, G.G. (2014) Multispecies acute
692 toxicity evaluation of wastewaters from different treatment stages in a coking
693 wastewater treatment plant. Environ. Toxicol. Chem. 33(9), 1967-1975. Doi:
694 http://dx.doi.org/10.1002/etc.2638

695 Zhu, S., Wu, H., Wei, C., Zhou, L. and Xie, J. (2016) Contrasting microbial community
696 composition and function perspective in sections of a full-scale coking wastewater
697 treatment system. Appl. Microbiol. Biotechnol. 100(2), 949-960. Doi:
698 http://dx.doi.org/10.1007/s00253-015-7009-z

699

700

Table 1

Correlation of all functional genes, organic pollutant degrading genes, and microbial community structure with environmental variables and performance of bioreactor as shown by Mantel test

		All functional genes		Organic pollutant degrading genes		Microbial community	
		r	*P*	*r*	*P*	*r*	*P*
Wastewater quality	COD	0.023	0.416	0.112	0.281	0.209	0.108
	Total Phenol	-0.183	0.882	-0.170	0.771	-0.039	0.514
	Total nitrogen	0.257	0.074	0.203	0.103	**0.621**	**0.001**
	Ammonia nitrogen	0.253	0.078	0.172	0.150	**0.571**	**0.006**
	Sulfate	0.023	0.434	0.142	0.213	**0.442**	**0.016**
Operational condition	pH	**0.31**	**0.038**	**0.319**	**0.040**	**0.534**	**0.005**
Performance of bioreactor	COD removal	0.072	0.289	-0.167	0.749	0.059	0.310
	Total Phenol removal	-0.233	0.957	-0.087	0.608	-0.098	0.738
	$d\mathrm{NH_4^+}$–N/COD	**0.456**	**0.004**	**0.489**	**0.002**	**0.535**	**0.002**
	$d\mathrm{SO_4^{2-}}$/COD						

r represents statistical correlation coefficient, *P* represents *P* value. Bold figures indicate significant correlation ($P < 0.05$).

Table 2

Correlation between abundances of organic pollutant degrading genes and the release of ammonia nitrogen ($d\mathrm{NH_4^+}$–N/COD) and sulfate ($d\mathrm{SO_4^{2-}}$/COD) as revealed by the Mantel test ($P < 0.05$)

Substrate	Enzyme	Genes	*r*	*P*
Aromatic carboxylic acid	Benzoyl-CoA reductase, subunit A	*bco*	0.581	0.002
	Salicylate hydroxylase	*nagG*	0.469	0.004
	Toluate 1,2-dioxygenase subunit-α	*xylXY*	0.810	0.001
Aromatics	Acylamide amidohydrolase	*amiE*	0.371	0.042
	Catechol 1,2 dioxygenase	*catechol*	0.378	0.023
	Nitrilase	*Nitrilase**	0.279	0.07
	Aromatic 1,2-dioxygenase	*one_ring_12diox*	0.367	0.038
	Protocatechuate 4,5-dioxygenase	*proO*	0.348	0.022
	Aniline dioxygenase	*tdnB**	0.285	0.078
	3-hydroxybenzoate 6-hydroxylase	*xlnD*	0.358	0.036
BTEX related aromatics	Methane/phenol/toluene hydroxylase	*tomA*	0.499	0.01
	Muconate cycloisomerase	*catB*	0.411	0.016
	Cresol dehydrogenase	*pchcf*	0.365	0.043
Chlorinated aromatics	Aniline dioxygenase	*tfdA*	0.382	0.032
Heterocyclic	Dibenzothiophene oxidase	*DbtAc*	0.324	0.047

aromatics				
Nitroaromatics	Nitroreductase	*nfsA_2*	0.345	0.033
	Nitroreductase	*nfsB_2*	0.448	0.005
	Nitrile hydratase	*nhh*	0.527	0.006
	Nitroreductase	*nitroreductase_1*	0.349	0.034
Polycyclic aromatics	2-oxo-4-hydroxypentanoate aldolase	*bphF1*	0.503	0.004
Xenobiotic related Hydrocarbons	Amidohydrolase	*trzA*	0.417	0.012
	Hydroxydechloro atrazine ethylaminohydrolase	*atzB*	0.383	0.037
	Cyanuric acid amidohydrolase	*atzD*	0.539	0.004
	Methylamine dehydrogenase small subunit	*mauAB*	0.649	0.001
Other hydrocarbons	Taurine dioxygenase	*sdsA*	0.516	0.001
Others	Cytochrome P450 monooxygenase	*p450aro*	0.495	0.008

r represents statistical correlation coefficient.*Significance level, $P < 0.1$

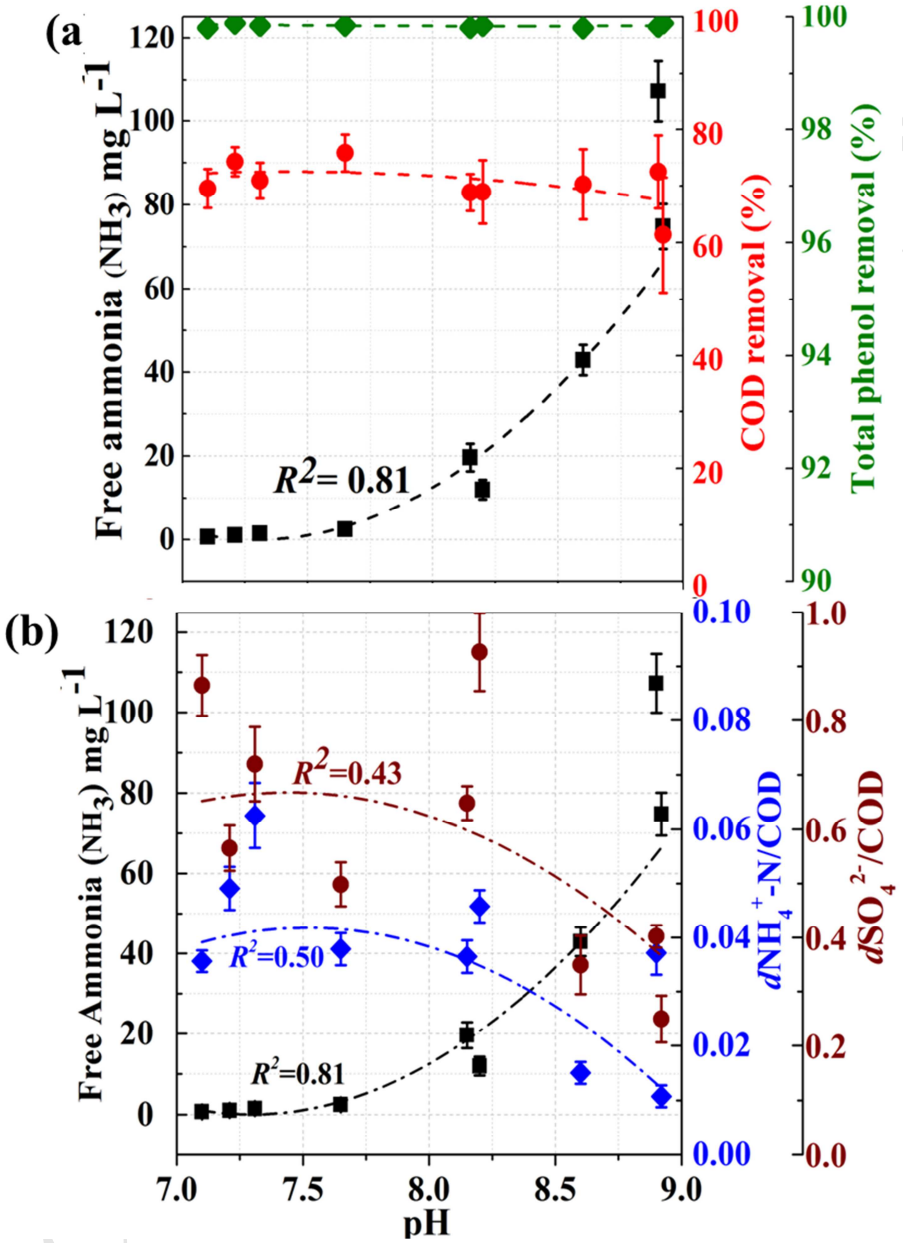


Fig. 1. Effect of pH and free ammonia (■) on (a) total phenol (◆) and COD (●) removal efficiency, (b) release of ammonia nitrogen ($d\mathrm{NH_4^+}$–N) (◆) and sulfate ($d\mathrm{SO_4^{2-}}$) (●) concentration with respect to influent COD during aerobic treatment. Dotted or dashed curve lines represent best possible polynomial curve fits.

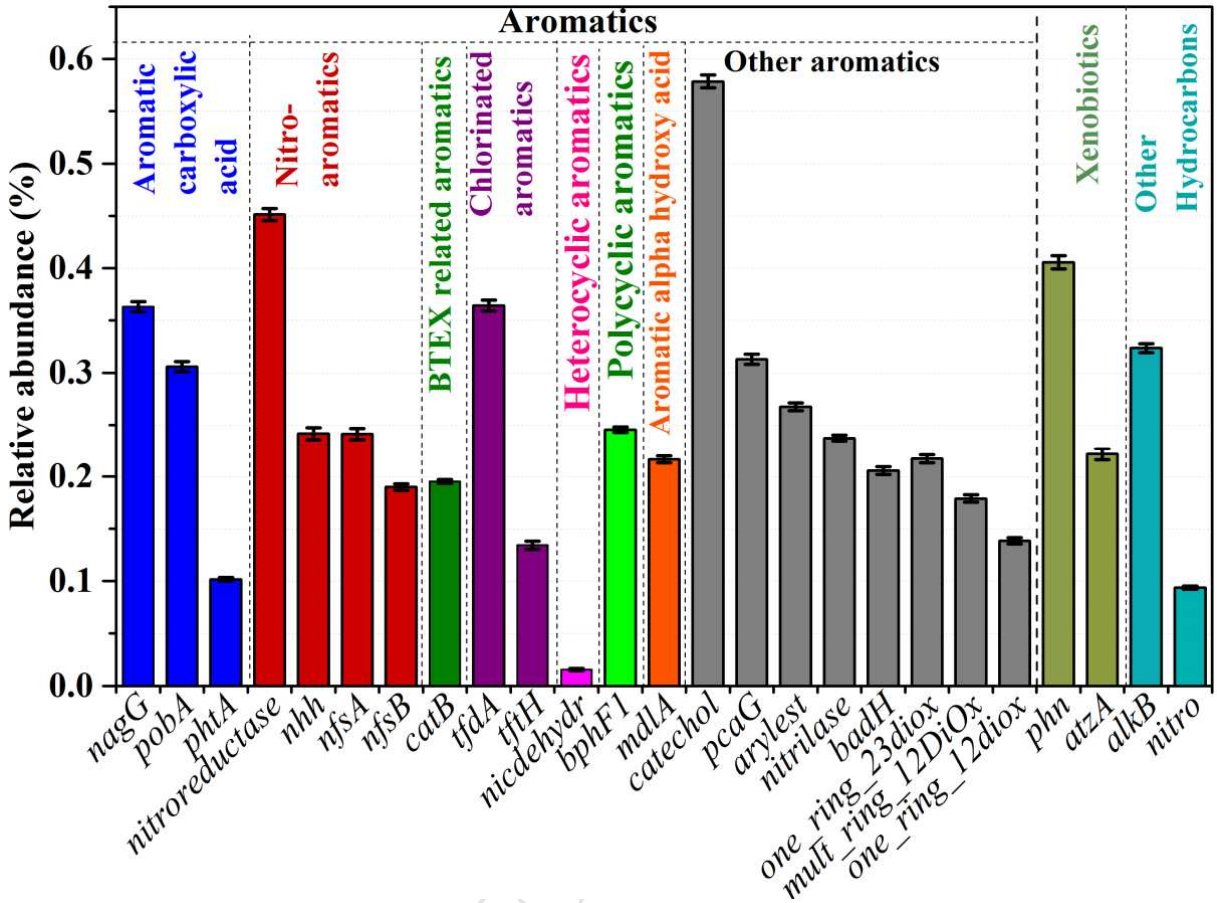


Fig. 2. Relative abundances of various organic pollutant degrading genes. The relative abundance was determined as percentage of the total signal intensities (normalized) of all genes detected by GeoChip 5.0. Data are presented as mean values from nine sludge samples and error bars represent standard deviation.

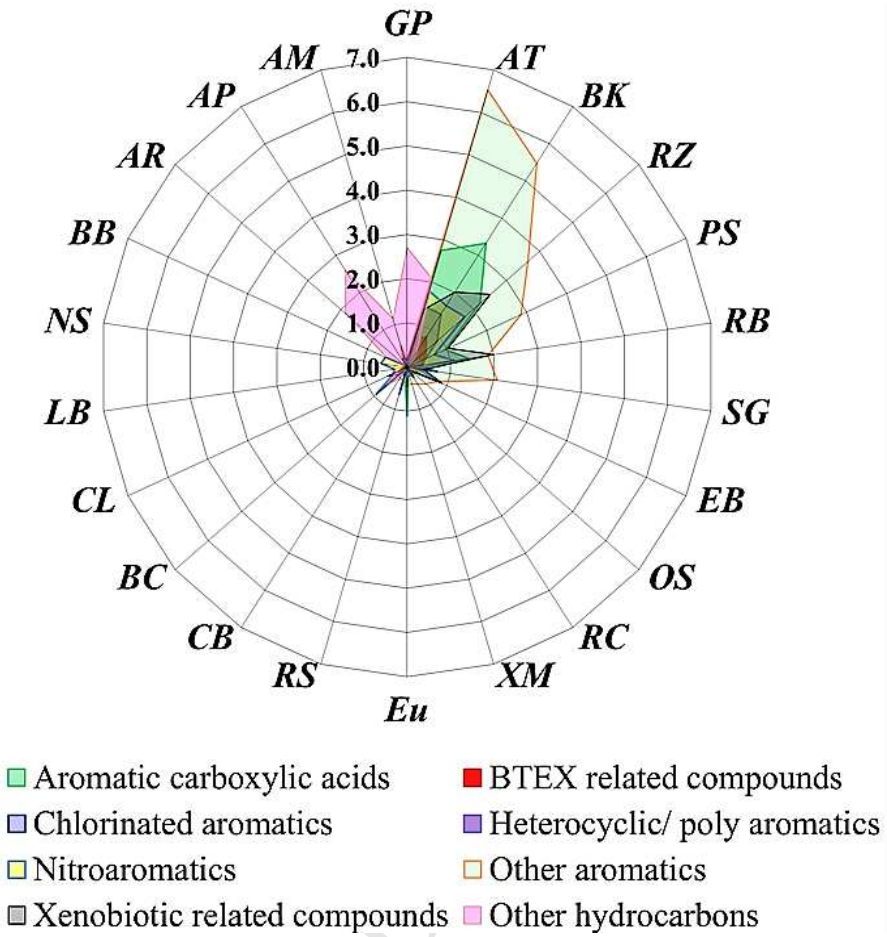


Fig. 3. Taxa-function relationships for organic pollutant degrading genes detected by GeoChip 5.0. Relative abundance was calculated as a percentage of normalized total signal intensities of gene categories derived from given taxa (order level) for each bioprocess. Mean values of relative abundances of all nine samples were plotted for the organic pollutant degrading functional genes. Abbreviation of microbial orders are: *AT = Actinomycetales, BK = Burkholderiales, RZ = Rhizobiales, PS = Pseudomonadales, SG = Sphingomonadales, RB = Rhodobacterales, EB = Enterobacteriales, OS = Oceanospirillales, RC = Rhodocyclales, XM = Xanthomonadales, Eu = Eurotiales, RS = Rhodospirillales, CB = Caulobacterales, BC = Bacillales, SD = Sordariales, AP = Other Alphaproteobacteria, AM = Alteromonadales, AR = Aeromonadales, GP = Other Gammaproteobacteria, CL = Clostridiales, LB = Lactobacillales, NS = Neisseriales, BB = Bifidobacteriales.*

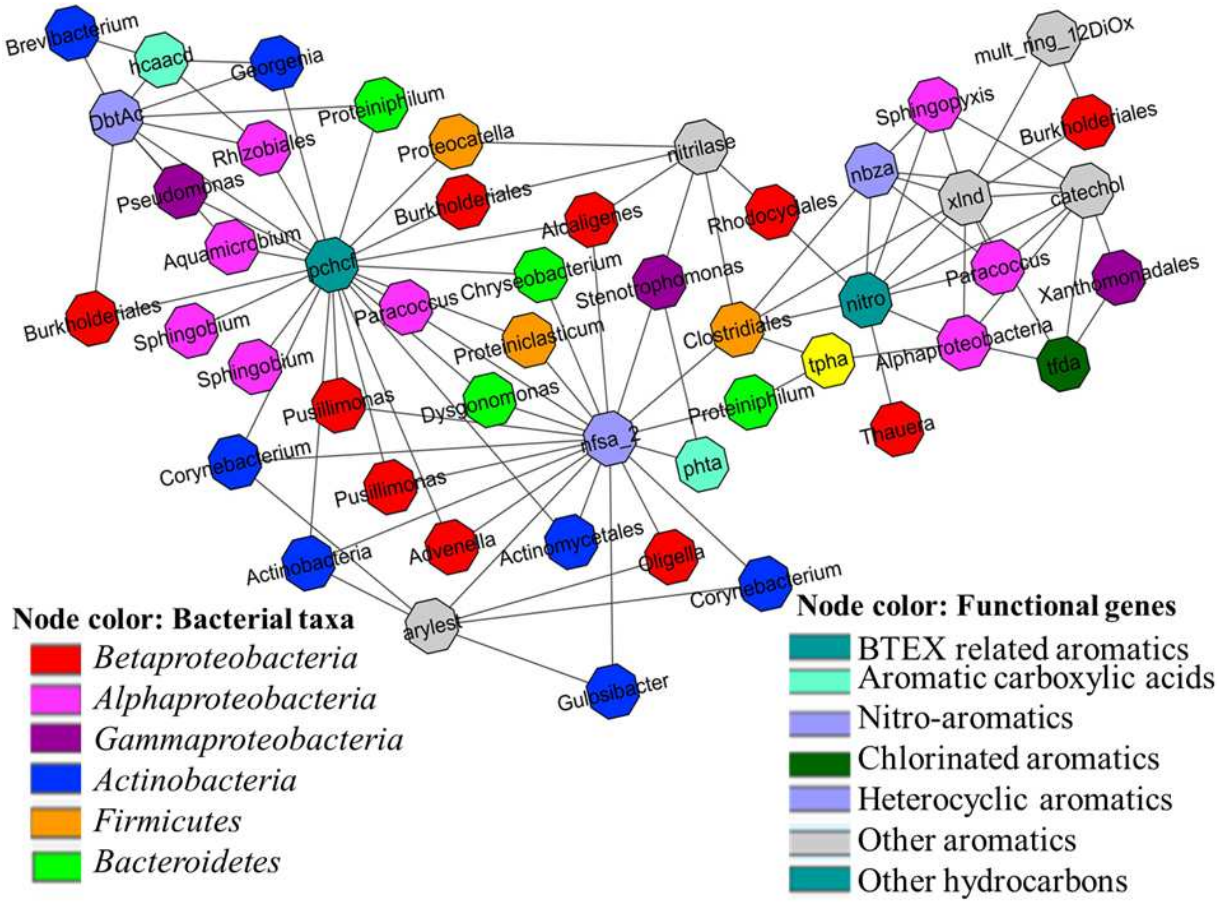


Fig. 4. Network analysis revealing the co-occurrence patterns between bacterial taxa obtained by 16S rRNA gene sequencing (Illumina MiSeq) and organic pollutant degrading genes (GeoChip 5.0). Each of the nodes represents either bacterial genera or functional genes. The solid line (edge) between nodes denotes the positive correlation ($p < 0.05$) between the abundances of linked taxa and genes. Network was visualized by *Cystoscape V3.3.0*.

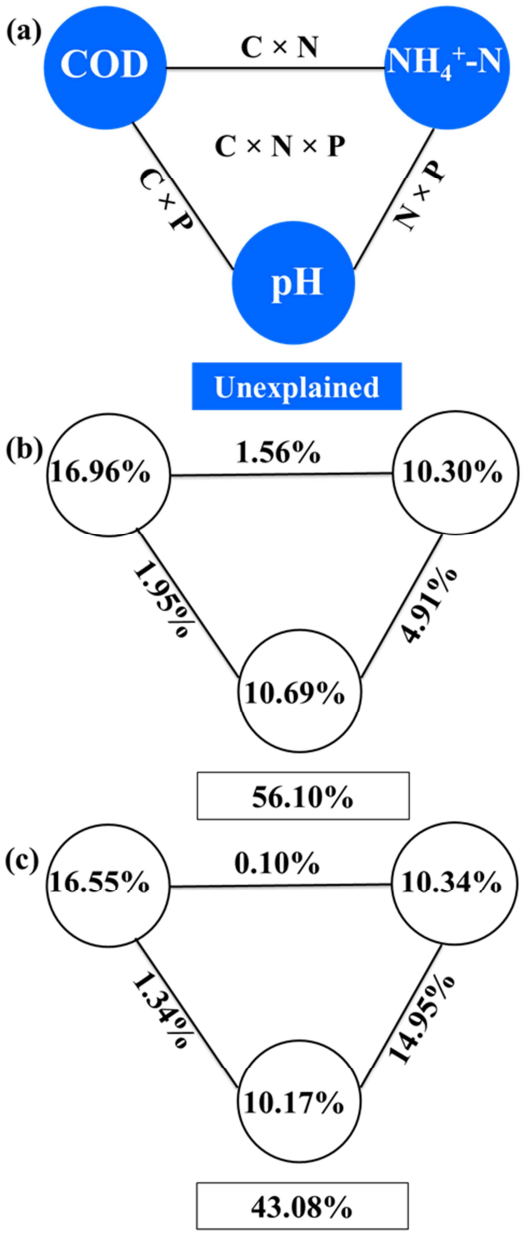


Fig. 5. Variation partitioning analysis of microbial diversity explained by influent COD (C), influent NH_4^+–N (N), and pH (P): (a) general outline, (b) all organic pollutant degrading genes as obtained from GeoChip 5.0 data, and (c) bacterial OTUs as obtained from Illumina MiSeq data. Each diagram represents the biological variation partitioned into the relative effects of each variable, in which geometric areas are proportional to the respective percentages of explained variation. Each node represents the variation explained by the respective variable alone. The edge represents the interaction between adjoining node variables.

Highlights

- Biotransformation of N and S- containing pollutants deteriorated at pH >8.0.
- pH and NH_4^+–N partially shaped variation in microbial community functional genes.
- Abundance of functional genes linked with biotransformation of N- and S- pollutants.
- *Bulkholderia*, *Actinomycetes*, *Pseudomonas* and *Thiobacillus* were key functional taxa.
- Aromatic dioxygenases were most abundant organic pollutant removal genes.