Full Paper

Whole gene transcriptomic analysis of PCB/biphenyl degrading *Rhodococcus jostii* RHA1

(Received March 15, 2018; Accepted August 9, 2018; J-STAGE Advance publication date: January 28, 2019)

Shazwana Sha'arani,^{1,†} Hirofumi Hara,^{2,*,†} Hiroya Araie,³ Iwane Suzuki,³ Megat Johari Megat Mohd Noor,¹ Fazrena Nadia MD Akhir,¹ Nor'azizi Othman,⁴ and Zuriati Zakaria¹

¹ Department of Environmental Engineering and Green Technology, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia

² Department of Chemical Process Engineering, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia

³ Graduate School of Life and Environmental Science, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572, Japan

⁴ Department of Mechanical Precision Engineering, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia

This study gives the first picture of whole RNA-Sequencing analysis of a PCB-degrading microbe, *Rhodococcus jostii* RHA1. Genes that were highly expressed in biphenyl-grown cells, compared with pyruvate-grown cells, were chosen based on the Reads Per Kilobase Million (RPKM) value and were summarized based on the criteria of RPKM ≥ 100 and fold change ≥ 2.0 . Consequently, 266 total genes were identified as genes expressed particularly for the degradation of biphenyl. After comparison with previous microarray data that identified highly-expressed genes, based on a fold change ≥ 2.0 and p -value ≤ 0.05 , 62 highly-expressed genes from biphenyl-grown cells were determined from both analytical platforms. As these 62 genes involve known PCB degradation genes, such as *bph*, *eth*, and *ebd*, the genes identified in this study can be considered as essential genes for PCB/biphenyl degradation. In the 62 genes, eleven genes encoding hypothetical proteins were highly expressed in the biphenyl-grown cells. Meanwhile, we identified several highly-expressed unannotated DNA regions on the opposite strand. In order to verify the encoded proteins, two regions were cloned into an expression vector. A protein was successfully obtained from one region at approximately 25 kDa from the unannotated strand. Thus, the genome se-

quence with transcriptomic analysis gives new insight, considering re-annotation of the genome of *R. jostii* RHA1, and provides a clearer picture of PCB/biphenyl degradation in this strain.

Key Words: biphenyl degradation; polychlorinated biphenyl (PCB); RHA1; *Rhodococcus jostii*

Introduction

Polychlorinated biphenyls (PCBs) are one of the most extensively distributed classes of chlorinated pollutants found in the environment (Tiedje et al., 1994). PCBs are recalcitrant compounds that persist in the environment. They are synthetically produced by the method of direct chlorination of biphenyl. Additionally, the term PCBs is assigned to a group of molecules that have a biphenyl backbone, in which one to ten of the hydrogens have been substituted with a chlorine atom (Furukawa, 2000; Pieper, 2005). Essentially, bacteria can degrade PCBs by two processes: aerobic degradation through the biphenyl pathway and anaerobic dechlorination (Fennell et al., 2004; Furukawa and Fujihara, 2008; Macková et al., 2007). A vast variety of microorganisms are capable of aerobically transforming some of the 209 possible PCB congeners through the *bph* pathway, which degrades biphenyl to benzoate and 2-hydroxypenta-2,4-dienoate (Abramowicz,

*Corresponding author: Hirofumi Hara, Department of Chemical Process Engineering, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Jalan Sultan Yahya, 54100, Kuala Lumpur, Malaysia.

Tel: (+60) 3-2203-1264 Fax: (+60) 3-2203-1266 E-mail: hhara@utm.my

[†]Both authors contributed equally to this work.

None of the authors of this manuscript has any financial or personal relationship with other people or organizations that could inappropriately influence their work.

1990; Furukawa, 2000). It has been demonstrated that PCBs follow the same catabolic pathway as biphenyl and utilize the same enzymes (Ellis et al., 2006). Biphenyl dioxygenase plays a critical role in PCB degradation by catalyzing the first step (Erickson and Mondello, 1992; Iwasaki et al., 2007). A lot of effort has been invested in studying the *bph* pathway to enhance our understanding of its enzymatic makeup, to discover the steps that limit PCB transformation, and to improve its PCB-degrading capabilities (Abramowicz, 1990; Erickson and Mondello, 1992; Kimbara et al., 1989; Taira et al., 1992). It cannot be denied that studies on PCB metabolism are entering a divergent era with the application of genetically engineered microorganisms, such as *Rhodococcus jostii* RHA1, for bioremediation processes.

R. jostii RHA1, which belongs to the Rhodococci group, has become a model bacterium to study genetic functions. *R. jostii* RHA1 is well suited for industrial purposes and bioremediation applications, because it can resist several kinds of environmental stresses, including osmotic, temperature, and desiccation stresses. This strain was isolated from lindane-contaminated soil, and it is well known for its ability to degrade PCBs (Seto et al., 1995). McLeod et al. (2006) found that the genome of *R. jostii* RHA1 was 9,702,737 base pairs in length, which makes it one of the largest known bacterial genomes, and that it was arranged in four linear replicons, consisting of one chromosome and three plasmids (pRHL1, pRHL2, and pRHL3). They determined that *R. jostii* RHA1 has a G+C content of 67% and 9,145 protein-coding genes, which are rich in ligases and oxygenases. Using biphenyl or pyruvate as the sole carbon source, Gonçalves et al. (2006) compared the transcript abundance of *R. jostii* RHA1. From the results, they saw that the *bph*, *ben*, and *cat* genes were significantly upregulated during biphenyl growth.

The development of microarray technology is a result of the accessibility of the complete genome sequences of many organisms and allows researchers to study gene expression on a single chip (Mockler et al., 2005). Next-generation sequencing (NGS) appears to be a realistic option in this modern era and is dramatically changing the revolution of DNA and RNA technologies. After learning where the genetic information was stored, researchers' attention turned to determinations of its content, expression, and regulation, which leads to important discoveries regarding the genes. RNA-Seq by NGS shows clear advantages over existing profiling technologies in terms of the amount of sequence coverage, the accuracy of defining the transcription level, the ability to reveal new transcriptomics insights, etc. Although NGS is becoming more accessible to researchers, as this technique does not require a clone library, microarray technology still remains reliable for many researchers for various reasons (Blow, 2009).

The NGS platform strength depends on its ability to place millions of DNA fragments onto one chip and process the fragments simultaneously. It results in throughput enhancement and allows a larger scope of input data. Besides, RNA-Seq by NGS is better, as it is more sensitive, specific, and accurate in terms of its technology, compared to microarrays, and can report strand specific sequencing

(Willenbrock et al., 2009). Furthermore, RNA-Seq by NGS always considers the fold change and expression level, based on the number of reads, compared to a microarray that only considers the fold change, based on the fluorescence intensity. The 'closed system' of a microarray regards only the sequences that are specific to the probes on the array. Meanwhile, NGS represents an 'open system,' which is relevant for cataloguing gene diversity, including the discovery of novel gene diversity, without any prior sequence information (Roh et al., 2010).

To gain a better understanding of the gene function of PCB/biphenyl degradation in *R. jostii* RHA1, we focused, in this study, on the genes with higher expression, after growth on biphenyl, compared with pyruvate, by analysis with RNA-Seq. We further compared the results with previous microarray data (Gonçalves et al., 2006). A comparison has been made between both platforms based on defined criteria. Further, several highly expressed unannotated DNA regions on the opposite strand were identified and characterized based on specified criteria as well. In the second part of this paper, using wet lab experiments, we identified a novel gene that might conclusively affect biphenyl metabolism in *R. jostii* RHA1.

Methods

Preparation of *R. jostii* RHA1 cells. Wild-type *R. jostii* RHA1 strains were maintained in 1/5 Luria Bertani (LB) agar plates at 30°C. Cells in the exponential growth phase with either 10 mM biphenyl or 10 mM pyruvate in W-minimum media were harvested after three days of culture, and they were used for total RNA extraction. For long-term storage, a *R. jostii* RHA1 stock culture was maintained in a 40% (v/v) glycerol stock and stored at -80°C.

RNA isolation. RNA was extracted from the cells, grown on biphenyl or pyruvate, and sequenced with the Ion Personal Genome Machine (PGM™) System (Life Technologies, USA) platform, following the protocols supplied by Ion PGM™ for the sequencing of mRNA samples. After extraction of the total RNA, the ribosomal RNA was removed by following the manufacturer's procedure for the Ribo-Zero rRNA Removal Kit (Epicentre). Prepared mRNA fragments were then used to synthesize the first and second strand cDNAs using reverse transcriptase (Invitrogen, USA) and random hexamer primers. The cDNA was treated in an end repair reaction with T4 DNA polymerase and Klenow DNA polymerase at the blunt ends. The 3' end of the blunt, phosphorylated DNA fragments was added using an "A" base. Subsequently, an adapter with a "T" base overhang at its 3' end was ligated to the end of the DNA fragment. A range of cDNA template sizes was selected, and the library was amplified using PCR for the enrichment of adapter-ligated fragments. The prepared libraries were quantified and validated for quality using a High Sensitivity Bioanalyzer Chip (Agilent Technologies, USA). Subsequently, cDNA template clusters were sequenced on an Ion PGM™.

RNA-Seq analysis. Single-ended sequence reads were scanned 3' to 5'. The FASTQ Quality Trimmer tool from the FASTX-tool kit was used to trim those with a Phred quality value below 20. Using Bowtie 2 in the end-to-end

mode, trimmed reads were aligned to the *R. jostii* RHA1 genome. Based on the gene annotations found using custom Perl scripts, protein-coding nucleotide sequences were extracted from the *R. jostii* RHA1 genome sequence by the NCBI Prokaryotic Genome Annotation Pipeline. Further, the generated data were analyzed using Artemis Software (Sanger Institute), followed by the identification of differentially expressed genes.

Identification of highly expressed genes by RPKM. As a starting point for data analysis, high expression with biphenyl compared with pyruvate in *R. jostii* RHA1, was examined based on the results of the transcriptomic data from NGS. The fold change and the Read Per Kilobase Million (RPKM) value from the transcriptomic data from NGS were calculated and characterized. This characterization was based on certain criteria that were significant for PCB degradation. A gene was characterized and selected based on the following criteria: RPKM ≥ 100 and fold change ≥ 2 , which are significant for PCB degradation. This criterion ensures that all *bph* genes that are important for PCB degradation are covered. High expression with biphenyl was reflected by the RPKM value, which was calculated using Eq. (1) as follows (Mortazavi et al., 2008):

$$\text{RPKM} = \frac{\text{total reads of the target gene (million)}}{\left\{ \frac{\text{total reads mapped in the genome (million)}}{\text{length of the target gene (kb)}} \right\}} \quad (1)$$

Normalized RPKM value for pyruvate were calculated by the raw RPKM value from pyruvate divided by (total read number of biphenyl divided by total read number of pyruvate). Genes that were significant for PCB degradation from previous data from a microarray (Gonçalves et al., 2006) and this study (RNA-Seq by NGS) were also compared, as the microarray only showed a fold change value, while RNA-Seq by NGS showed both expression levels and a fold change value.

Identification of unannotated DNA regions. Several DNA regions may not have matched based on the annotation of the Artemis data, as it showed expression of mRNA on the opposite strand. Thus, several candidates with high expression in unannotated DNA regions were selected from the opposite strand, based on high expression with biphenyl, but not with pyruvate. The criteria chosen were based on RPKM ≥ 100 and fold change ≥ 1.5 . Among several candidates, two DNA regions, which were named as CD01 (Supplemental Fig. 1) and CD02 (Supplemental Fig. 2), were chosen that were highly expressed with biphenyl and not pyruvate, based on certain criteria for further analysis.

DNA manipulation. For the DNA regions, CD01 and CD02, primers were designed, depending on the transcriptomic data from NGS. Primer design was conducted depending on high expression with biphenyl but not with pyruvate. The primers were designed to ensure coverage of the translation of the open reading frames of CD01 and CD02. The forward and reverse primers with

the sequences 5'-CTACTGCCAGCCGATAGTG-3' and 3'-TTCACGTTCTCGAATCCG-5', respectively, were used to specifically amplify CD01, and the sequences 5'-CTACCCGCTGCAGTGCGGTT-3' and 3'-TGAGAAGTAACCAGTGCTGA-5' were used to amplify CD02.

A reaction mixture (19.0 L) was prepared containing 4.0 L of 5× Green GoTaq Flexi buffer, 0.4 of dNTP mix, 0.1 of GoTaq Flexi DNA polymerase, 0.5 L of the CD01/CD02 forward and CD01/CD02 reverse primers, each at a concentration of 20 μM , 1.0 L of MgCl_2 , and 12.5 L of autoclaved dH_2O . Further, 1.0 L of the extracted DNA was added to the prepared reaction mixture. The PCR reaction used the GoTaq polymerase enzyme protocol from Promega Corporation, USA, to amplify the target gene, as it generates A-tailed fragments used for TA cloning. All PCR assays were performed with a thermal cycler (ProFlexTM PCR System; Applied Biosystems) and analyzed by 1.0% agarose gel electrophoresis. PCR products were gel excised and purified using the GenepHlowTM Gel/PCR extraction kit protocol, before further use in cloning and transformation. For construction of the plasmid, the target gene was ligated into pGEM-T under the control of the lac operon promoter. *Escherichia coli* JM109 was used as the host. *E. coli* cells, harboring the engineered plasmids, were screened and grown at 37°C in LB broth, supplemented with antibiotics (50 g mL^{-1} ampicillin), with shaking at 160 rpm. The plasmid was extracted using the QIAprep Spin Miniprep Kit (Qiagen, USA).

Protein induction with isopropyl- β -D-1-thiogalactopyranoside (IPTG). For the evaluation of protein production, recombinant cells were grown overnight, while incubating at 37°C in LB media and being continuously shaken at 160 rpm. On the next day, prior to induction with IPTG, the optical density was measured, and the cells were collected after centrifugation at $10,000 \times g$ for 2 min. When the culture reached an OD_{600} of ~ 0.5 , protein expression was induced with 1.0 mM IPTG. The concentration of IPTG and the optical density were optimized. The protein concentration was determined by a Qubit[®] 2.0 Fluorometer (Invitrogen). The optical density was checked every two hours, and cells were collected using the protocol adopted by Sambrook and Russell (2001) until the best expression levels were obtained, when analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). All of the chemicals used to conduct SDS-PAGE were purchased from Bio-Rad Laboratories, USA.

Results

Transcriptomic analysis of RHA1 with biphenyl or pyruvate

To investigate the genes that are involved in PCB degradation and are highly expressed with biphenyl compared with pyruvate, we performed whole transcriptome analysis using RNA-Seq. *R. jostii* RHA1 was grown on W-minimum media with biphenyl or pyruvate until log phase and the total RNA was extracted. After removal of rRNA from the total RNA and library preparation, RNA-Seq analysis was performed. Ion PGMTM sequencing provided

Table 1. List of selected genes from comparison of RNA-seq and microarray analysis except for known PCB/biphenyl degradation genes.

ro id	RPKM in biphenyl from RNA-Seq	RPKM in pyruvate from RNA-Seq	Fold change (biphenyl/pyruvate) from RNA-Seq	Annotated function	Fold change (biphenyl/pyruvate) from microarray	p-value from microarray data
ro01050	116.522	44.061	2.64	possible succinate dehydrogenase	2.06	0.00985
ro01372	119.094	49.592	2.40	ATP-dependent protease proteolytic subunit	67.34	0.00011
ro02267	119.246	42.088	2.83	hypothetical protein	4.40	0.00354
ro02282	568.100	145.256	3.91	hypothetical protein	3.03	0.000599
ro02375	145.019	19.603	7.40	intradiol ring-cleavage dioxygenase	4.12	3.88E-05
ro03278	146.716	69.952	2.10	hypothetical protein	2.46	0.0187
ro03798	116.334	31.069	3.74	hypothetical protein	12.32	1.64E-06
ro04011	162.786	72.299	2.25	protein-methionine-S-oxide reductase	15.88	0.000163
ro04316	1525.329	356.301	4.28	possible transcriptional regulator, WhiB family	2.23	0.00812
ro04427	171.811	38.043	4.52	hypothetical protein	3.51	1.24E-06
ro04428	136.773	28.657	4.77	probable thiocyanate hydrolase gamma subunit	11.49	5.85E-08
ro04431	102.972	37.897	2.72	hypothetical protein	2.02	0.025
ro06086	106.190	29.630	3.58	hypothetical protein	2.42	0.00982
ro08019	305.528	1.617	188.95	aldehyde dehydrogenase	2.02	0.00017
ro08046	262.910	2.824	93.09	quinone oxidoreductase	13.60	8.55E-09
ro08048	105.052	2.753	38.15	hypothetical protein	48.330	3.14E-08
ro08049	1019.512	113.812	8.96	enoyl-CoA hydratase	30.58	8.73E-09
ro08050	915.581	84.225	10.87	medium-chain acyl-CoA ligase	10.03	3.88E-05
ro08053	4097.418	13.656	300.04	hypothetical protein	42.82	0.000368

4,460,433 reads for biphenyl-grown cells (mean lead length: 145 bp; total aligned bases: 593 Mb; depth of reference: 61.1X; aligned bases: 92%) and 5,031,570 reads for pyruvate-grown cells (mean lead length: 126 bp; total aligned bases: 584 Mb; depth of reference: 60.2X; aligned bases: 92%); these reads were used for RPKM calculations for each of the sampled conditions.

Identification of highly-expressed genes by RPKM

In our previous research, we identified 770 genes out of total ORF in RHA1 that showed a fold change ≥ 2 and p -value < 0.05 after growth with biphenyl, compared with pyruvate, by using a microarray (Gonçalves et al., 2006). In this research, we identified highly-expressed genes based on the fold change and expression level from RNA-Seq. A gene was considered highly expressed when the mRNA was highly transcribed with biphenyl compared with pyruvate in *R. jostii* RHA1. A total of 266 genes, as presented in Supplementary Table 1, were found to be particularly transcribed for the degradation of PCB from RNA-Seq, based on our defined criteria ($\text{RPKM} \geq 100$ and fold change ≥ 2), when comparing biphenyl with pyruvate. We have set these criteria, as $\text{RPKM} \geq 100$ is considered high for expression relatively compared with the RPKM value from all ORFs, and a fold change ≥ 2 is based on previous microarray data. Ninety-eight genes encoding hypothetical protein were identified based on our defined criteria. Among these genes, we identified *sdhB1* and *sdhA1* for succinate dehydrogenase; *fabD*, *acpM*, *fabF* for acyl carrier protein; *pccB1* acts as propionyl-CoA carboxylase subunit; *cysH* as a phosphoadenylyl-sulfate reductase; *ectC* for L-ectoine synthase; *pcaI1* for 3-oxoacid CoA-transferase; *clpX*, *clpC1* and *clpC2* acts as ATP binding subunit; *tig* for FKBP-type bacterial trigger factor; *tufA* for elongation factor EF1A; *rpoC* and *rpoB* for RNA polymerase; ATP genes like *atpC*, *atpD*, *atpG*, *atpA* and *atpF* as proton transporter; *metE1* for 5-methyltetrahydropteroyltrimethylglutamate-homocysteine S-methyltransferase; *dldH1* function as dihydrolipoyl dehydrogenase; *sodA* for superoxide dismutase; *cspA1* for cold shock protein A; *sufB* and *sufC* which annotated as FeS assembly protein for biphenyl dioxygenase and *thcA* genes as aldehyde dehydrogenase. Forty-one genes have $\text{RPKM} \geq 1000$ in biphenyl grown cells which include all thirty-six genes encoding known PCB degradation genes. For degradation on biphenyl, such as *bph*, *etb*, *ben* and *cat* genes, that have been previously reported in *R. jostii* RHA1 (Gonçalves et al., 2006; Iwasaki et al., 2007; Masai et al., 1995; McLeod et al., 2006; Takeda et al., 2004a, b). In the case of *bph* genes in RNA-Sequencing (e.g., *bphA*, *bphB*, *bphC*, *bphE*, *bphF*, *bphG*) all the expression is ($\text{RPKM} > 4000$) while fold change is (> 200), which is quite consistent. Meanwhile, in a microarray that considers only the fold change is (> 20) that is underestimated. In case of *etb* genes (e.g., *etbAd*, *etbAa1*, *etbAb1*, *etbC*) all the expression is ($\text{RPKM} > 1000$) in RNA-Sequencing. However, the fold change is low (> 2.0) compared with a microarray that has a high fold change (> 10). While other known genes that are important for PCB degradation (*ebd*, *ben*, *cat* genes) shows consistencies with the RNA-Sequencing platform since in microarray detected high fold

change. In RNA-Sequencing the expression level of *ben* genes (e.g., *benA*, *benB*, *benC*, *benD*, *benK*) is (RPKM > 100) and fold change is (>2.0–4.0) while a microarray gives a fold change (>1.0–10.0). Meanwhile, *cat* genes (e.g., *catA1*, *catB*, *catC*) shows expression level (RPKM > 600) and fold change (>5.0) in RNA-Sequencing, whereas in microarray platform gives fold change (>10.0). To screen more confidently, the microarray data and RNA-Seq data from NGS were compared, because these two methods generated data for the comparison of biphenyl- and pyruvate-grown cells. After correlation with the microarray data for highly upregulated genes (fold change ≥ 2.0 and p -value ≤ 0.05), 62 genes were highly expressed out of all the genes, based on our criteria, and included all genes which encode known biphenyl degradation, indicating that they were involved in PCB/biphenyl degradation. After removal of known PCB degradation genes, eleven genes were identified as hypothetical proteins, while the rest were the putative enzymes and regulatory proteins shown in Table 1.

Identification of highly-expressed unannotated DNA regions by RPKM

Unannotated DNA region represents the DNA region that does not have any annotation information in which an unidentified novel gene might be encoded on an opposite strand. It was observed that several highly-expressed DNA regions, from growth with biphenyl, did not match because there were no encoded proteins. In this study, we have characterized several unannotated DNA regions on the antisense strand. We identified around 22 ‘candidates’ (DNA regions) that have a high expression of unannotated DNA regions (antisense strand), based on a RPKM ≥ 100 and fold change ≥ 1.5 , as presented in Supplementary Table 2. Among the 22 candidates, two candidates, the opposite strand of ro01178 (Supplementary Fig. 1) and the opposite strand of ro02144 (Supplementary Fig. 2), named CD01 and CD02, respectively, were randomly chosen for further analysis because of high expression in the antisense strand.

Validation of unannotated DNA regions as encoded proteins

The two candidates chosen (CD01 and CD02) were expressed in *E. coli* to determine if either expressed any protein. After induction with IPTG, a protein with an apparent molecular weight of approximately 25 kDa was successfully obtained for CD01 depend on the time after induction, (Supplementary Fig. 3, lane 7–9) in the antisense strand of ro01178, but not from the sense strand of it (Supplementary Fig. 3, lane 2–4). In this study, we found an encoded protein (CD01) on the antisense strand, as it is highly expressed with biphenyl but not with pyruvate on the bottom strand. For the CD02 region (antisense of ro02144), there was no protein observed (data not shown) with forward and reverse direction of the promoter at a molecular weight around 3 kDa or 6.5 kDa. The putative nucleotide and amino acid sequence of CD01 are shown in Supplementary Fig. 4 as we identify protein around 25 kDa.

Discussion

This study aimed to show a clear picture of gene expression pattern required for biphenyl degradation by *R. jostii* RHA1, based on transcriptomic analysis by NGS. In this study, we also compared the data from two platforms: microarray and our data from RNA-Seq by NGS. The genes involved in PCB degradation, such as *bph*, *etb*, *ebd*, *ben*, and *cat* were included in this criterion, and this proves that genes in our data can be considered convincing enough to be involved in biphenyl degradation. It can obviously be seen that these genes are highly expressed, as determined by both RNA-Seq by NGS and the previous microarray. *R. jostii* RHA1 co-metabolizes PCBs through the biphenyl catabolic pathway. In this pathway, biphenyl is transformed to benzoate and 2-hydroxy-penta-2,4-dienoate (HPD) by the *bphABCD* gene products, and the resulting HPD is further metabolized to pyruvate and acetyl-coenzyme A (CoA) by successive reactions, catalyzed by the HPD metabolic pathway enzymes, including HPD hydratase (HPDH), 4-hydroxy-2-oxovalerate aldolase (HOVA), and acetaldehyde dehydrogenase (acylating) (AADH) (Furukawa, 1994). A study by Masai et al. (1995) showed that PCB degradation by *R. jostii* RHA1 and the transcription of *bph* genes, which are responsible for the degradation of biphenyl, were activated in the presence of biphenyl. They have discovered that, in the presence of benzoate, transcriptional activation of the *bph* genes is repressed. Benzoate is among the metabolites of biphenyl degradation by *R. jostii* RHA1, whereas the substrate that affects the repression is known as catechol. Catechol is a lower metabolite of benzoate. A study by Ito et al. (2013) demonstrated that the repression of transcriptional activation of *bphAa* by catechol was released by the overexpression of the *catA* gene and improved the growth of *R. jostii* RHA1 on biphenyl.

It is worth noting that gene ro04316 appeared to be upregulated in the data from both platforms, previous microarray (Gonçalves et al., 2006) and NGS. Interestingly, the function of this gene is indicated as a possible transcriptional regulator, from the WhiB family. The exact role of WhiB is not yet clear, but a mutation in the gene produces white, tightly-coiled aerial hyphae in *Streptomyces* (Kormanec and Homerova, 1993). The WhiB family has been reported in some species, such as *Streptomyces*, *Nocardia* and *Mycobacterium* (Cole et al., 1998; Lasker et al., 2014; Soliveri et al., 2000). It is hypothesized that this gene might play an important function in the physiological change of *R. jostii* RHA1 during PCB degradation. This study broadens our understanding of how *Rhodococci* can form several physiological changes. Nevertheless, this finding is in contrast with the report from Alvarez (2010), which demonstrated that all *Actinomycetes*, including *R. jostii* RHA1, are nonsporulating bacteria. Meanwhile, eleven hypothetical genes were upregulated. A hypothetical protein is a protein whose existence has been predicted, but there is still inadequate experimental evidence (Galperin and Koonin, 2004). It is common that in any sequenced bacterial genome, approximately 30–40% genes will not have an assigned function (Bork, 2000). The experimental charac-

terization of a hypothetical protein is expected to reveal new, crucial aspects of microbial biology and could also lead to a better functional prediction for PCB degradation.

Atago et al. (2016) had previously chosen eight BphT1-targeting genes that take part in biphenyl metabolism based on a ChIP-chip assay, and we compared their results with our experimental data, as shown in Supplementary Table 3. BphS1 and BphT1 is a two-component system that plays the main role in the gene activation of *R. jostii* RHA1 for biphenyl degradation (Takeda et al., 2004b). The two-component regulatory system (BphST) has been described as the transcriptional regulatory system in the *Rhodococcus* sp. strains M5 (Labbé et al., 1997) and *R. jostii* RHA1 (Takeda et al., 2004b) for biphenyl degradation. A comparison with our study shows that five genes are on the list as being highly expressed, except ro02841, ro08231, and ro08628, based on our criteria (RPKM ≥ 100 and fold change ≥ 2). Even our data shows five out of eight genes are consistent with previous ChIP-chip analysis, and this also showed that genes in our data might be involved in the biphenyl degradation in RHA1.

To predict which Open Reading Frame (ORF) encodes a protein around 25 kDa from the antisense strand for CD01, the JavaScript program Sequence Manipulation Suite was used to analyze the possible ORF. It is possible that some stop codon can also encode selenocysteine. Since, there are no homologs of selenocysteine synthase in *R. jostii* RHA1 genome, we conclude that there is a possible unknown mechanism involved which needed further investigation in order to evaluate the protein functions. BLAST analysis of CD01 showed >50% identity with the hypothetical proteins from *Rhodococcus* sp. RD6.2 and *Rhodococcus* sp. B7740. Meanwhile, an analysis by SWISS-MODEL homology modeling showed that the name of the possible protein, based on the first hits, is the putative phenazine biosynthesis protein PhcZ/Phz. The phenazines are a class of over 150 bacterial secondary metabolites with redox properties that enable them to act as broad specificity antibiotics, as well as virulence and survival factors (Guttenberger et al., 2017; Okegbe et al., 2012). However, for CD02, there was no protein expressed in either the forward or reverse directions. Therefore, there are two possibilities that this ORF cannot be expressed in *E. coli*, or might encode a small regulatory RNA under the growth of biphenyl. According to an article reviewed by Georg and Hess (2011), increasing the number of small RNAs encoded on the opposite strand of established coding sequences has been found to impact the translation or mRNA stability of the fully complementary sense gene.

Our data shown here provides a clear picture of the whole transcription pattern in the cells grown on biphenyl of a PCB-degrading microbe, *R. jostii* RHA1. From our analysis, genome sequences and the annotation of microbes may not be enough for a complete understanding. To solve this issue, we also need to combine transcription information with transcriptomic data to correct the annotations, based on the expression data from RNA-Seq. Besides, wet lab experiments are also important for confirmation. Thus, a combination of genome data and transcriptomic information from RNA-Seq may give complete and sufficient information about the microbes.

Acknowledgments

SS was financially supported by an MJIT scholarship. This work was supported by the Malaysian-Japan International Institute of Technology, Universiti Teknologi Malaysia.

Supplementary Materials

Supplementary figures and tables are available in our J-STAGE site (<http://www.jstage.jst.go.jp/browse/jgam>).

References

- Abramowicz, D. A. (1990) Aerobic and anaerobic biodegradation of PCBs: A review. *Crit. Rev. Biotechnol.*, **10**, 241–251.
- Alvarez Héctor, M. (ed.) (2010) *Biology of Rhodococcus*, Springer Science & Business Media.
- Atago, Y., Shimodaira, J., Araki, N., Othman, N. A., Zakaria, Z. et al. (2016) Identification of novel extracellular protein for PCB/biphenyl metabolism in *Rhodococcus jostii* RHA1. *Biosci., Biotechnol., Biochem.*, **80**, 1012–1019.
- Blow, N. (2009) Transcriptomics: The digital generation. *Nature*, **458**, 239–242.
- Bork, P. (2000) Powers and pitfalls in sequence analysis: The 70% hurdle. *Genome Res.*, **10**, 398–400.
- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C. et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, **393**, 537–544.
- Ellis, L. B., Hershberger, C. D., and Wackett, L. P. (2006) The University of Minnesota Biocatalysis/Biodegradation database: microorganisms, genomics and prediction. *Nucleic Acids Res.*, **28**, 377–379.
- Erickson, B. D. and Mondello, F. J. (1992) Nucleotide sequencing and transcriptional mapping of the genes encoding biphenyl dioxygenase, a multi-component polychlorinated biphenyl-degrading enzyme in *Pseudomonas* sp. strain LB400. *J. Bacteriol.*, **174**, 2903–2912.
- Fennell, D. E., Nijenhuis, I., Wilson, S. F., Zinder, S. H., and Haggblom, M. M. (2004) *Dehalococcoides ethenogenes* strain 195 reductively dechlorinates diverse chlorinated aromatic pollutants. *Environ. Sci. Technol.*, **38**, 2075–2081.
- Furukawa, K. (1994) Molecular genetics and evolutionary relationship of PCB-degrading bacteria. *Biodegradation*, **5**, 289–300.
- Furukawa, K. (2000) Biochemical and genetic bases of microbial degradation of polychlorinated biphenyls (PCBs). *J. General Appl. Microbiol.*, **46**, 283–296.
- Furukawa, K. and Fujihara, H. (2008) Microbial degradation of polychlorinated biphenyls: biochemical and molecular features. *J. Biosci. Bioeng.*, **105**, 433–449.
- Galperin, M. Y. and Koonin, E. V. (2004) “Conserved hypothetical” proteins: Prioritization of targets for experimental study. *Nucleic Acids Res.*, **32**, 5452–5463.
- Georg, J. and Hess, W. R. (2011) cis-antisense RNA, another level of gene regulation in bacteria. *Microbiol. Mol. Biol. Rev.: MMBR*, **75**, 286–300.
- Gonçalves, E. R., Hara, H., Miyazawa, D., Davies, J. E., Eltis, L. D. et al. (2006) Transcriptomic assessment of isozymes in the biphenyl pathway of *Rhodococcus* sp. strain RHA1. *Appl. Environ. Microbiol.*, **72**, 6183–6193.
- Guttenberger, N., Blankenfeldt, W., and Breinbauer, R. (2017) Recent developments in the isolation, biological function, biosynthesis, and synthesis of phenazine natural products. *Bioorganic Medicinal Chem.*, **25**, 6149–6166.
- Ito, T., Endo, G., Fukuda, M., and Miyauchi, K. (2013) Enhancement of the transcriptional activity of the PCB degradation genes in *Rhodococcus jostii* RHA1 by the introducing of catechol degradation gene. *J. Environ. Eng. Res.*, **69**, 223–229.
- Iwasaki, T., Takeda, H., Miyauchi, K., Yamada, T., Masai, E. et al. (2007) Characterization of two biphenyl dioxygenases for biphenyl/PCB degradation in A PCB degrader, *Rhodococcus* sp. strain RHA1. *Biosci. Biotechnol. Biochem.*, **71**, 993–1002.
- Kimbara, K., Hashimoto, T., Fukuda, M., Koana, T., Takagi, M. et al. (1989) Cloning and sequencing of two tandem genes involved in

- degradation of 2,3-dihydroxybiphenyl to benzoic acid in the polychlorinated biphenyl-degrading soil bacterium *Pseudomonas* sp. strain KKS102. *J. Bacteriol.*, **171**, 2740–2747.
- Kormanec, J. and Homerova, D. (1993) *Streptomyces aureofaciens* *whiB* gene encoding putative transcription factor essential for differentiation. *Nucleic Acids Res.*, **21**, 2512.
- Labbé, D., Garnon, J., and Lau, P. C. K. (1997) Characterization of the genes encoding a receptor-like histidine kinase and a cognate response regulator from a Biphenyl/Polychlorobiphenyl-Degrading Bacterium, *Rhodococcus* sp. strain. *J. Bacteriol.*, **179**, 2772–2776.
- Lasker, B. A., Bell, M., Spro, H. K. C., Schumann, P., and Brown, J. M. (2014) *Nocardia vulneris* sp. nov., isolated from wounds of human patients in North America. *Antonie van Leeuwenhoek*, **106**, 543–553.
- Macková, M., Vrchotová, B., Francová, K., Sylvestre, M., Tomaniová, M. et al. (2007) Biotransformation of PCBs by plants and bacteria—consequences of plant-microbe interactions. *Eur. J. Soil Biol.*, **43**, 233–241.
- Masai, E., Yamada, A., Healy, J. M., Hatta, T., Kimbara, K. et al. (1995) Characterization of biphenyl catabolic genes of gram-positive polychlorinated biphenyl degrader *Rhodococcus* sp. strain RHA1. *Appl. Environ. Microbiol.*, **61**, 2079–2085.
- McLeod, M. P., Warren, R. L., Hsiao, W. W., Araki, N., Myhre, M. et al. (2006) The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc. Nat. Acad. Sci.*, **103**, 15582–15587.
- Mockler, T. C., Chan, S., Sundaresan, A., Chen, H., Jacobsen, S. E. et al. (2005) Applications of DNA tiling arrays for whole-genome analysis. *Genomics*, **85**, 1–15.
- Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, **5**, 1–8.
- Okegbe, C., Sakhtah, H., Sekedat, M. D., Price-Whelan, A., and Dietrich, L. E. P. (2012) Redox eustress: Roles for redox-active metabolites in bacterial signaling and behavior. *Antioxidants Redox Signaling*, **16**, 658–667.
- Pieper, D. H. (2005) Aerobic degradation of polychlorinated biphenyls. *Appl. Microbiol Biotechnol.*, **67**, 170–191.
- Roh, S. W., Guy, C. J., Abell, G. C. J., Kim, K., Nam, Y. et al. (2010) Comparing microarrays and next generation sequencing technologies for microbial ecology research. *Trends Biotechnol.*, **28**, 291–299.
- Sambrook, J. and Russell, R. W. (2001) *Molecular Cloning: A Laboratory Manual*. 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Seto, M., Kimbara, K., Shimura, M., Hatta, T., Fukuda, M. et al. (1995) A novel transformation of polychlorinated biphenyls by *Rhodococcus* sp. strain RHA1. *Appl. Environ. Microbiol.*, **61**, 3353–3358.
- Soliveri, J. A., Gomez, J., Bishai, W. R., and Chater, K. F. (2000) Multiple paralogous genes related to the *Streptomyces coelicolor* developmental regulatory gene *whiB* are present in *Streptomyces* and other actinomycetes. *Microbiology*, **146**, 333–343.
- Taira, K., Hirose, J., Hayashida, S., and Furukawa, K. (1992) Analysis of *bph* operon from the polychlorinated biphenyl-degrading strain of *Pseudomonas pseudoalcaligenes*. *J. Biol. Chem.*, **267**, 4844–4853.
- Takeda, H., Hara, N., Sakai, M., Yamada, A., Miyauchi, K. et al. (2004a) Biphenyl-inducible promoters in a polychlorinated biphenyl-degrading bacterium, *Rhodococcus* sp. RHA1. *Biosci., Biotechnol., Biochem.*, **68**, 1249–1258.
- Takeda, H., Yamada, A., Miyauchi, K., Masai, E., and Fukuda, M. (2004b) Characterization of transcriptional regulatory genes for biphenyl degradation in *Rhodococcus* sp. strain RHA1. *J. Bacteriol.*, **186**, 2134–2146.
- Tiedje, J. M., Quensen, J. F., Chee-Sanford, J., Schimel, J. P., and Boyd, S. A. (1994) Microbial reductive dechlorination of PCBs. *Biodegradation*, **4**, 231–240.
- Willenbrock, H., Salomon, J., Sokildel, R., Barken, K. B., Hansen, T. N. et al. (2009) Quantitative miRNA expression analysis: Comparing microarrays with next-generation sequencing. *RNA*, **15**, 2028–2034.