Genomic Analysis of *Rhodococcus* sp. Br-6, a Bromate Reducing Bacterium Isolated From Soil in Chiba, Japan

Kohei Ito[1,2], Masafumi Harada[1,2], Nobuyoshi Nakajima[3], Shigeki Yamamura[3,4], Masaru Tomita[1,2], Haruo Suzuki[1,2✉], Seigo Amachi[5]

1. Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, Japan
2. Faculty of Environment and Information Studies, Keio University, Fujisawa, Kanagawa, Japan
3. Center for Environmental Biology and Ecosystem Studies, National Institute for Environmental Studies, Tsukuba, Ibaraki, Japan
4. Center for Regional Environmental Research, National Institute for Environmental Studies, Tsukuba, Ibaraki, Japan
5. Graduate School of Horticulture, Chiba University, Matsudo City, Chiba, Japan

✉ Corresponding author: Haruo Suzuki, phone/fax number: +81-466-47-5099, email: haruo@sfc.keio.ac.jp

Abstract

Bromate is a byproduct of the ozone disinfection of drinking water. It is a genotoxic carcinogen and causes renal cell tumors in rats. Physicochemical removal of bromate is very difficult, making microbial reduction of bromate to bromide a promising approach to eliminate bromate from water. *Rhodococcus* sp. Br-6, isolated from soil, can efficiently reduce bromate by using acetate as an electron donor. We determined the draft genome sequence of *Rhodococcus* sp. Br-6 for the potential practical application of the bromate-reducing bacterium. Core genome phylogeny suggests that the Br-6 strain is most closely related to *R. equi*. The Br-6 genome contains genes encoding multiple isoforms of diaphorase, previously found to be involved in Br-6-mediated bromate reduction. The genes identified in the present study could be effective targets for experimental studies of microbial bromate reduction in the future.

Key words: Bromate reducing bacterium; *Rhodococcus* sp. Br-6; Genome; Phylogeny; Diaphorase.

Introduction

Chemical compounds are purified by multiple treatment steps at water treatment plants. Common filtration treatments do not remove fungal odor and agricultural chemicals so ozone treatment is commonly used. Bromate is a disinfection byproduct of ozone treatment and is formed from the bromide ion [1]. Bromate is a genotoxic carcinogen that causes renal cell tumors in rats. However, as long as ozone treatment is carried out, it is very difficult to completely remove bromate from the water. The microbial reduction of bromate to bromide is a promising way to remove bromate from drinking water and bromate reduction in biologically activated carbon (BAC) filters can reduce bromate [2]. Several studies have indicated that bacterial communities catalyze bromate reduction. However, to date, few bacteria from such communities have been isolated as pure cultures, and the mechanism of bacterial bromate reduction has not been elucidated. A soil bacterium *Rhodococcus* sp. Br-6 was previously isolated from soil in Matsudo-city, Chiba, Japan [3]. Strain Br-6 completely reduced 250 μM bromate within 4 days in the presence of acetate as an electron donor under microaerophilic conditions. Interestingly, the bromate reduction rate of strain Br-6 was much faster than that described for other bromate-reducing bacteria [3].

The genus *Rhodococcus* comprises more than 40 species, and is widely distributed in the environment [4]. *Rhodococcus* species have been used for various biotechnological applications including the production of acrylates and bioactive steroids, and fossil fuel desulfurization due to their biodegradative abilities [5]. We determined the draft genome

sequence of *Rhodococcus* sp. Br-6 to understand its detailed metabolism and for the future practical application of bromate-reducing bacteria. We used multiple databases to identify genes involved in bromate reduction by strain Br-6 and assigned functional protein annotations. We also compared the genome sequence of Br-6 with that of sequenced *Rhodococcus* strains to reconstruct the genome-based phylogenetic tree and to assess the conservation of Br-6 genes in other *Rhodococcus* genomes.

Materials and Methods

Genome sequencing and assembly

Rhodococcus sp. Br-6 was grown aerobically in a minimal salt medium containing acetate as the carbon source [3], and DNA was extracted using a DNeasy blood and tissue kit (Qiagen, Hilden, Germany). Whole-genome sequencing was performed using the Illumina MiSeq sequencing platform as per manufacturer's instructions. The sequencer produced 300-bp paired-end reads that were obtained from 550-bp inserts. Quality control and genome assembly were performed as described previously [6].

Genome annotation

The genome was annotated using Prokka v1.11 [7]. A total of 5,186 protein coding DNA sequences (CDSs) were predicted using Prokka for the Br-6 strain. We performed similarity searches of the 5,186 Br-6 proteins against the UniProt Reference Clusters UniRef90 (Release: 07-Sep-2016; Number of clusters: 44,448,796) [8] using BLASTP [9] with the E value cutoff of 1e-05 and assigned the most similar (best hit) protein sequence information. The genome was also annotated using KAAS (KEGG Automatic Annotation Server) [10] for which gene data sets from three *Rhodococcus* spp. (*R. jostii*, *R. erythropolis* PR4, and *R. opacus* B4) and the default set of organisms

(abbreviated as "hsa, dme, cel, ath, sce, cho, eco, nme, hpy, rpr, bsu, lla, cac, mge, mtu, ctr, bbu, syn, bth, dra, aae, mja, ape, rha, rer, and rop" at http://www.genome.jp/tools/kaas/) were selected.

Comparative genomic analysis

For comparative analysis with *Rhodococcus* sp. Br-6, RefSeq data for the genome sequences of *Rhodococcus* species including *R. defluvii* and *R. equi* [11][12] (Table 1) were downloaded from NCBI FTP Site (ftp://ftp.ncbi.nlm.nih.gov/genomes/) [13] in December 2016.

Homologous groups of genes (representing "protein families") from multiple *Rhodococcus* genomes were built using the pan-genome analysis pipeline Roary with the minimum BLASTP percentage identity of 80 [14]. Nucleotide sequence alignments for genes present in a single copy in every genome ("core-genome") were produced using MAFFT [15]. A phylogenetic tree for the concatenated core gene alignment was reconstructed using the GTR+CAT model of FastTree version 2.1.9 [16]. Based on previous phylogenetic studies [12][17], *R. triatomae*, *R. jostii*, *R. opacus*, and *R. erythropolis* were used as an outgroup for tree rooting. The "drop.tip" function of the package APE (version 5.0) in R (version 3.4.3) was used to remove the outgroup of the phylogenetic tree. The phylogenetic tree was drawn using FigTree version 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

To assess the conservation of the protein-encoding genes of strain Br-6 in other *Rhodococcus* strains, we performed TBLASTN searches with the E-value cutoff of 1e-5 [9] to compare strain Br-6 protein sequences with whole nucleotide sequences of 13 *Rhodococcus* strains. This approach can detect genes that were missed due to differences in gene finding algorithms.

Table 1. Genomic features of the *Rhodococcus* strains analyzed

Accession No.	Organism	% GC	Size (bp)	CDS
GCF_000196695.1	*R. equi* 103S	68.8	5,043,170	4,512
GCF_000738775.1	*R. defluvii* strain Ca11	68.7	5,134,337	4,791
GCF_002094305.1	*R. equi* strain DSM 20307	68.8	5,199,710	4,803
This study	*Rhodococcus* sp. Br-6	68.7	5,496,476	5,186
GCF_000341795.1	*R. triatomae* BKS 15-14	69.0	5,824,349	5,269
GCF_000454045.1	*R. erythropolis* CCM2595	62.5	6,371,421	5,828
GCF_000696675.2	*R. erythropolis* R138	62.3	6,806,506	6,130
GCF_000975175.1	*R. erythropolis* strain BG43	62.3	6,865,205	6,158
GCF_000010105.1	*R. erythropolis* PR4	62.3	6,895,538	6,437
GCF_001685605.1	*R. opacus* strain 1CP	67.1	8,637,535	7,921
GCF_000010805.1	*R. opacus* B4	67.6	8,834,939	8,203
GCF_000599545.1	*R. opacus* PD630	67.2	9,169,032	8,942
GCF_000014565.1	*R. jostii* RHA1	67.0	9,702,737	9,145

% GC: G+C content defined as $100 \times (G+C)/(A+T+G+C)$.
CDS: the number of protein-coding DNA sequences.

Results and Discussion

Genomic features

The draft genome sequence of *Rhodococcus* sp. Br-6 contains 23 contigs consisting of 5,496,476 bp, with a G+C content of 68.66 % (Table 1). Among the 13 *Rhodococcus* strains analyzed, genome size (Mb) and G+C content (%) varied from 5.0 (*R. equi* 103S) to 9.7 (*R. jostii*), and 62 (*R. erythropolis*) to 69 (*R. triatomae*), respectively. The genome size and G+C content of strain Br-6 were similar to those of the ingroup taxa (*R. equi* and *R. defluvii*).

Genome phylogeny

Previously, based on the 16S rRNA gene sequence analysis, the Br-6 strain was reported to be most closely related to *R. equi* [3]. In the present study, 839 core genes were used to infer the phylogenetic relationships between the 13 *Rhodococcus* strains. The phylogenetic tree, inferred from concatenation of the core genes (Figure 1), demonstrated that strain Br-6 and *R. equi* strains 103S and DSM 20307^T formed a clade to which *R. defluvii* is the sister species. The core genome phylogeny suggests that strain Br-6 belongs to the species *R. equi*.

Gene conservation

We performed TBLASTN searches [9] to assess the conservation of Br-6 genes in other *Rhodococcus* genomes. Of the 5,167 protein-coding genes detected in the Br-6 genome, 4,716 were conserved in all *R. equi* strains (103S and DSM 20307^T), and 3,671 were conserved in all the *Rhodococcus* strains examined. The largest numbers of Br-6 genes were conserved in the *R. equi* strains, ranging from 4,753 to 4,763, followed by *R. defluvii* with 4,452, and finally the outgroup, ranging from 4,178 (*R. triatomae*) to 4,295 (*R. jostii*). Thus, the conservation of Br-6 genes in the *Rhodococcus* strains roughly follows the pattern of their phylogenetic relationships (Figure 1).

Gene functions

The draft genome sequence of strain Br-6 contained 5,186 protein coding DNA sequences (CDS), of which 1,452 were functionally unknown, with a product name of "hypothetical protein" (Table S1). Of the 5,186 proteins, 4,947 (95.4%) matched with 4,894 unique records in the UniRef90 database, and 2,044 (39.4%) matched with 1,481 unique KEGG orthology (KO) identifiers.

We searched for genes involved in various biological functions of strain Br-6. The Br-6 genome contained at least 13 genes coding for *c*-type cytochromes and at least 19 genes putatively involved in arsenic resistance or arsenic metabolism: several genes encoding HTH-type, ArsR family transcriptional regulators, *arsB* encoding the membrane arsenic pump protein (locus_tag: Br6_00718), *aioB* encoding the arsenite oxidase subunit AioB precursor (locus_tag: Br6_00766), *arsA* encoding the arsenic pump-driving ATPase (locus_tag: Br6_00780 and Br6_04451), *arsC* encoding the arsenate reductase (locus_tag: Br6_01479), and a cluster of three genes (locus_tag: Br6_04407, Br6_04408, and Br6_04409; UniProt: E4WCD3, E4WCD4, and E4WCD5) which appear to form the *arsRBC* operon [18]. The draft genome contains a nitrate reductase gene cluster (locus_tag: Br6_04375, Br6_04376, Br6_04377, and Br6_04378) and denitrification regulatory proteins (locus_tag: Br6_04870 and Br6_04921), but does not contain nitrogenase genes. As expected by the fact that *Rhodococcus* is non-motile, the Br-6 genome did not contain any genes involved in motility.
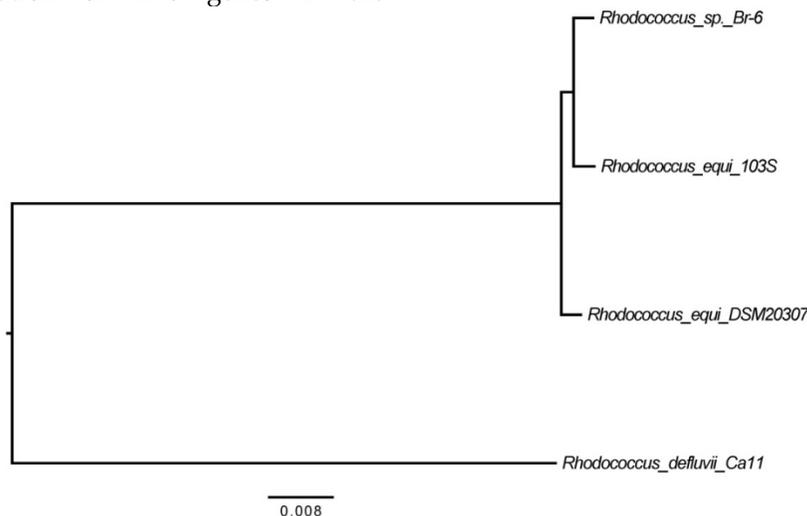
Figure 1. Phylogenetic tree obtained from a concatenated nucleotide sequence alignment of the 868 core genes of 13 *Rhodococcus* strains. The horizontal bar at the base of the figure represents 0.008 substitutions per nucleotide site. All the internodes exhibited highest local support values of 1.0.

Table 2. Br-6 strain *lpdC* gene homologues identified by BLASTP search.

locus_tag	Gene	Product	BLAST search statistics		
			Alignment length	% Identity	E-value
Br6_03586	*lpdC*	dihydrolipoyl dehydrogenase	467	100.0	0
Br6_02233	*mtr*	mycothione reductase	475	31.2	1.E-56
Br6_00169	*lpdA*	NAD(P)H dehydrogenase	472	29.5	5.E-51
Br6_01549	*merA*	mercuric reductase	487	28.3	3.E-39
Br6_00125	*cdr*	coenzyme A disulfide reductase	347	28.0	1.E-21
Br6_00870	*ahpF*	alkyl hydroperoxide reductase subunit F	201	29.4	3.E-08
Br6_03974	*trxB_2*	thioredoxin reductase	333	23.7	1.E-07
Br6_03365	*thcD_3*	rhodocoxin reductase	199	28.6	2.E-07
Br6_00086	*camA*	putidaredoxin reductase	191	26.2	6.E-07
Br6_03025	*yumC*	ferredoxin--NADP reductase 2	320	25.0	2.E-06
Br6_00577	*thcD_1*	rhodocoxin reductase	239	28.5	2.E-06

Tamai *et al.* (2016) suggested that diaphorase is involved in bromate reduction by strain Br-6, and that strain Br-6 has multiple isoforms of diaphorase [3]. Some dihydrolipoyl dehydrogenase (DLD) enzymes have NADH-dependent diaphorase activity [19]. The Br-6 genome contains the *lpdC* gene, which encodes DLD (locus_tag: Br6_03586; UniProt: E4W8R5). The DLD protein sequence was used as a query in a BLASTP search (E-value < 1e-5) against all the strain Br-6 predicted protein sequences to identify DLD homologues. The *lpdC* protein has low percentage identity (ranging from 25% to 31%) with all 10 reductase encoding homologs (Table 2), including the *lpdA* gene, annotated as "NAD(P)H dehydrogenase" (locus_tag: Br6_00169) or "flavoprotein disulfide reductase" (UniProt: E9T344), and the *merA* gene, annotated as "mercuric reductase" (Br6_01549) or "putative dihydrolipoyl dehydrogenase" (UniProt: E4WE66). The *lpdC*, *lpdA*, and *merA* genes were also annotated as "dihydrolipoamide dehydrogenase [EC:1.8.1.4]" (KEGG: K00382). These results suggest that these homologous genes encode multiple diaphorase isoforms. The genes identified here could be effective targets for future experimental studies.

Nucleotide sequence accession number

The whole *Rhodococcus* sp. Br-6 genome shotgun sequence has been deposited at DDBJ/EMBL/GenBank under the accession number BDGK00000000. The version described in this paper is the second version, BDGK02000000.

Supplementary Material

Table S1. *Rhodococcus* sp. strain Br-6 genes. The columns are as follows: locus_tag, length in amino acids (Laa), KEGG orthology identifiers (ko), gene and product names, and the most similar sequence annotation in the UniRef90 database (FASTA header and organism name).
http://www.jgenomics.com/v06p0122s1.csv

Acknowledgments

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

1. von Gunten U. Ozonation of drinking water: Part II. Disinfection and by-product formation in presence of bromide, iodide or chlorine. *Water Res*. 2003;37(7):1469-1487.
2. Kirisits MJ, Snoeyink VL, Inan H, Chee-sanford JC, Raskin L, Brown JC. Water quality factors affecting bromate reduction in biologically active carbon filters. *Water Res*. 2001;35(4):891-900.
3. Tamai N, Ishii T, Sato Y, et al. Bromate Reduction by Rhodococcus sp. Br-6 in the Presence of Multiple Redox Mediators. *Environ Sci Technol*. 2016;50(19):10527-10534.
4. Larkin MJ, Kulakov LA, Allen CC. Biodegradation and Rhodococcus – masters of catabolic versatility. *Curr Opin Biotechnol*. 2005;16(3):282-290.
5. Gürtler V, Mayall BC, Seviour R. Can whole genome analysis refine the taxonomy of the genus Rhodococcus? *FEMS Microbiol Rev*. 2004;28(3):377-403.
6. Yuliana T, Nakajima N, Yamamura S, Tomita M, Suzuki H, Amachi S. Draft Genome Sequence of Roseovarius sp. A-2, an Iodide-Oxidizing Bacterium Isolated from Natural Gas Brine Water, Chiba, Japan. *J genomics*. 2017;5:51-53.
7. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068-2069.
8. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, UniProt Consortium. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*. 2015;31(6):926-932.
9. Boratyn GM, Camacho C, Cooper PS, et al. BLAST: a more efficient report with usability improvements. *Nucleic Acids Res*. 2013;41(Web Server issue):W29-33.
10. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*. 2007;35(Web Server):W182-W185.
11. Letek M, González P, Macarthur I, et al. The genome of a pathogenic rhodococcus: cooptive virulence underpinned by key gene acquisitions. *PLoS Genet*. 2010;6(9):e1001145.
12. Anastasi E, MacArthur I, Scortti M, Alvarez S, Giguère S, Vázquez-Boland JA. Pangenome and Phylogenomic Analysis of the Pathogenic Actinobacterium Rhodococcus equi. *Genome Biol Evol*. 2016;8(10):3140-3148.
13. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2016;44(D1):D7-D19.
14. Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. 2015;31(22):3691-3693.
15. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30(14):3059-3066.

16. Price MN, Dehal PS, Arkin AP, Rojas M, Brodie E. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. Poon AFY, ed. *PLoS One.* 2010;5(3):e9490.

17. Creason AL, Davis EW, Putnam ML, Vandeputte OM, Chang JH, Chang JH. Use of whole genome sequences to develop a molecular phylogenetic framework for Rhodococcus fascians and the Rhodococcus genus. *Front Plant Sci.* 2014;5:406.

18. Achour-Rokbani A, Cordi A, Poupin P, Bauda P, Billard P. Characterization of the ars gene cluster from extremely arsenic-resistant Microbacterium sp. strain A33. *Appl Environ Microbiol.* 2010;76(3):948-955.

19. Kianmehr A, Mahdizadeh R, Oladnabi M, Ansari J. Recombinant expression, characterization and application of a dihydrolipoamide dehydrogenase with diaphorase activity from Bacillus sphaericus. *3 Biotech.* 2017;7(2):153.