

Note (Genome Announcement)

Genome sequence of *Caballeronia sordidicola* strain PAMC 26510 isolated from *Psoroma* sp., an Antarctic lichen

Jhung Ahn Yang¹, Soon Gyu Hong², and Hyun-Myung Oh^{1*}

¹Department of Marine-Bio Convergence Science, Specialized Graduate School Science & Technology Convergence, Pukyong National University, Busan 48547, Republic of Korea

²Division of Polar Life Sciences, Korea Polar Research Institute, Incheon 21990, Republic of Korea

남극 지의류에서 분리한 *Caballeronia sordidicola* 균주 PAMC 26510의 유전체 서열 분석

양정안¹ · 홍순규² · 오현명^{1*}

¹부경대학교 과학기술융합전문대학원, ²극지연구소

(Received June 1, 2017; Accepted June 8, 2017)

Caballeronia sordidicola strain PAMC 26510 was isolated from *Psoroma* sp., a lichen material, collected from Barton Peninsula of King George Island in Antarctica. The draft genome sequence of PAMC 26510 consisted of 224 contigs and they was 7,872,143 base pairs with 59.7% G+C content. The genome included 7,580 protein coding sequences and 6 ribosomal RNA genes and 46 tRNA genes. The strain PAMC 26510 is also a metabolic generalist as we have observed in previous genomic studies in the arctic strain of *Caballeronia sordidicola*. The draft genomic sequences of PAMC 26510 had six CRISPR arrays on six contigs, and there were two clusters of CRISPR-associated genes that were linked with respective CRISPR arrays.

Keywords: *Caballeronia sordidicola*, *Psoroma* species, Antarctic, lichen

The strain PAMC 26510 was identified as *Burkholderia sordidicola* based on 16S rRNA sequence analysis and the strain was associated with a *Psoroma* species (Lee *et al.*, 2012,

2014). The genus *Burkholderia* contains animal and plant pathogenic species, endophytes, and legume nodulators as well as environmental bacteria in soil and water (Angus *et al.*, 2014). However the *Burkholderia* was proposed for animal and plant pathogens and a new generic name *Paraburkholderia* was coined to include so-called environmental isolates by the molecular typing and phylogenetic analysis (Sawana *et al.*, 2014). The *Caballeronia* genus was proposed to include twelve *Burkholderia* and *Paraburkholderia* species with distinct phylogenetic lineage from the main phylogenetic lineages of *Burkholderia* and *Paraburkholderia* (Dobritsa and Samadpour, 2016). Members of the genus *Caballeronia* has been isolated from soil, wastewater-treatment system, fungus, or moss (Dobritsa and Samadpour, 2016). So the strain PAMC 26510 was assigned from *Burkholderia sordidicola* to *Caballeronia sordidicola* during the genome analysis.

Here we report the draft genome sequence of the strain PAMC 26510 (Table 1). It was isolated from *Psoroma* sp., a lichen collected from Barton Peninsula of King George Island in the Antarctic (Lee *et al.*, 2014). The draft genome sequence was determined using Illumina and was assembled with Ray

*For correspondence. E-mail: marinebio@pknu.ac.kr;
Tel.: +82-51-629-6869; Fax: +82-51-629-6865

Table 1. *Caballeronia sordidicola* strain PAMC 26510 genome assembly and its general features

Item	Description
Genome Assembly Data	
Assembly method	Ray v. 2.3.1
Genome coverage	54.5x
Sequencing technology	Illumina
MIGS Data	
Assembly	Ray v. 2.3.1
Biome	Lichen
Collection_date	2010-02-06
Env_package	Missing
Feature	Lichen
Geo_loc_name	Antarctica; South Shetland Islands; King George Island; Barton Peninsula
Investigation_type	bacteria_archaea
Isol_growth_condt	NA
Lat_lon	-62.22605 -58.78356667
Material	Lichen
Project_name	<i>Caballeronia sordidicola</i> PAMC 26510
Seq_meth	Illumina
Experimental_factor	NA
Depth	NA
Alt_elev	NA
Num_replicons	NA
Ref_biomaterial	NA
Source_mat_id	PAMC 26510
Pathogenicity	NA
Biotic_relationship	Symbiont
Trophic_level	chemoheterotroph
Rel_to_oxygen	aerobe
Assembly_name	NA
Finishing_strategy	draft;54.5x coverage; 224 contigs
Annot_source	NA

software (v.2.3.1) (Boisvert *et al.*, 2010). Assembly of genome analysis were carried out with 23-mer search option, which produced longest N50 (121,862 bp) and 225 contigs. Bedtools (bedtools genomecov; aka “genomeCoverageBed” v. 2.25.0) (Quinlan and Hall, 2010) was used to calculate genome coverage (Kim *et al.*, 2017) and aligned Illumina reads corresponded to 54.5x coverage depth (Table 1). Finally 224 contigs were submitted to GenBank and the draft genome of PAMC 26510 was comprised of 7,872,143 base pairs with 59.7% GC content. One contig was filtered because it was

contaminant sequence with no hits to any known nucleotides in the GenBank database.

The genome included 7,580 protein coding sequences annotated by the methods described in the previous study (Kim *et al.*, 2017). Annotation of some genes of interest was confirmed using BLAST search tool against National Center for Biotechnology Information (NCBI) database. The genome had 6 ribosomal RNA genes (four 5S rRNA genes) and 46 tRNA genes. Major genes covered varieties of enzymes involved in metabolism of carbohydrate, amino acid, nitrogen/sulfur, aromatic compound, and vitamins. Glucose metabolism genes included Embden-Meyerhof-Parnas Pathway, Entner-Doudoroff Pathway, and Pentose phosphate pathway. Glycogen or starch metabolic genes and maltose phosphorylase and β -phosphoglucomutase were presented. Chitobiose could not be utilizable but N-acetylglucosamine could be processed by deacetylase and deaminases to glucose. PAMC 26510 had two kinds of lactate dehydrogenases that can form D-lactate and L-lactate from pyruvate from carbohydrate metabolisms. Acetyl-CoA fermentation to butyric acid is impaired because butanoyl-CoA dehydrogenase was missing for crotonyl-CoA that was converted from acetyl-CoA. Propanoate could be metabolized to succinate through propanoyl-CoA formation. Environmental response genes and regulators included stress response proteins, membrane transporters, antibiotic resistance genes, multidrug resistance efflux pumps, and heavy metal resistances. Twenty three phage-related genes were observed and there were three regions with transposable elements. Overall metabolic features were not different from PAMC26592 from the same species with poles apart location and different host lichen (Kim *et al.*, 2017). By using the CRISPRDetect program for CRISPR array analysis (Biswas *et al.*, 2016), there are several contigs that have clustered regularly interspaced short palindromic repeat (CRISPR) sequence on them, which is differentiating PAMC 26510 from previous arctic strain PAMC 26592 that had no CRISPR array (Kim *et al.*, 2017). There are six CRISPR arrays by CRISPRDetect program and two contigs shared the same spacer sequences (GenBank Accessions: NBTY01000155.1 and NBTY01000222.1) (Table 2). CRISPR-related protein genes were adjacent to two contigs respectively (NBTY01000120.1 and NBTY01000155.1).

Table 2. CRISPR array list from draft genome of *C. sordidicola* strain PAMC 26510 in this study: (.) consensus sequences, (-) missing sequences

Contig (Accession)	Position	Repeat	%identity	Spacer	Repeat sequence	Spacer sequence	Insertion/Deletion [position]
NBTY01000120.1	1	32	100	34	CATGCCACCCTACCCATGCCACCCTAATTTAGAT	
	67	32	100	34	GAGCGGGCAATCGACGAGGTTCTGAAGACCGGGGA	
	133	32	100	35	TCAAACCTGACTTCGGACAAATCGATTTCGACGGT	
	200	32	100	35	GGTTGCTAGGGCGTGGGCATACGCAAAACGCAAC	
	267	32	100	34	ACCTCGGCTTTAGAGGATGCAGCGAAAGCCGGAC	
	333	32	100	36	TATAGGTTTCGACGCGCTACTCTGGGTCTACATATTC	
	401	32	100	35	AGAGCACGGCCAACTCGTCCGACTACGGTCGTAGC	
	468	32	100	34	CGCTGCCGACCGTATTGTCGATGAGGTGATCGAG	
	534	32	100	33	CAATCCGGCCAGTTTCTCGGCGCCATCGTGGTA	
	599	32	100	33	CGCGCATTTCACCGCCGATATCGTTGCCATATGA	
	664	32	100	35	GATGGTAGCCGTTGCCGTTGAATCAAACGAATGGT	
	731	32	100	0		
	12	32	100	35		GTTTCAACCCACGCGCCCGTGAAGAGCGCGAC	
NBTY01000120.1	9725	31	100	35	GTCCAAGCGGTCCGAGCTTCAACGTGCCAAACG	
	9659	31	100	34	TTGCATCCAAAGCCCGTTGAGCCAATAAGCCGAC	
	9594	31	100	34	GCAAGAACTCGCGGGCTCCACCGCGGTGAGGT	
	9529	31	100	35	GACACAGTCAGCAAACTAGACGTCACGCGCCCGAG	
	9463	31	100	34	GCTTGATCTCAGGGCGCGCGCTCGAACTGTGG	
	9398	31	100	36	GCCGCCGCAAAAAACAAGCTTTGATTGACTACGA	
	9331	31	100	35	CGACAATTCGTCGAGACGTGCGATACATCGGCTAA	
	9265	31	96.8	35C.....	TGAATTGTGCTGGCTTGGCTATCGCCAAAGCGC	
	9199	31	100	36	ACCATCATGGCCGCGTGTGGACTCAGGCGGGCAC	
	9132	31	100	36	GCCACTGGCCGGGCTGTAGCGAAGCCATTTACACC	
	9065	31	100	35	TATCGGGCCCTATGGGTGTATGGCAGCAAGCCGTT	CGCGGGTG [9039]
	8991	31	87.1	0C.....A...T...G		
	12	31	98.7	35		GTCGCGTCTCAGGGCGCGTGGGTTGAAAC	
NBTY01000155.1	55095	28	100	32	AACCAGACACCAGCAGCGCAAAAAACATGGG	
	55155	28	100	32	TGACGATACACCAAGCGCGGATCTGTACGCC	
	55215	22	78.6	0		
3	28	92.9	32		GTTTGCTGCCGCAAGGCAGCTCAGAAA		
NBTY01000222.1	1	26	92.9	32	--.....	ATGAGGATCGTGGACGGCAGCTACCAACTT	
	59	28	100	32	AGCGCGGCGTCATCGAAATGCAACTCGATGTA	
	119	28	100	32	GTCTTGCTCAACGACATAGCAGCGCCGATCG	
	179	28	100	32	ACCCGGAAGGCATCGAAAATTGAAGCCATT	
	239	28	100	32	ACAAAATATCGACCGATCAGCGCGTGGCCGC	
	299	28	100	32	AGAAGCCGAGGAGAGCTTTGAGGTGGCCGCA	
	359	28	100	32	TGCAGCAGCCGACAGCTTTGTCGATGATGTC	
	419	28	89.3	0C.....G.....C		
8	28	97.8	32		GTTTGCTGCCGCAAGGCAGCTCAGAAA		
NBTY01000194.1	56456	31	100	32	ATCATCTCGTCGAGCAGCGGCTCCGCATCGTC	
	56519	31	100	34	TTCATGTAAAGCGGACCACTCATCGCCGCTTTTT	
	56584	31	100	33	GATATTCGGTGCCGGCGTGGTTCTGGGGACGG	
	56648	31	100	35	AGCTTGAGGCCCTATCTATATGGGCGTTGAAACC	
	56714	31	100	36	AGCTTGCCCTGATATGTGAGTATCCGCGCCCTCCGC	
	56781	31	100	36	GATGACAGTGACCGGCGGGAGTGGCACCAGGCAC	
	56848	31	100	35	GTCGACGGCCGTAGCCGGTGTATCTTCCCGCGCT	
	56914	31	100	35	GTCCAAGCGGTCGCCAGTCTTCAACGTGCCAAACG	
	56980	31	100	34	TTGCATCCAAAGCCCGTTGAGCCAATAAGCCGAC	
	57045	31	100	34	GCAAGAACTCGCGGGCTCCACCGCGGTGAGGT	
	57110	31	100	35	GACACAGTCAGCAAACTAGACGTCACGCGCCCGAG	
	57176	30	96.8	0	-	Deletion [57206]
	57206	31	64.5	4	CG..TT..AT.....C..C...CT	GTGG	Deletion [57241]
	57241	31	100	36	GCCGCCCGCAAAAAACAAGCTGTTGATTGACTACGA	
	57308	31	100	0		
15	31	97.4	30		GTCGCGTCTCAGGGCGCGTGGGTTGAAAC		
NBTY01000066.1	123058	32	100	36	TTCAATAGTTCAAGTTTCATTAACCTCCCAATAAGC	
	122990	32	100	36	CATGCGAGCCACCGTAATCCGTCCCGCCGATACCG	
	122922	32	100	35	GAATCCAGCAATGCGCGTCATCAGTTCGGCGCGGC	
	122855	32	96.9	35T.....	TGAAGCGCATCAACTGACCGCGGTGACGGAGCAA	
	122788	32	100	35	TGCCAACGCTCGACAAGGCCGATGCGGTTATCAGC	
	122721	32	100	34	GTGGGCAATATCTACGGTGTCAATCGCGATCAAAG	
	122655	32	90.6	0A..C.G...		C [122625]
	7	32	98.2	35		GTCGCGTCTTTCAGGGCGCGTGGGTTGAAAC	

Nucleotide sequence accession number(s)

This whole genome shotgun project has been deposited under the accession NBTY00000000 at GenBank. The version described in this paper is version NBTY 01000000 (NBTY 01000001-NBTY01000224). Small subunit rRNA sequence of *Caballeronia sordidicola* strain PAMC 26510 was deposited at GenBank under accession number KJ606826 according to previous study.

적 요

Caballeronia sordidicola 균주 PAMC 26510은 남극 킹조지섬의 바톤반도에서 채집된 지의류인 *Psoroma* sp.에서 분리되었다. 균주 PAMC 26510의 초안 유전체 서열은 224개의 콘티그로 이루어졌으며, 총 7,872,143 염기쌍은 59.7% G+C 함량을 나타냈다. 유전체는 7,580개의 단백질 유전자, 6개의 rRNA 유전자와 46개의 tRNA 유전자를 포함하였다. 균주 PAMC 26510는 이전에 연구한 *Caballeronia sordidicola*에 속하는 북극 균주와 마찬가지로 대사능력에 있어서 광범위한 능력을 가지고 있는 것으로 여겨진다. PAMC 26510의 초안 유전체는 6개의 CRISPR arrays를 각각의 콘티그 상에 가지고 있으며, 이중 두 개의 콘티그 상에 CRISPR-관련 유전자가 연결되어 있었다.

Acknowledgements

This work was supported by the Korea Polar Research Institute (KOPRI, Grant PE17090) and was supported by a grant from the Collaborative Genome Program (20140428) funded by the Ministry of Oceans and Fisheries, Korea.

References

- Angus, A.A., Agapakis, C.M., Fong, S., Yerrapragada, S., Estrada-de los Santos, P., Yang, P., Song, N., Kano, S., Caballero-Mellado, J., de Faria, S.M., *et al.* 2014. Plant-associated symbiotic *Burkholderia* species lack hallmark strategies required in mammalian pathogenesis. *PLoS One* **9**, e83779.
- Biswas, A., Staals, R.H.J., Morales, S.E., Fineran, P.C., and Brown, C.M. 2016. CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics* **17**, 356.
- Boisvert, S., Laviolette, F., and Corbeil, J. 2010. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J. Comput. Biol.* **17**, 1519–1533.
- Dobritsa, A.P. and Samadpour, M. 2016. Transfer of eleven species of the genus *Burkholderia* to the genus *Paraburkholderia* and proposal of *Caballeronia* gen. nov. to accommodate twelve species of the genera *Burkholderia* and *Paraburkholderia*. *Int. J. Syst. Evol. Microbiol.* **66**, 2836–2846.
- Kim, J., Kwon, K.K., Kim, B.K., Hong, S.G., and Oh, H.M. 2017. Genome sequence of *Caballeronia sordidicola* strain PAMC 26592 isolated from an arctic lichen species. *Korean J. Microbiol.* **53**.
- Lee, Y.M., Kim, G., Jung, Y.J., Choe, C.D., Yim, J.H., Lee, H.K., and Hong, S.G. 2012. Polar and Alpine Microbial Collection (PAMC): a culture collection dedicated to polar and alpine microorganisms. *Polar Biol.* **35**, 1433–1438.
- Lee, Y.M., Kim, E.H., Lee, H.K., and Hong, S.G. 2014. Biodiversity and physiological characteristics of Antarctic and Arctic lichens-associated bacteria. *World J. Microbiol. Biotechnol.* **30**, 2711–2721.
- Quinlan, A.R. and Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842.
- Sawana, A., Adeolu, M., and Gupta, R.S. 2014. Molecular signatures and phylogenomic analysis of the genus *Burkholderia*: proposal for division of this genus into the emended genus *Burkholderia* containing pathogenic organisms and a new genus *Paraburkholderia* gen. nov. harboring environmental species. *Front Genet.* **5**, 429.