

Genome Announcement

Genome sequence of *Caballeronia sordidicola* strain PAMC 26592 isolated from an arctic lichen species

Junghye Kim¹, Kae Kyoung Kwon², Byung Kwon Kim³, Soon Gyu Hong⁴, and Hyun-Myung Oh^{1*}

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

²Marine Biotechnology Research Division, Korea Institute of Ocean Science and Technology, Ansan 15627, Republic of Korea

³OmicPia, Co. Ltd., Daejeon 34867, Republic of Korea

⁴Division of Polar Life Sciences, Korea Polar Research Institute, Incheon 21990, Republic of Korea

북극 지의류에서 분리한 *Caballeronia sordidicola* 균주 PAMC 26592의 유전체 서열 분석

김정희¹ · 권개경² · 김병권³ · 홍순규⁴ · 오현명^{1*}

¹부경대학교 과학기술융합전문대학원, ²해양과학기술원, ³(주) OmicsPia, ⁴극지연구소

(Received March 3, 2017; Revised March 6, 2017; Accepted March 7, 2017)

Caballeronia sordidicola strain PAMC 26592 was isolated from *Umbilicaria* sp., a lichen material collected from Svalbard Archipelago in the Arctic Ocean. We report the draft genome sequence of the strain PAMC 26592, a metabolic generalist. As we have observed in previous genomic studies in the genus *Caballeronia* draft genomic sequences of PAMC 26592 had an assortment of genes of ecological importance and of biotechnical potentials, which include diverse metabolic genes for carbohydrates, aromatic compounds, amino acids, and vitamins, and genes for nitrogen / sulfur metabolisms, stress responses, membrane transporters, antibiotic resistance, and heavy metal resistance.

Keywords: *Caballeronia sordidicola*, *Umbilicaria*, arctic, genome, lichen

Burkholderia is such a diverse and dynamic genus that they contain pathogenic species as well as species that form complex interactions with their host organisms, plants (Angus *et al.*,

2014). It has been proposed to split the genus *Burkholderia* into two genera according to phylogenetic clustering: the genus retaining the name *Burkholderia* consisted mainly of animal and plant pathogens and the genus *Paraburkholderia* included so-called environmental bacteria (Dobritsa and Samadpour, 2016). The genus *Caballeronia* has 12 species from genera *Burkholderia* and *Paraburkholderia* (Dobritsa and Samadpour, 2016) that includes *Caballeronia sordidicola*. In previous study the strain PAMC 26592 had reportedly been associated with a *Umbilicaria* species (Lee *et al.*, 2014) and strain PAMC 26592 was deposited under Polar and Alpine Microbial Collection (PAMC) project by Korean Polar Research Institute (KOPRI) (Lee *et al.*, 2012). The closest known species of strain PAMC 26592 was *Caballeronia sordidicola* based on the 16S rRNA sequence analysis (Lee *et al.*, 2014).

Here we report the draft genome sequence of the strain PAMC 26592 (Table 1). Strain PAMC 26592 was isolated from a lichen collected from Svalbard Archipelago in the Arctic Ocean (Lee *et al.*, 2014). The draft genome sequence was determined using the Illumina sequencing and assembly

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

program was Ray (v.2.3.1) (Boisvert *et al.*, 2010). Assembly of genome analysis was carried out with k-mer at 23 where N50 and contigs were longest. Bowtie-2 (Version 2.2.6) was used to generate a *.sam file (Langmead and Salzberg, 2012) and samtools (v. 0.1.19) was used to generate a *.bam file (Li *et al.*, 2009). Bedtools (bedtools genomcov; aka “genomeCoverageBed” v. 2.25.0) was used to calculate genome coverage (Quinlan and Hall, 2010); Illumina reads aligned in *.bam file corresponded to 64.9X coverage depth (Table 1). Three contigs with no CDS features and thirty contigs less than 500 base pairs in lengths were excluded. Finally 95 contigs were submitted to GenBank and N50 value was 218,950. The draft genome of PAMC

26592 comprised of 7,765,253 base pairs with %GC content was 59.8.

All protein coding sequences were 6,850 in number out of 7,259 genes that were annotated by RAST server (Aziz *et al.*, 2008) according to previous method (Yang *et al.*, 2016). Some genes of interest were confirmed using blast using National Center for Biotechnology Information (NCBI) database. Noncoding genes including 16S rRNA gene sequences were 10 in number and tRNA-coding genes were 52 in number. Most abundant features of metabolic categories by the RAST system included myriads of carbohydrate genes, amino acid metabolisms, aromatic compound metabolisms, vitamins biosynthesis, nitrogen / sulfur genes, stress genes, membrane transporters, antibiotic resistance genes (β -lactamase, fluoroquinolone resistance, and multidrug resistance efflux pumps), and heavy metal resistance genes (arsenic, chromium, cobalt, copper, mercury, zinc). More than 752 carbohydrate-related genes could be annotated and butanol / butyrate biosynthesis genes plus propionate genes. Genes for one carbon metabolisms included D-arabino-3-hexulose-6-phosphate formaldehyde lyase in ribulose mono-phosphate pathways and serine-glyoxylate cycle genes (glyoxylate regeneration, serine hydroxymethyl transferase, and associated PhaCAB genes). Plant-like oxidative C₂ cycle genes and sugar-utilizing genes covered diversity of substrates; arabinose, ribose, xylose, fucose, galactarate, galactonate, gluconate, N-acetylglucosamine / chitin, inositol, mannose, mannitol, lactose, trehalose, and glycogen. The membrane transporter genes were 25 in number and protein secretion systems were diversely presented (Type I-VII) with no sugar phosphotransferase system genes. Genes for potassium homeostasis, nitrogen, phosphate, and iron regulation were could be found. Such crowded genes of diverse metabolic enzymes may well demarcate the strain PAMC 26592 as a featured metabolic generalist living with its lichen host adapted in Arctic environment. By using the CRISPRFinder program analysis (Grissa *et al.*, 2008), there is no contig that has clustered regularly interspaced short palindromic repeat (CRISPR) sequence of the strain PAMC 26592. Limited number of 11 phage-related genes and no transposable elements were observed.

Table 1. *Caballeronia sordidicola* strain PAMC 26592 genome assembly and its general features

Item	Description
Genome Assembly Data	
Assembly Method	Ray v. 2.3.1
Genome Coverage	64.9X
Sequencing Technology	Illumina
MIGS Data	
Assembly	Ray v. 2.3.1
Biome	Lichen
Collection_date	2010-07-13
Env_package	missing
Feature	Lichen
Geo_loc_name	Europe: Norway, Svalbard
Investigation_type	bacteria_archaea
Isol_growth_condt	NA
Lat_lon	78.912 11.951433333
Material	Lichen
Project_name	<i>Burkholderia sordidicola</i> PAMC 26592
Seq_meth	Illumina
Experimental_factor	NA
Depth	NA
Alt_elev	NA
Num_replicons	NA
Ref_biomaterial	NA
Source_mat_id	PAMC 26592
Pathogenicity	NA
Biotic_relationship	Symbiont
Trophic_level	chemoheterotroph
Rel_to_oxygen	aerobe
Assembly_name	NA
Finishing_strategy	draft;64.9X coverage; 95 contig
Annot_source	NA

Nucleotide sequence accession number(s)

This Whole Genome Shotgun project *Caballeronia sordidicola*

strain PAMC 26592 has been deposited under the accession MSRG00000000 at DDBJ / ENA / GenBank. The version described in this paper is version MSRG01000000. Small subunit rRNA sequence of PAMC 26592 was deposited under GenBank accession number AF512827 in previous investigation described in the above text.

Acknowledgements

This work was supported by the Research Grant of Pukyong National University.

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.
- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., *et al.* 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75.
- 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.
- Grissa, I., Vergnaud, G., and Pourcel, C. 2008. CRISPRcompar: a website to compare clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **36**, W145–W148.
- Langmead, B. and Salzberg, S.L. 2012. Fast gapped-read alignment with bowtie 2. *Nat. Methods* **9**, 357–359.
- 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.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, G.P.D.P. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
- Quinlan, A.R. and Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842.
- Yang, J.A., Kang, I., Moon, M., Ryu, U.C., Kwon, K.K., Cho, J.C., and Oh, H.M. 2016. Complete genome sequence of *Celeribacter marinus* IMCC12053^T, the host strain of marine bacteriophage P12053L. *Marine Genomics.* **26**, 5–7.