



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

# Draft genome sequence data of *Paenibacillus Polymyxa* strain TH2H2, isolated from a tomato flower in Korea

Gyeongjun Cho<sup>a</sup>, Da-Ran Kim<sup>b</sup>, Chang-Wook Jeon<sup>a</sup>,  
Youn-Sig Kwak<sup>b,\*</sup>

<sup>a</sup> Division of Applied Life Science (BK21Plus), Gyeongsang National University, Jinju 52828, Republic of Korea

<sup>b</sup> Department of Plant Medicine and RILS, Gyeongsang National University, Jinju 52828, Republic of Korea

## ARTICLE INFO

## Article history:

Received 20 March 2020

Revised 29 May 2020

Accepted 2 June 2020

Available online 10 June 2020

## Keywords:

Antibiotic genes

Biocontrol

*Paenibacillus polymyxa*

Phytophthora

Tomato

## ABSTRACT

Members of the genus *Paenibacillus* are known for their production of useful substances, and some species of the genus are recognized to be plant growth-promoting rhizobacteria. *Paenibacillus polymyxa* TH2H2, isolated from a tomato flower, had antifungal activity. Here, the draft genome sequence of *Paenibacillus polymyxa* TH2H2 is reported. The 5,983,104-bp genome, with a G+C content of 45.31%, comprised 5,221 protein-coding genes, 64 ribosomal RNA and 100 transfer RNA. Three intact antibiotic biosynthesis gene clusters were identified using antiSMASH. These encoded the antifungal agent fusaricidin and two antibacterial agents, tridecaptin and polymyxin. Sequence data have been deposited in the DDBJ/ENA/GenBank database under the accession number RPDG01000000. The version described in this paper is RPDG00000000.1. The BioProject ID in the GenBank database is PRJNA505713.

© 2020 The Authors. Published by Elsevier Inc.  
This is an open access article under the CC BY license.  
(<http://creativecommons.org/licenses/by/4.0/>)

\* Corresponding author: Youn-Sig Kwak  
E-mail address: [kwak@gnu.ac.kr](mailto:kwak@gnu.ac.kr) (Y.-S. Kwak).

## Specifications table

Subject	Microbiology
Specific subject area	Phytobiome and a keystone taxon
Type of data	Table, figure
How data were acquired	Genome sequencing with MiSeq paired-end protocol at Chunlab Inc. Republic of Korea
Data format	Raw and Analyzed
Parameters for data collection	Bacterial genomic DNA was extracted from a culture of <i>Paenibacillus polymyxa</i> TH2H2
Description of data collection	Genome features (Table 1), genome map (Fig. 1a), EggNOG functional category (Fig. 1b), genome annotation and COG (Supplementary data 1), secondary metabolite gene cluster (Supplementary data 2)
Data source location	<i>Paenibacillus polymyxa</i> TH2H2 was obtained from a healthy tomato flower in the city of Jinju, Republic of Korea (GPS: 35.2109N, 128.1164E)
Data accessibility	Repository name: DDBJ/ENA/GenBank Data identification number: RPDG01000000. The version described in this paper is RPDG00000000.1 Direct URL to data: <a href="https://www.ncbi.nlm.nih.gov/nucore/RPDG00000000.1">https://www.ncbi.nlm.nih.gov/nucore/RPDG00000000.1</a> The BioProject ID in GenBank is PRJNA505713 ( <a href="https://www.ncbi.nlm.nih.gov/search/all/?term=PRJNA505713">https://www.ncbi.nlm.nih.gov/search/all/?term=PRJNA505713</a> ) The data processing R code is available in a GitHub repository ( <a href="https://github.com/gyeongjunCho/TH2H2_draft_genome">https://github.com/gyeongjunCho/TH2H2_draft_genome</a> )
Related research article	Da-Ran Kim, Jun-Taek Lee, Hye sun Kim, Chang Wook Jeon, Youn-Sig Kwak. Selection of biocontrol agent of tomato gray mold disease from flower and pollinator hive. The Korean Journal of Pesticide Science 21 (2017) 90-96. <a href="https://doi.org/10.7585/kjps.2017.21.1.90">https://doi.org/10.7585/kjps.2017.21.1.90</a>

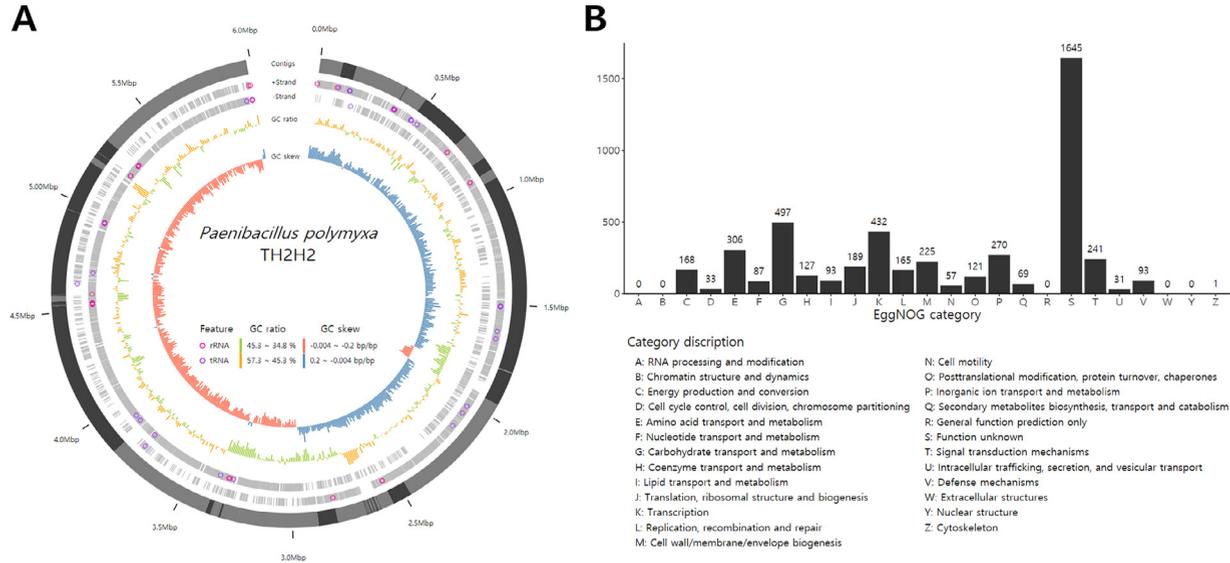
## Value of the data

- The complete genome sequence of *P. polymyxa* TH2H2 provides essential information about the strain that can be applied to plant protection research and biological control of plant.
- In the genome of *P. polymyxa* TH2H2, 37 antibiotic-related metabolite gene clusters were predicted, indicating that the strain could be valuable in investigations of plant-microbe interactions.
- *P. polymyxa* TH2H2 genome data provides information about species of the genus *Paenibacillus* that will be useful to the wider microbial research community.

## Data Description

Species of the genus *Paenibacillus* are rod-shaped, aerobic or facultatively anaerobic, endospore-forming bacteria with Gram-positive and Gram-variable attributes. They inhabit various sites including soil, sediment, sewage, caves, compost, water, plant and animal tissues, and so on. Some species of the genus *Paenibacillus* are reported to be plant growth-promoting rhizobacteria that fix nitrogen, solubilize phosphate and act as antagonistic agents for plant pathogens [1,2]. Members of the genus *Paenibacillus* produce a range of molecules such as amylases, cellulases, lipases, pectinases, oxygenases, dehydrogenases, lignin-modifying enzyme, exopolysaccharides and antibiotic agents, and these products have applications in the food, paper, biofuel, textiles, agriculture, detergents and medicine industries [1]. *Paenibacillus polymyxa* TH2H2 was previously isolated from a tomato flower [3]. The strain had antifungal activity, accompanied by high levels of cellulase and proteinase activity, towards the tomato pathogen *Botrytis cinera* [3].

The *P. polymyxa* TH2H2 genome was sequenced to 5,983,104 bp, and comprised 113 contigs and a G+C content of 45.31% (Fig. 1A, Table 1, supplementary data 1). There were 5,221 coding sequences (CDS), 64 ribosomal RNA and 100 transfer RNA. Cluster of Orthologous Groups (COG) analysis categorized 4,646 genes (supplementary data 1). With the exception of genes with unknown functions, the most frequently observed category was carbohydrate transport and metabolism, followed by transcription, then energy production and conversion (Fig. 1B). Thirty-



**Fig. 1.** Pseudogenome map and Clusters of Orthologous Groups (COG) analysis of *P. polymyxa* TH2H2 coding sequences. (A) In the draft genome map, the gray outermost ring is segmented by contig length; light gray and dark gray refer to odd and even numbers of contigs, respectively. The next two gray rings indicate CDS in each DNA direction. The remaining bar graph rings are GC ratio and GC skew per 10,000 bp. (B) EggNOG functional category classification. The x-axis represents gene functional category and the y-axis represents the frequency of each functional category.

**Table 1**  
Genome features of *Paenibacillus polymyxa* TH2H2.

Genome feature	Value
Genome size	5,983,104 bp
G+C content	45.31%
GC skew	-0.004 bp/bp
Number of rRNA genes	64
Number of tRNA genes	100
Number of ORFs	5,221
Number of contigs	113
Sequencing depth of coverage	139.97×
N50	466,974
L50	5

seven secondary metabolite gene clusters were predicted using antiSMASH (supplementary data 2). Among these, three complete non-ribosomal peptides (NRP) clusters were identified. The first NRP was fusaricidin biosynthetic gene cluster [4]; fusaricidin is a recognized antifungal agent. The other two NRPs were tridecaptin and polymyxin biosynthetic gene clusters [5,6], both of which are antibacterial agents targeting Gram-negative bacteria.

## Experimental Design, Materials and Methods

*P. polymyxa* TH2H2 was cultivated as previously described [3]. Genomic DNA was extracted using the CTAB method [7]. Illumina Miseq paired-end (2 × 300 bp) sequencing of *P. polymyxa* TH2H2 was performed by Chunlab Inc. (Seoul, Korea). Briefly, genomic DNA was processed into a library using NEBNext dsDNA fragmentase (NEB, Hitchin, UK) and TruSeq RNA Library Prep Kit v2 (Illumina, Inc., San Diego, CA, USA). The sequencing generated 4,895,864 reads of 1,102,733,884 bp when adapter sequences were removed. The reads were assembled into 113 contigs of 137.97 × coverage using SPAdes (version 3.10.1) [8]. The absence of sequence contamination was confirmed (supplementary Figure 1) using ContEst16 (<https://www.ezbiocloud.net/tools/contest16s>) [9]. COG analysis was conducted using the EggNOG database [10]. The above analysis procedures were automatically performed with CLgenomics (version 1.55). Secondary metabolite gene clusters were predicted using antiSMASH (version 4.2.0; <https://antismash.secondarymetabolites.org>) [11]. All data were graphed in R (version 3.6.3), and the R code used was shared on a GitHub repository ([https://github.com/gyeongjunCho/TH2H2\\_draft\\_genome](https://github.com/gyeongjunCho/TH2H2_draft_genome)).

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

## Acknowledgments

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (Grant number: 2020R1A2C2004177)

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.dib.2020.105824](https://doi.org/10.1016/j.dib.2020.105824).

## References

- [1] E.N. Grady, J. MacDonald, L. Liu, A. Richman, Z.-C. Yuan, Current knowledge and perspectives of *Paenibacillus*: a review, *Microb. Cell Fact.* 15 (2016) 203, doi:[10.1186/s12934-016-0603-7](https://doi.org/10.1186/s12934-016-0603-7).
- [2] B. Weselowski, et al., Isolation, identification and characterization of *Paenibacillus polymyxa* CR1 with potentials for biopesticide, biofertilization, biomass degradation and biofuel production, *BMC Microbiol* 16 (2016) 244, doi:[10.1186/s12866-016-0860-y](https://doi.org/10.1186/s12866-016-0860-y).
- [3] D.-R. Kim, et al., Selection of biocontrol agent of tomato gray mold disease from flower and pollinator hive, *Korean J. Pestic. Sci.* 21 (2017) 90–96, doi:[10.7585/kjps.2017.21.1.90](https://doi.org/10.7585/kjps.2017.21.1.90).
- [4] J. Li, S.E. Jensen, Nonribosomal biosynthesis of fusaricidins by *Paenibacillus polymyxa* PKB1 involves direct activation of a d-amino acid, *Chem. Biol.* 15 (2008) 118–127, doi:[10.1016/j.chembiol.2007.12.014](https://doi.org/10.1016/j.chembiol.2007.12.014).
- [5] C.T. Lohans, et al., Biochemical, structural, and genetic characterization of tridecaptin A1, an antagonist of *Campylobacter jejuni*, *ChemBiochem* 15 (2014) 243–249, doi:[10.1002/cbic.201300595](https://doi.org/10.1002/cbic.201300595).
- [6] S.-K. Choi, et al., Identification of a polymyxin synthetase gene cluster of *Paenibacillus polymyxa* and heterologous expression of the gene in *Bacillus subtilis*, *J. Bacteriol.* 191 (2009) 3350–3358 <https://doi.org/10.1128/JB.01728-08>.
- [7] S. William, H. Feil, A. Copeland, *Bacterial genomic DNA isolation using CTAB*, *Sigma* 50 (2012) 6876.
- [8] A. Bankevich, et al., SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing, *J. Comput. Biol.* 19 (2012) 455–477, doi:[10.1089/cmb.2012.0021](https://doi.org/10.1089/cmb.2012.0021).
- [9] I. Lee, et al., ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences, *Int. J. Syst. Evol. Microbiol* 67 (2017) 2053–2057, doi:[10.1099/ijsem.0.001872](https://doi.org/10.1099/ijsem.0.001872).
- [10] S. Powell, et al., eggNOG v4.0: nested orthology inference across 3686 organisms, *Nucleic Acids Res* 42 (2014) D231–D239, doi:[10.1093/nar/gkt1253](https://doi.org/10.1093/nar/gkt1253).
- [11] M.H. Medema, et al., antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences, *Nucleic Acids Res* 39 (2011) W339–W346, doi:[10.1093/nar/gkr466](https://doi.org/10.1093/nar/gkr466).