Data in Brief 54 (2024) 110285

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data on annotation and analysis of genome sequence of *Paenibacillus elgii* YSY-1.2, a promising chitinase-producing, plant-growth-promoting, and biocontrol agent

Dinh Minh Tran[a,*], Phuong Thi Pham[b], Bich Thuy Vu[b], Le Nguyen Tieu Ngoc[a]

[a] *Institute of Biotechnology and Environment, Tay Nguyen University, Buon Ma Thuot, Dak Lak 630000, Viet Nam*
[b] *Faculty of Natural Science and Technology, Tay Nguyen University, Buon Ma Thuot, Dak Lak 630000, Viet Nam*

ARTICLE INFO

Article history:
Received 7 January 2024
Revised 26 January 2024
Accepted 26 February 2024
Available online 2 March 2024

Dataset link: Whole genome sequence of Paenibacillus elgii YSY-1.2, a promising chitinase-producing, plant-growth-promoting, and biocontrol agent (Original data)
Dataset link: Paenibacillus elgii strain YSY-1.2, whole genome shotgun sequencing project (Original data)

Keywords:
Genome sequence
Paenibacillus elgii
Chitin-degrading system
CAZymes
Antimicrobial metabolites

ABSTRACT

The bacterium *Paenibacillus elgii* YSY-1.2 was recently isolated from soil collected from Yok Don National Park in Vietnam. Previous experiments showed this bacterium possesses high chitin-degrading activity, plant-growth promotion, and biocontrol capacity. Here, we report the draft genome sequence of strain YSY-1.2 for further characterizations related to crop production. The genome sequencing was performed using the DNBSeq-G99 with the Illumina platform. The draft genome of *P. elgii* YSY-1.2 has 8,240,519 bp in length and comprises 135 contigs. It has an N50 of 315,408 bp and a GC% of 52.8%. The genome contains 7498 protein-coding genes, 87 tRNA genes, and 1 rRNA gene. Among the protein-coding sequences, 6610 were assigned by COG, while 3230 were assigned by KEGG. The genome possesses at least 61 genes involved in environmental adaptation and plant growth promotion. Additionally; there are 258 carbohydrate-active enzymes deduced from the genome; among them, at least 14 may contribute to the biocontrol capacity. The chitin-degrading system of strain YSY-1.2 contains 16 chitinolytic enzymes, comprising 10 chitinases, 4 β-*N*-acetylhexosaminidases, and 2 auxiliary activities. Furthermore, 32 gene clusters encoding antimicrobial metabolites were identified from the genome, with 17 show-

* Corresponding author.
E-mail address: tmdinh@ttn.edu.vn (D.M. Tran).

https://doi.org/10.1016/j.dib.2024.110285

ing no sequence similarities to reported clusters. Data provide an insight into the genomic information of strain YSY-1.2 and could lead to valuable further explorations and applications in crop production. This is the first report describing the genome sequence of *P. elgii* isolated from Vietnam.

Specifications Table

Subject	Microbiology: Applied Microbiology
Specific subject area	Molecular biology, Bioinformatics
Data format	Raw, Filtered, and Analyzed
Type of data	Figures, Tables
Data collection	The genomic DNA was extracted using the QIAamp DNA mini kit (Qiagen, Germany). The genomic library was prepared using the NEBNext dsDNA Fragmentase, NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, USA), and sequenced using DNBSeq-G99 (MGI) with the Illumina platform (2×150 PE). The assembly was performed using Unicycler v0.4.8. The annotation was analyzed against DFAST, COG, and KEGG. CAZymes and biosynthesis gene clusters were identified by the dbCAN3 metaserver and antiSMASH v7.0, respectively.
Data source location	Institution: Institute of Biotechnology and Environment, Tay Nguyen University City/Province/Country: Buon Ma Thuot/Dak Lak/Vietnam
Data accessibility	1. Raw sequences Repository name: Mendeley Data Data identification number: doi: 10.17632/8z9jjcwmk8.1 Direct URL to data: https://data.mendeley.com/datasets/8z9jjcwmk8/1 2. Genome sequence Repository name: DDBJ/GenBank/EMBL Data identification number: BTYA00000000 Direct URL to data: https://www.ncbi.nlm.nih.gov/nuccore/BTYA00000000.1

1. Value of the Data

- Data provide an insight into the genomic information of the chitinolytic *P. elgii* YSY-1.2.
- Data elucidate capacities for chitin degradation, phytopathogenic biocontrol, and plant-growth promotion of *P. elgii* YSY-1.2.
- Data can be valuable for future explorations in crop production and related fields using gene cloning and expression.

2. Background

P. elgii YSY-1.2 was originally isolated from the soil sample collected from Yok Don National Park in Vietnam. Our previous results showed that this bacterium possessed high chitinase activity (3.09 U per mg proteins), high inhibiting activity (63.33%) against the growth of phytopathogenic fungi, produced extracellular enzymes (protease, cellulase, amylase), phytohormones (IAA, GA3, and Zeatin), and siderophores. Therefore, this bacterium has the potential for chitin degradation, phytopathogenic biocontrol, and plant growth promotion [1]. However, the genome sequence of strain YSY-1.2 has yet to be reported. This work aimed to sequence and analyze the draft genome of strain YSY-1.2 for further characterizations related to crop production using gene expression.

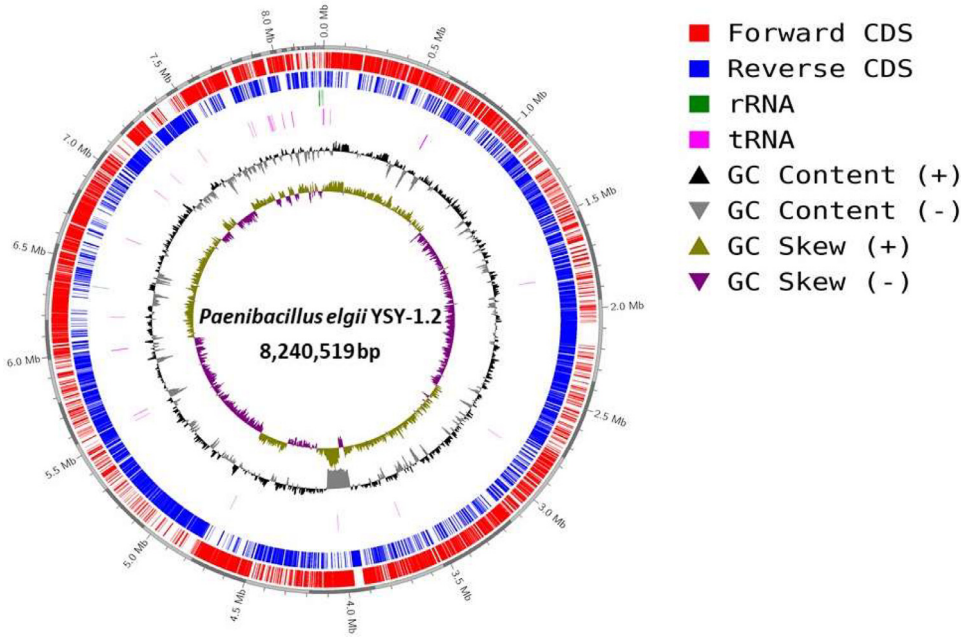

Fig. 1. Circular representation of the draft genome of *Paenibacillus elgii* YSY-1.2.

Table 1
General genomic features of *Paenibacillus elgii* YSY-1.2.

Feature	Value
Genome size (bp)	8,240,519
G+C content (%)	52.8
No. of contigs	135
The largest contig (bp)	620,115
The shortest contig (bp)	174
N50	315,408
Coding sequences	7498
Gene assigned to COG	6110
Gene assigned to KEGG	3230
rRNA	1
tRNA	87
CRISPR arrays	4

3. Data Description

Assembly showed that the genome (Fig. 1) of strain YSY-1.2 has 8,240,519 bp in size and a 52.8% G+C content. It was generated from 135 contigs, with the maximum contig size of 620,115 bp and the minimum contig size of 174 bp. The genome annotation using DFAST predicted 7498 protein-coding genes, 1 rRNA gene, 87 tRNA genes, and 4 CRISPR arrays. The genome of strain YSY-1.2 shares a 95.65% ANI value with that of *P. elgii* AC13 (PYHP00000000.1), indicating that it is a new member of the species *P. elgii*. The COG annotation showed that 6610 gene annotations were obtained, accounting for 81.49% of the total genes. In contrast, the KEGG annotation revealed that 3230 genes were assigned, accounting for 43.1% of the total genes (Table 1 and Fig. 2). The raw read sequence was submitted to Mendeley Data and is available at https://data.mendeley.com/datasets/8z9jjcwmk8/1. The genome sequence of *P. elgii* YSY-1.2 is

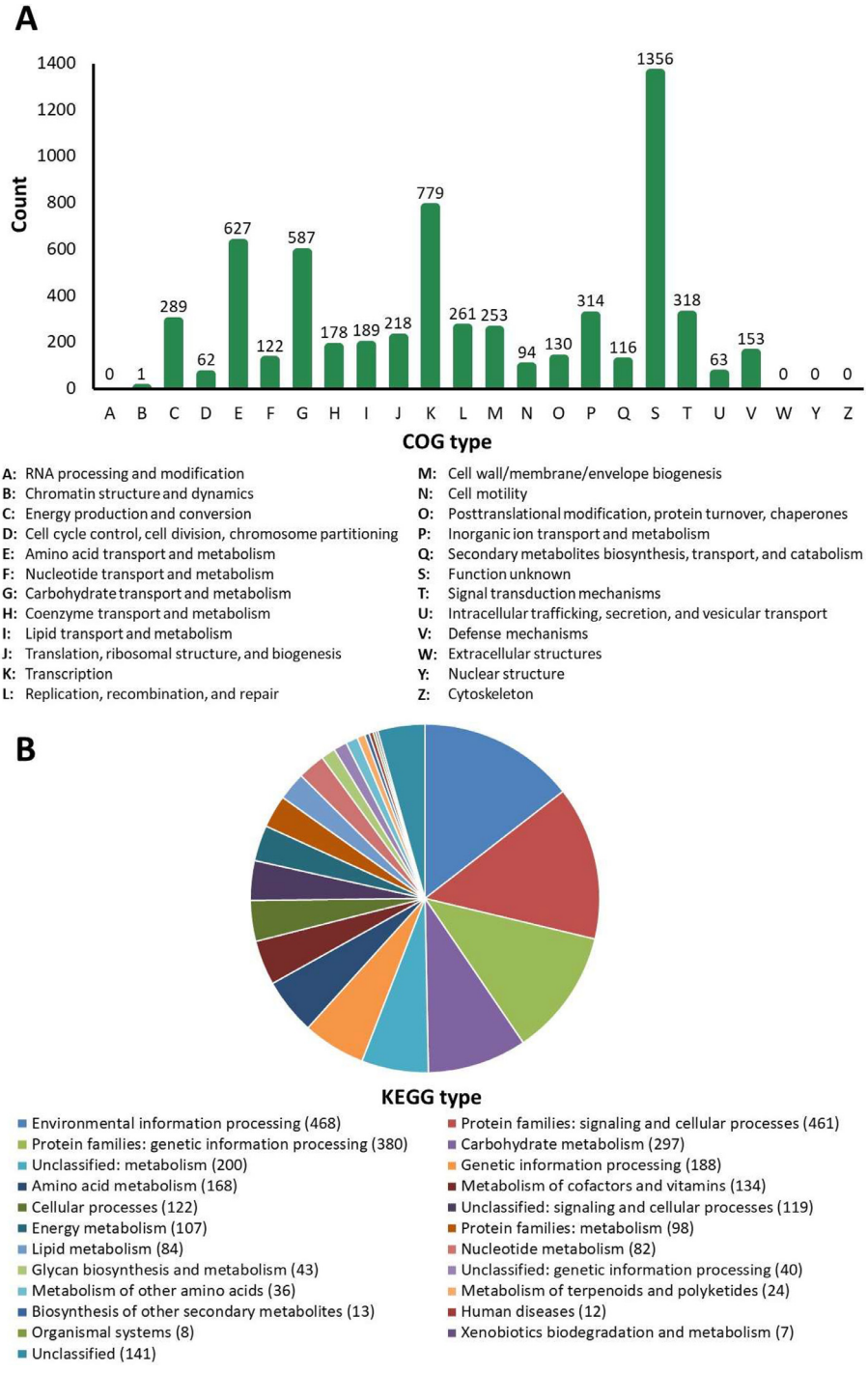

Fig. 2. The COG and KEGG annotation of coding sequences of *Paenibacillus elgii* YSY-1.2. A, the COG annotation of coding sequences; B, the KEGG annotation of the sequences.

Table 2
Putative genes involved in environmental adaptation and plant growth promotion of *Paenibacillus elgii* YSY-1.2.

Role	Product	Locus tag
Heavy metal resistance		
Cd^{2+}, Co^{2+}, Zn^{2+}	Cation diffusion facilitator family transporter	Elgi_28150
As^{3+}, As^{5+}	Arsenic transporter	Elgi_46280, Elgi_56740
Cr^{6+}	Chromate transporter	Elgi_18630, Elgi_18640, Elgi_48290
Phosphate solubilization	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	Elgi_01780
Zinc solubilization	Zinc-dependent sulfurtransferase SufU	Elgi_05320
	Zinc-transporting ATPase	Elgi_22940
	Zinc uptake transcriptional repressor Zur	Elgi_63460
Potassium solubilization	TrkA family potassium uptake protein	Elgi_12430, Elgi_72370
	TrkH family potassium uptake protein	Elgi_12440, Elgi_15740
Indole-3-acetic acid production	Tryptophan synthase subunit alpha	Elgi_45180
	Threonine synthase	Elgi_18460
	Anthranilate phosphoribosyltransferase	Elgi_45140, Elgi_57810
	Anthranilate synthase component 1	Elgi_45130
	Aminodeoxychorismate/anthranilate synthase component 2	Elgi_68040
	Indole-3-glycerol phosphate synthase TrpC	Elgi_45150
	Chorismate mutase	Elgi_45120, Elgi_45100
	3-dehydroquinate synthase	Elgi_45110
Nitrate transport and reduction	Nitrite reductase small subunit NirD	Elgi_17080, Elgi_13790
	NADPH-nitrite reductase	Elgi_13780, Elgi_17090
	Nitrate/nitrite transporter	Elgi_17100
	Nitrate transporter NarK	Elgi_46520
ACC biosynthesis	Acyl-CoA dehydrogenase FadE	Elgi_23140
	Acyl-CoA dehydrogenase	Elgi_47060, Elgi_54500, Elgi_67150
	Acyl-CoA dehydrogenase family protein	Elgi_51250
	Acyl-CoA dehydrogenase AcdA	Elgi_54520
Iron uptake	Iron ABC transporter permease	Elgi_01060, Elgi_13560, Elgi_13570, Elgi_14480, Elgi_14540, Elgi_15350, Elgi_16730, Elgi_25990, Elgi_26000, Elgi_27700, Elgi_29320, Elgi_29330, Elgi_33950, Elgi_46590, Elgi_47580, Elgi_51160, Elgi_54360, Elgi_73750
Siderophore production	Siderophore ABC transporter substrate-binding protein	Elgi_12010, Elgi_16420, Elgi_26010, Elgi_29340, Elgi_53630, Elgi_63050, Elgi_74300

available at the DDBJ/GenBank/EMBL database under accession number BTYA00000000 and can be retrieved at https://www.ncbi.nlm.nih.gov/nuccore/BTYA00000000.1.

Table 2 shows that the genome of strain YSY-1.2 contains 61 genes involved in environmental adaptation and plant growth promotion, including heavy metal resistance (6 genes), phosphate solubilization (1 gene), zinc solubilization (3 genes), potassium solubilization (4 genes), indole-3-acetic acid biosynthesis (10 genes), nitrate transport and reduction (6 genes), ACC biosynthesis (6 genes), iron uptake (18 genes), and siderophore production (7 genes). It was reported that

Table 3
Carbohydrate-active enzymes of *Paenibacillus elgii* YSY-1.2.

Class	Family (No.)
Glycoside hydrolases	GH2 (4), GH3 (3), GH4 (4), GH5 (2), GH8 (2), GH9 (1), GH10 (1), GH13 (6), GH15 (4), GH16 (3), GH18 (9), GH19 (1), GH20 (2), GH23 (1), GH26 (3), GH29 (4), GH30 (1), GH31 (2), GH32 (3), GH35 (1), GH38 (2), GH39 (1), GH43 (1), GH46 (1), GH57 (1), GH63 (1), GH65 (1), GH68 (2), GH73 (1), GH85 (2), GH87 (1), GH88 (4), GH93 (1), GH94 (1), GH95 (3), GH105 (1), GH109 (1), GH115 (1), GH117 (1), GH123 (1), GH125 (1), GH129 (1), GH136 (2), GH151 (1), GH154 (1), GH171 (1), GH176 (1), GH177 (9), GH179 (20)
Glycosyltransferase	GT1 (6), GT2 (22), GT4 (16), GT5 (1), GT17 (1), GT26 (1), GT28 (5), GT35 (1), GT51 (4), GT101 (1)
Polysaccharide lyases	PL9 (2), PL12 (1), PL15 (1), PL17 (2), PL31 (2), PL35 (2), PL38 (1)
Carbohydrate esterases	CE1 (9), CE4 (24), CE9 (5), CE14 (5), CE17 (1)
Auxiliary activities	AA1 (2), AA3 (1), AA4 (1), AA6 (1), AA7 (1), AA10 (2)
Carbohydrate-binding modules	CBM9 (1), CBM12 (1), CBM13 (3), CBM32 (4), CBM35 (1), CBM54 (2), CBM96 (3)

indole-3-acetic acid is a phytohormone that contributes to plant growth and development and the remediation of contaminated soil [2]. Genes involved in heavy metal resistance, phosphate solubilization, zinc solubilization, potassium solubilization, nitrate transport and reduction, ACC biosynthesis, iron uptake, and siderophore production can contribute to the adaptation of bacteria to its environment and its ability to promote plant growth [3]. Therefore, our findings highlighted the potential of strain YSY-1.2 as a plant-growth-promoting agent. Further characterizations of these genes and strain YSY-1.2 concerning crop production are very necessary.

Carbohydrate-active enzymes (CAZymes) are crucial in the synthesis, degradation, and modification of carbohydrates and glycoconjugates [4]. In this work, 258 CAZymes were identified from the genome of strain YSY-1.2, including 122 glycoside hydrolases (GH), 58 glycosyltransferases (GT), 11 polysaccharide lyases (PL), 44 carbohydrate esterases (CE), and 8 auxiliary activities (AA), and 15 carbohydrate-binding modules (Table 3). Among these enzymes, 14 (3 GH16 glucanases, 9 GH18 chitinases, one GH19 chitinase, and one GH46 chitosanase) may possess activities against phytopathogens, such as fungi and nematodes.

Chitin-degrading enzymes play a crucial role in the hydrolysis of chitin polymers into small oligosaccharides [5]. As shown in Fig. 3, the genome of strain YSY-1.2 harbours 16 genes coding for chitinolytic enzymes, including 9 genes related to family 18 chitinases, one gene to family 19 chitinase, 2 genes to family 3 β-*N*-acetylhexosaminidases, 2 genes to family 20 β-*N*-acetylhexosaminidases, and 2 genes to auxiliary activity family 10 proteins. Of those, 6 possess multiple functional domains other than the catalytic domain. For instance, each enzyme (PeChiA, PeLPMO10A, PeLPMO10B, PeChiF, and PeChiH) contains two fibronectin type III domain (FN3) and a carbohydrate-binding module family 12 (CBM12), and PeChiM contains a fibronectin type III domain (FN3) and a carbohydrate-binding module family 5 (CBM5) in their primary structure. It is clear that chitin is a major component of the cell walls of fungi and the exoskeletons of arthropods [5]; therefore, further experiments are very necessary to express and characterize the main chitinases of strain YSY-1.2 toward controlling phytopathogens.

Antimicrobial metabolites produced by microorganisms are important for controlling phytopathogens [6]. In this report, 32 gene clusters responsible for antimicrobial metabolite biosynthesis were identified from the *P. elgii* YSY-1.2 genome. Among them, 17 showed no identities to the known clusters (Table 4). This result indicated that *P. elgii* YSY-1.2 has the potential to produce novel antimicrobial metabolites.

4. Experimental Design, Materials and Methods

A single colony of strain YSY-1.2 was inoculated into 5 mL of Luria–Bertani medium and incubated at 30 °C overnight (16 h), with shaking (150 rpm). Bacterial cells were then collected

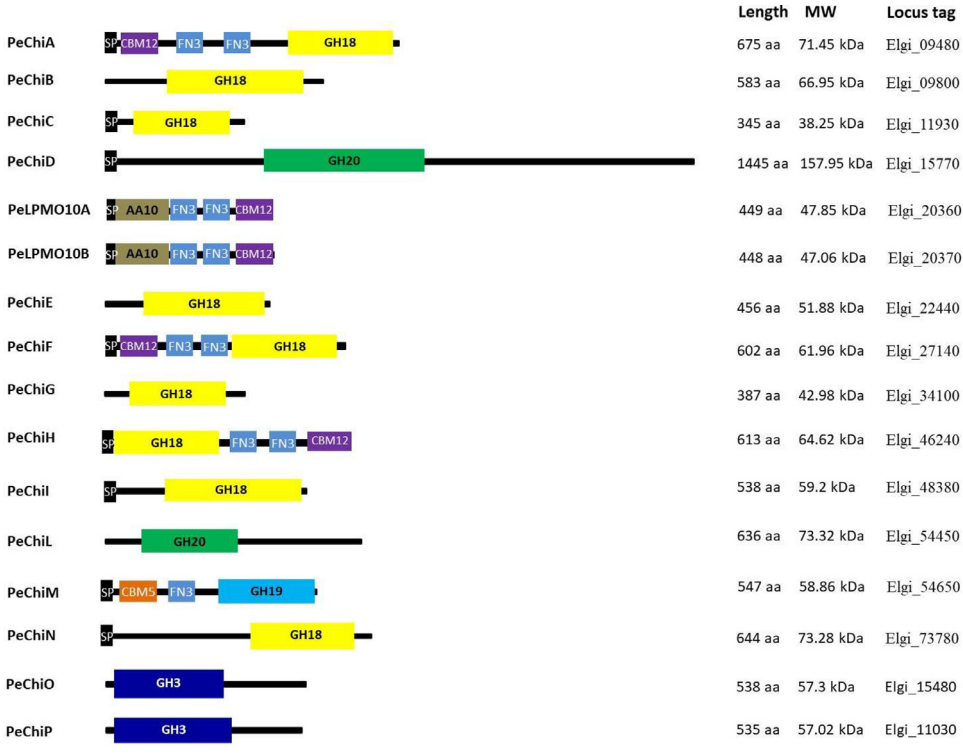

Fig. 3. Chitin-degrading system of *Paenibacillus elgii* YSY-1.2. SP, signal peptide sequence; GH18, glycoside hydrolase family 18; GH20, glycoside hydrolase family 20; AA10, auxiliary activity family 10; GH19, glycoside hydrolase family 19; GH3, glycoside hydrolase family 3. CBM12, carbohydrate-binding module family 12; FN3, fibronectin type III domain; CBM5, carbohydrate-binding module family 5. MW, molecular weight; aa, amino acid; kDa, kilodaltons.

using centrifugation (13,000 rpm, 4 °C, 10 min). The genomic DNA of strain YSY-1.2 was isolated using the QIAamp DNA mini kit (Qiagen, Germany) per the manufacturer's instructions. The whole genome library was prepared using the NEBNext dsDNA Fragmentase, NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, USA) under the manufacturer's instructions. Finally, prepared libraries were sequenced using DNBSeq-G99 (MGI) with the Illumina platform (2×150 PE) [1,7]. The raw reads were quality filtered by Fastp v0.23.1 [8]. The filtered reads were assembled by Unicycler v0.4.8 [9] to generate the draft genome sequence. Genomic sequence annotation was performed using DFAST, the web-based annotation pipeline [10]. The average nucleotide identity (ANI) value between the genome of strain YSY-1.2 and that of *P. elgii* AC13 (PYHP00000000.1) was calculated as described previously [11]. Functional annotation was examined using the eggNOG-mapper v2.0 [12] and the KEGG database [13]. Domain structures of deduced proteins were analyzed by SMART v9.0 [14] and Pfam v35.0 [15], respectively. CAZymes were analyzed by the dbCAN3 metaserver [16]. Phylogeny of deduced chitinases was computed by MEGA v6.0 [17]. Biosynthesis gene clusters were identified by antiSMASH v7.0 [18].

Table 4
Putative secondary metabolite gene clusters of *Paenibacillus elgii* YSY-1.2.

Cluster	Size (bp)	Type	Most similar known cluster	Similarity (%)
Cluster 1	97,604	NRPS/Lanthipeptide-class-i	Penisin	100
Cluster 2	46,957	NRPS	–	–
Cluster 3	73,623	Thioamide-NRP/HR-T2PKS	–	–
Cluster 4	20,522	Cyclic-lactone-autoinducer	–	–
Cluster 5	22,147	Opine-like-metallophore	Bacillopaline	100
Cluster 6	24,952	NRPS	Octapeptin C4	64
Cluster 7	20,558	Cyclic-lactone-autoinducer	–	–
Cluster 8	21,379	Cyclic-lactone-autoinducer	–	–
Cluster 9	21,911	Terpene	–	–
Cluster 10	22,071	Redox-cofactor	Lankacidin C	13
Cluster 11	83,200	NRPS/T1PKS	Puwainaphycin A/ puwainaphycin B/ puwainaphycin C/ puwainaphycin D	20
Cluster 12	101,393	NRPS/transAT-PKS	Paenilipoheptin	7
Cluster 13	21,618	NRPS	Fusaricidin B	87
Cluster 14	48,586	NRPS	–	–
Cluster 15	58,038	NRPS/T1PKS	Cystothiazole A	11
Cluster 16	24,751	Phosphonate	–	–
Cluster 17	91,131	TransAT-PKS	Macrobrevin	20
Cluster 18	32,593	LAP	–	–
Cluster 19	22,274	NRPS	–	–
Cluster 20	97,761	NRPS/NRP-metallophore	Bacillibactin	100
Cluster 21	40,607	NRPS	Laterocidine	5
Cluster 22	23,070	NRPS-like	–	–
Cluster 23	23,025	NRPS	Octapeptin C4	47
Cluster 24	10,640	NRPS	Paenilarvin A/paenilarvin B/paenilarvin C	75
Cluster 25	4735	NRPS	–	–
Cluster 26	4572	NRPS	–	–
Cluster 27	3265	NRPS	–	–
Cluster 28	3137	NRPS	–	–
Cluster 29	2536	NRPS	–	–
Cluster 30	2314	NRPS	–	–
Cluster 31	2074	NRPS	Thermoactinoamide A	100
Cluster 32	2071	NRPS	Thermoactinoamide A	100

Limitations

Not applicable.

Ethics Statement

The current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Data Availability

Whole genome sequence of Paenibacillus elgii YSY-1.2, a promising chitinase-producing, plant-growth-promoting, and biocontrol agent (Original data) (Mendeley Data)

Paenibacillus elgii strain YSY-1.2, whole genome shotgun sequencing project (Original data) (NCBI)

CRediT Author Statement

Dinh Minh Tran: Conceptualization, Methodology, Investigation, Formal analysis, Software, Data curation, Validation, Visualization, Writing – original draft, Writing – review & editing; **Phuong Thi Pham:** Investigation, Formal analysis; **Bich Thuy Vu:** Investigation, Formal analysis; **Le Nguyen Tieu Ngoc:** Investigation, Formal analysis, Software.

Acknowledgments

This research was funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant No. 106.04-2019.337.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] D.M. Tran, T.U. Huynh, T.H. Nguyen, T.O. Do, Q.V. Nguyen, A.D. Nguyen, Novel resources of chitinolytic bacteria isolated from Yok Don National Park, Vietnam, J. Appl. Microbiol. 134 (2023) lxad141, doi:10.1093/jambio/lxad141.
[2] M. Li, R. Guo, F. Yu, X. Chen, H. Zhao, H. Li, J. Wu, Indole-3-acetic acid biosynthesis pathways in the plant-beneficial bacterium *Arthrobacter pascens* ZZ21, Int. J. Mol. Sci. 19 (2) (2018) 443, doi:10.3390/ijms19020443.
[3] D. Chlebek, T. Płociniczak, S. Gobetti, A. Kumor, K. Hupert-Kocurek, M. Pacwa-Płociniczak, Analysis of the genome of the heavy metal resistant and hydrocarbon-degrading rhizospheric *Pseudomonas qingdaonensis* ZCR6 strain and assessment of its plant-growth-promoting traits, Int. J. Mol. Sci. 23 (1) (2021) 214, doi:10.3390/ijms23010214.
[4] B. Cobucci-Ponzano, M. Rossi, M. Moracci, Carbohydrate-active enzymes from hyperthermophiles: biochemistry and applications, in: Extremophiles Handbook, Springer, Tokyo, 2011, pp. 427–441, doi:10.1007/978-4-431-53898-1_20.
[5] V. Poria, A. Rana, A. Kumari, J. Grewal, K. Pranaw, S. Singh, Current perspectives on chitinolytic enzymes and their agro-industrial applications, Biology 10 (12) (2021) 1319 (Basel), doi:10.3390/biology10121319.
[6] Z. Narayanan, B.R. Glick, Secondary metabolites produced by plant growth-promoting bacterial endophytes, Microorganisms 10 (10) (2022) 2008, doi:10.3390/microorganisms10102008.
[7] D.M. Tran, T.H. Nguyen, A.D. Nguyen, Complete genome sequence data of chitin-degrading *Bacillus velezensis* RB.IBE29, Data Br. 51 (2023) 109815, doi:10.1016/j.dib.2023.109815.
[8] S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (17) (2018) i884–i890, doi:10.1093/bioinformatics/bty560.
[9] R.R. Wick, L.M. Judd, C.L. Gorrie, K.E. Holt, Unicycler: resolving bacterial genome assemblies from short and long sequencing reads, PLoS Comput. Biol. 13 (2017) e1005595, doi:10.1371/journal.pcbi.1005595.
[10] Y. Tanizawa, T. Fujisawa, Y. Nakamura, DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication, Bioinformatics 34 (2018) 1037–1039, doi:10.1093/bioinformatics/btx713.
[11] S.H. Yoon, S.M. Ha, J. Lim, S. Kwon, J. Chun, A large-scale evaluation of algorithms to calculate average nucleotide identity, Antonie Leeuwenhoek 110 (10) (2017) 1281–1286, doi:10.1007/s10482-017-0844-4.
[12] C.P. Cantalapiedra, A. Hernández-Plaza, I. Letunic, P. Bork, J. Huerta-Cepas, eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale, Mol. Biol. Evol. 38 (12) (2021) 5825–5829, doi:10.1093/molbev/msab293.
[13] M. Kanehisa, M. Furumichi, Y. Sato, M. Kawashima, M. Ishiguro-Watanabe, KEGG for taxonomy-based analysis of pathways and genomes, Nucleic Acids Res. 51 ((D1) (2023) D587–D592, doi:10.1093/nar/gkac963.
[14] I. Letunic, S. Khedkar, P. Bork, SMART: recent updates, new developments and status in 2020, Nucleic Acids Res. 49 (D1) (2021) D458–D460, doi:10.1093/nar/gkaa937.
[15] J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G.A. Salazar, E.L.L. Sonnhammer, S.C.E. Tosatto, L. Paladin, S. Raj, L.J. Richardson, R.D. Finn, A. Bateman, Pfam: the protein families database in 2021, Nucleic Acids Res. 49 (D1) (2021) D412–D419, doi:10.1093/nar/gkaa913.
[16] J. Zheng, Q. Ge, Y. Yan, X. Zhang, L. Huang, Y. Yin, dbCAN3: automated carbohydrate-active enzyme and substrate annotation, Nucleic Acids Res. 51 (W1) (2023) W115–W121, doi:10.1093/nar/gkad328.
[17] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0, Mol. Biol. Evol. 30 (12) (2013) 2725–2729, doi:10.1093/molbev/mst197.
[18] K. Blin, S. Shaw, H.E. Augustijn, Z.L. Reitz, F. Biermann, M. Alanjary, A. Fetter, B.R. Terlouw, W.W. Metcalf, E.J.N. Helfrich, G.P. van Wezel, M.H. Medema, T. Weber, antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation, Nucleic Acids Res. 51 (W1) (2023) W46–W50, doi:10.1093/nar/gkad344.