



## Comparative Genomic Analysis of Pathogenic Factors of *Pectobacterium* Species Isolated in South Korea Using Whole-Genome Sequencing

Samnyu Jee <sup>1,2\*</sup>, In-Jeong Kang<sup>2,3</sup>, Gyeryeong Bak<sup>1</sup>, Sera Kang<sup>1</sup>, Jeongtae Lee<sup>1</sup>, Sunggi Heu<sup>4</sup>, and Ingyu Hwang<sup>2</sup>

<sup>1</sup>Highland Agriculture Research Institute, National Institute of Crop Science, Rural Development Administration, Pyeongchang 25342, Korea

<sup>2</sup>Department of Agricultural Biotechnology, Seoul National University, Seoul 08826, Korea

<sup>3</sup>Division of Crop Cultivation and Environment Research, National Institute of Crop Science, Suwon 16613, Korea

<sup>4</sup>Department of Plant Science, College of Agriculture and Life Science, Seoul National University, Seoul 08826, Korea

(Received on September 30, 2021; Revised on December 15, 2021; Accepted on December 21, 2021)

In this study, we conducted whole-genome sequencing with six species of *Pectobacterium* composed of seven strains, JR1.1, BP201601.1, JK2.1, HNP201719, MYP201603, PZ1, and HC, for the analysis of pathogenic factors associated with the genome of *Pectobacterium*. The genome sizes ranged from 4,724,337 bp to 5,208,618 bp, with the GC content ranging from 50.4% to 52.3%. The average nucleotide identity was 98% among the two *Pectobacterium* species and ranged from 88% to 96% among the remaining six species. A similar distribution was observed in the carbohydrate-active enzymes (CAZymes) class and extracellular plant cell wall degrading enzymes (PCWDEs). HC showed the highest number of enzymes in CAZymes and the lowest number in the extracellular PCWDEs. Six strains showed four subsets, and HC demonstrated three subsets, except *hasDEF*, in type I secretion system, while the type II secretion system of the seven strains was conserved. Components of human pathogens, such as

*Salmonella* pathogenicity island 1 type III secretion system (T3SS) and effectors, were identified in PZ1; T3SSa was not identified in HC. Two putative effectors, including *hrpK*, were identified in seven strains along with *dspEF*. We also identified 13 structural genes, six regulator genes, and five accessory genes in the type VI secretion system (T6SS) gene cluster of six *Pectobacterium* species, along with the loss of T6SS in PZ1. HC had two subsets, and JK2.1 had three subsets of T6SS. With the GxSxG motif, the phospholipase A gene did locate among *tssID* and *duf4123* genes in the T6SSa cluster of all strains. Important domains were identified in the *vgrG/paar* islands, including *duf4123*, *duf2235*, *vrr-nuc*, and *duf3396*.

**Keywords** : genomic analysis, *Pectobacterium* species, putative phospholipase effector, type secretion system

\*Corresponding author.

Phone) +82-33-330-1920, FAX) +82-33-330-1590

E-mail) jsnimp@korea.kr

ORCID

Samnyu Jee

https://orcid.org/0000-0001-7834-4303

Handling Editor : Kihyuck Choi

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Articles can be freely viewed online at [www.ppjonline.org](http://www.ppjonline.org).

The classification of the genus *Pectobacterium*, a causal agent of soft rot, has recently been revised due to its phenotypic and genotypic heterogeneity. This genus has been reclassified into 20 species: *P. aroidearum* (Par), *P. aquaticum*, *P. atrosepticum* (Pat), *P. actinidiae* (Pac), *P. betavasculorum*, *P. brasiliense* (Pbr), *P. cacticida*, *P. carotovorum* (Pca), *P. fontis*, *P. odoriferum* (Pod), *P. parmentieri* (Ppa), *P. parvum* (Ppv), *P. peruvienne*, *P. polaris* (Ppo), *P. polonicum*, *P. punjabense*, *P. quasiquaticum*, *P. versatile* (Pve), *P. wasabiae*, and *P. zantedeschiae* (Jee et al., 2020; Moussa et al., 2021; Pasanen et al., 2020; Portier et al., 2020). Pasanen et al. (2020) suggested a new species, Ppv, differentiated from Ppo with low average nucleotide

identity (ANI) coverage and reported a type III secretion system (T3SS) of *Salmonella* pathogenicity island 1 (SPI-1). Among the 20 species, 8, (Pat, Pac, Pbr, Pca, Pod, Ppa, Ppo, and Pve) were isolated from various hosts in South Korea (Jee et al., 2020; Lee et al., 2014, 2017; Park et al., 1999). Due to the heterogeneous genome of *Pectobacterium* strains, new species may be observed upon further examination, as confirmed previously (Portier et al., 2020). We previously identified that Pca, Pod, Pbr, and Pve were the prevalent species isolated from potatoes in South Korea (Jee et al., 2020). This result was specific to South Korea because Pod is known as a soft rot pathogen that primarily causes the disease in vegetables.

Attempts have been made to understand pathogenic characteristics via whole-genome sequencing and comparative analysis of *Pectobacterium* (Arizala and Arif, 2019; Li et al., 2018, 2019). Although some species were sequenced incompletely, such as *P. fontis* and *P. betavascularum*, at least one strain of the prevalent species has been completely sequenced at the whole-genome level. *Pectobacterium* demonstrates pathogenic factors, such as plant cell wall degrading enzymes (PCWDEs), secretion systems, and quorum sensing systems, which aid in the infection of host plants. PCWDEs include cellulases, proteases, and pectin-degrading enzymes, including pectin lyases, pectin esterases, and polygalacturonases. *Pectobacterium* is known to effectively degrade plant cell walls using these enzymes. Type I secretion system (TISS) and type II secretion system (T2SS) are used by pathogens for the extracellular secretion of enzymes. TISS is known for one-step transport, while T2SS, located in the outer membrane, requires Sec or Tat secretion pathway to deliver the target protein to the periplasm (Green and Mecsas, 2016). First identified in 1994, T3SS could transfer toxic proteins acting as effectors directly into the host cell. T3SS is considered an important factor for pathogenicity as it aids gram-negative bacterial pathogens in infecting a variety of hosts, ranging from plants to humans. Type VI secretion system (T6SS) was first reported in 2006, and its apparatus resembles a phage structure (Bernal et al., 2018). T6SS targets eukaryotic and prokaryotic cells for host manipulation and interbacterial competition.

Jee et al. (2020) evaluated the phenotype of pathogenic factors according to temperature with *Pectobacterium* strains isolated in South Korea. To identify these pathogenic factors in the genome, seven strains from South Korean collection isolated by Jee et al. (2020) were selected as representatives for each species. We conducted complete whole-genome sequencing and comparative analysis with seven representatives of South Korean *Pectobacterium*.

The strains isolated in South Korea for full genome sequencing are Pbr strain PCC21 and three Pac strains: KKH3, ICMP 19971, and ICMP 19972 (Lee et al., 2017; Park et al., 2012; Visnovsky et al., 2017). The objective of this study was to update the genome sequence information of *Pectobacterium* species that occur in South Korea and understand the pathogenicity factors associated with their genomes.

---

## Materials and Methods

**Bacterial strains and genomic DNA extraction.** Seven bacterial strains were selected as the representative strains of the *Pectobacterium* species isolated in South Korea. The seven bacterial strains consisted of six species: Pca, Pod, Pbr, Pve, Ppv, and Ppa. Jee et al. (2020) isolated five strains from potato (*Solanum tuberosum*), *P. carotovorum* JR1.1 from radish (*Raphanus sativus*), and *P. odoriferum* JK2.1 from kimchi cabbage (*Brassica rapa* subsp. *pekinensis*). The seven strains were cultivated in Luria-Bertani medium (BD, Franklin Lakes, NJ, USA) at 28°C for 16 h. Genomic DNA of seven *Pectobacterium* strains was extracted using the G-spin Genomic DNA Extraction Kit (Intron Biotechnology, Seongnam, Korea) according to the manufacturer's instructions.

**Whole-genome sequencing and annotation.** The extracted DNA was examined via electrophoresis using 1.0% agarose to assess DNA quality. The complete genomes of seven *Pectobacterium* strains were sequenced at Phyzen Co. Ltd. (Seongnam, Korea) using the PacBio RS II platform with a Template Prep Kit from the SMART bell library. The constructed library was validated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). *De novo* assembly was performed using PBcR (ver. 8.3.rc2), SMARTdenovo (ver. 1.0.0), and Flye (ver. 2.7-b1587). The completeness of each genome was measured with Benchmarking Universal Single Copy Orthologs (BUSCO) score using bacteria\_odb9 library. Gene prediction was performed using the NCBI Prokaryotic Genome Annotation Pipeline (Haft et al., 2018; Li et al., 2021; Tatusova et al., 2016). GenBank accession numbers were assigned to each strain in this study (Table 1).

**Comparative genome analysis.** All genome sequences of *Pectobacterium* were obtained from the NCBI database. To compare homology patterns among the seven *Pectobacterium* strains, genome sequences were aligned with the progressive Mauve algorithm using Mauve (ver. 2.4.0, <http://darlinglab.org/mauve/>) (Darling et al., 2004).

ANIs were evaluated using the ANI/AAI-Matrix (<http://enve-omics.ce.gatech.edu/g-matrix/>), and *in silico* DNA-DNA hybridization (*isDDH*) values were calculated using the default settings of the Genome-to-Genome Distance Calculator (GGDC 3.0, <http://ggdc.dsmz.de/ggdc.php>) (Meier-Kolthoff et al., 2021). A pan-genome phylogenetic tree was constructed with the bacterial pan-genome analysis pipeline software (BPGA, <https://iicb.res.in/bpga/>) using recommended settings and MEGA7 program using neighbor-joining method with 1,000 bootstraps (Chaudhari et al., 2016). Thirteen reference species were retrieved from the GenBank database (Supplementary Table 1). For functional genomics analysis among the strains, the coding sequences (CDSs) were classified into 25 functional groups with reference to orthologous groups (EggNOG 4.5, <http://eggdb.embl.de>) (Powell et al., 2014).

### CAZyme and pathogenesis-related gene analysis.

Carbohydrate-active enzymes (CAZymes) of the seven *Pectobacterium* strains were analyzed using the dbCAN meta server (<http://ccb.unl.edu/dbCAN2/>) (Zhang et al., 2018). Pathogenesis-related genes, such as those encoding PCWDEs and type secretion systems, including T1SS, T2SS, T3SS, and T6SS from one to six in each genome, were identified and compared using BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), such as BLASTN, BLASTX, and BLASTP, and UniProt (<https://www.uniprot.org/>) with previously known genes of *Pectobacterium* (Li et al., 2018, 2019). Signal peptides and subcellular locations of proteins were identified using UniProt. For accuracy, each BLAST hit of an open reading frame was limited to over 40% query coverage, the lowest e-value, and the highest percent identity. This study identified three types of enzymes: cellulases, proteases including metalloproteases and serine proteases, and pectin-degrading enzymes including pectate

lyases, pectin lyases, pectin esterases, rhamnogalacturonate lyase, and polygalacturonases. The motif of the putative effector of T6SS was identified using MEME (ver. 5.3.3, <https://meme-suite.org>) (Bailey et al., 2015).

## Results and Discussion

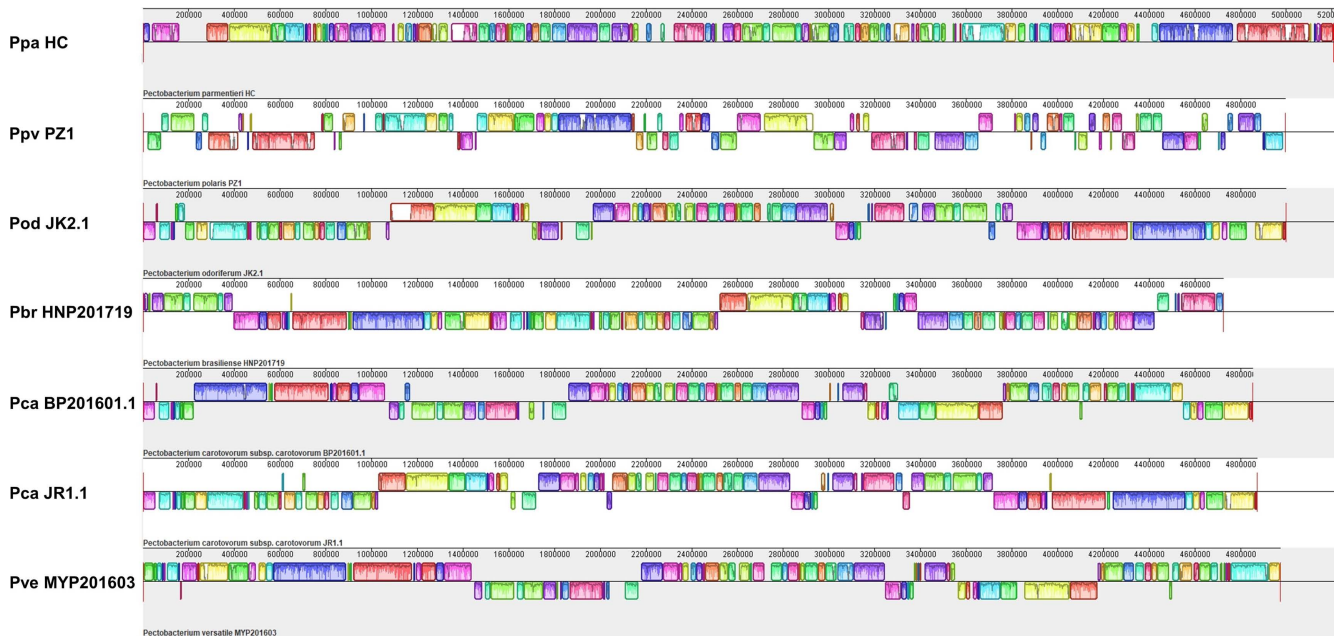
### Comparison of the whole-genome sequences of South Korean *Pectobacterium* strains.

Whole-genome sequence analysis was conducted with seven *Pectobacterium* strains considered representatives of South Korean *Pectobacterium* (Table 1). Since JR1.1 was isolated from radish, a rare host, we selected two strains of Pca for comparison with the strain isolated from potato. To obtain a single chromosome, *de novo* assembly was performed with three programs for each strain. The best results were selected considering the low contig number, high BUSCO score, and longest sequence length with an average coverage of 300×. The BUSCO score of the seven genomes ranged from 96.0% of MYP201603 to 98.0% of JR1.1. The genome sizes of the strains ranged from 4,724,337 bp observed with HNP201719 to 5,208,618 bp observed with HC, with a GC content ranging from 50.4% in HC to 52.3% in HNP201719 (Table 1, Fig. 1). The average number of CDSs of the seven strains varied, including 4,563 in JR1.1, BP201601.1, and HNP201719; 4,815 in JK2.1; and 5,161 in MYP201603, PZ1, and HC. The number of repeat regions ranged from 135 in HNP201719 to 226 in PZ1. A plasmid was identified in JK2.1 and MYP201603. Two plasmids showed a size of approximately 10 Mb and 120 CDSs. Li et al. (2018) reported that Ppa had the lowest average GC content, and Pbr had the highest average GC content among the six species. These findings are similar to the results of this study, but the GC content (52.3%) observed in HNP201719 was slightly higher than the 52.18%

**Table 1.** Genomic data of seven *Pectobacterium* strains identified in South Korean collection

Species	Strain	Type of sequence	No. contig	Dep. of coverage	Size (bp)	GC (%)	CDSs	rRNA	tRNA	Repeat regions	Accession no.
<i>P. parmentieri</i>	HC	Chromosome	1	295	5,208,618	50.4	5,164	22	77	163	CP046376.1
<i>P. parvum</i>	PZ1	Chromosome	1	223	4,994,870	51.0	5,132	22	76	226	CP046377.1
<i>P. odoriferum</i>	JK2.1	Chromosome	1	176	4,997,932	51.5	4,815	22	77	195	CP034938.1
	p.Jk2_1	Plasmid	1	176	102,326	48.1	120	-	-	-	CP034939.1
<i>P. brasiliense</i>	HNP201719	Chromosome	1	321	4,724,337	52.3	4,520	22	76	135	CP046380.1
<i>P. carotovorum</i>	BP201601.1	Chromosome	1	190	4,853,176	52.1	4,583	21	78	137	CP034236.1
<i>P. carotovorum</i>	JR1.1	Chromosome	1	240	4,872,902	51.9	4,587	22	76	182	CP034237.1
<i>P. versatile</i>	MYP201603	Chromosome	1	655	4,974,461	51.7	5,187	22	77	189	CP051628.1
	p.MYP201603	Plasmid	1	355	104,776	49.4	122	-	-	-	CP051629.1

CDS, coding sequences.

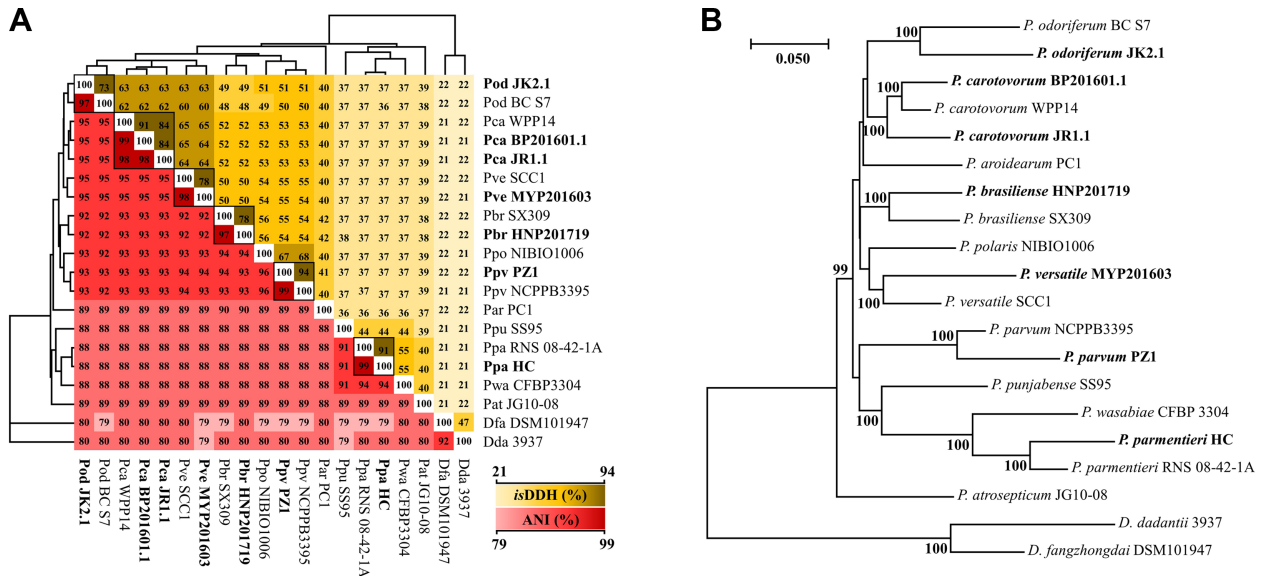


**Fig. 1.** Comparison of whole-genome sequences of seven *Pectobacterium* strains by progressive Mauve alignment. Same local collinear blocks denote conserved parts between genomes. In the genome of a strain, white segments indicate low-similarity or unique regions compare with other genomes. Under boxes of the central line indicate inverted regions. Ppa, *Pectobacterium parmentieri*; Ppv, *P. parvum*; Pod, *P. odoriferum*; Pbr, *P. brasiliense*; Pca, *P. carotovorum*; Pve, *P. versatile*.

observed in other studies (Li et al., 2018, 2019). Complete whole-genome sequencing was performed with the genomes of seven strains, and it was expected to serve as an

additional reference for whole-genome analysis of the *Pectobacterium* genus.

To evaluate conservation and variations among South



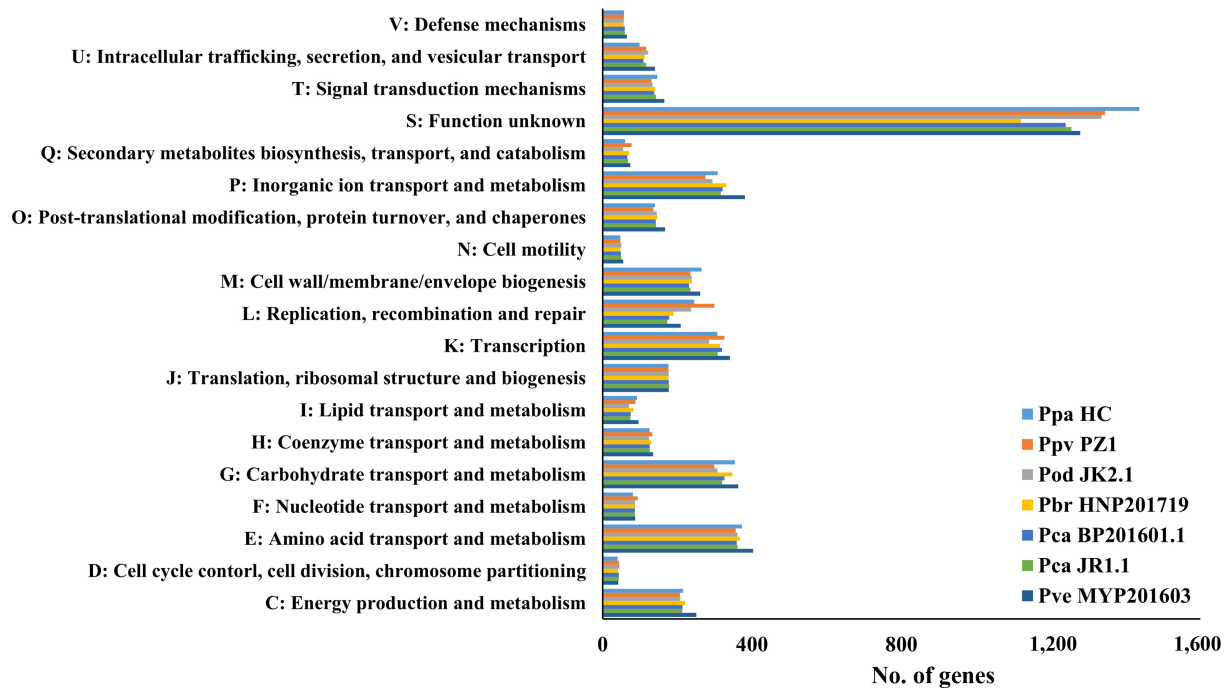
**Fig. 2.** Analysis of average nucleotide identity (ANI) values, *in silico* DNA-DNA hybridization (*isDDH*) values, and phylogeny of the seven isolated strains. (A) Yellow indicates *isDDH*, and red indicates ANI. Two values were calculated as percentages among 20 strains. (B) Phylogenetic tree was constructed with 20 strains including 13 reference strains using binary pan-matrix. Isolated strains are indicated in bold. Pod, *Pectobacterium odoriferum*; Pca, *P. carotovorum*; Pve, *P. versatile*; Pbr, *P. brasiliense*; Ppo, *P. polaris*; Ppv, *P. parvum*; Par, *P. aroidearum*; Ppu, *P. punjabense*; Ppa, *P. parmentieri*; Pwa, *P. wasabiae*; Pat, *P. atrosepticum*; Dfa, *Dickeya fangzhongdai*; Dda, *D. dadantii*.



Korean *Pectobacterium* strain collections, whole-genome sequences were aligned using the Mauve software. As this comparison was made at the species level, low-similarity regions were frequently observed among the genomes (Fig. 1). BP201601.1 showed a large inversion of local collinear blocks compared to JR1.1; however, the ANI value (%) was 98%, as observed with BP201601.1 and JR1.1 (Fig. 2A). The ANI values ranged from 88% to 96% at the species level of *Pectobacterium* in this study. Pve and Pod strains clustered more closely with Pca than with other strains. The Ppa strain, HC, showed the lowest ANI values compared to the other five species. PZ1 and NCPPB3395 considered as type strains of Ppv showed 99% identity in ANI value and 94% identity in the *is*DDH value. PZ1 was also negative for citrate utilization (Supplementary Table 2). Pan-genome phylogeny analysis was also performed with 13 reference strains for the accurate classification of seven isolated strains (Fig. 2B, Supplementary Table 1). BP201601.1 was more similar to WPP14 than to JR1.1, but both were classified as the same Pca species with an *is*DDH value of 70% (Li et al., 2019; Pasanen et al., 2020). Pasanen et al. (2020) reported 96.0-96.2% ANI identity and 66.1-68.4% *is*DDH value between Ppo and Ppv, and this report was similar with the result in this study. On comparing the utilization of four carbon sources, only Ppv isolates could not utilize citrate, among seven species of *Pectobac-*

*terium* including Ppo (Pasanen et al., 2020). The negative response of the PZ1 strain to citrate utilization was considered another piece of evidence supporting the re-classification of PZ1 from Ppo to Ppv (Supplementary Table 2). In this study, PZ1 was newly grouped with Ppv strain NCPPB 3395 and separated from Ppo type strain NIBIO 1006 upon examination of BIOLOG test, ANI value, *is*DDH value, and phylogenetic tree.

**Cluster of orthologous group analysis of South Korean *Pectobacterium* collection.** Cluster of orthologous group (COG) analysis was conducted with the seven *Pectobacterium* strains isolated in South Korea to compare the distribution of functional genes (Fig. 3). An average of 90.7% of the predicted CDSs was classified using the 25 COG categories, and approximately 29.3% were hypothetical proteins. No strain showed genes in five COGs including the general function prediction only group (R), the chromatin structure and dynamics group (B), the extracellular structures group (W), the nuclear structure group (Y), and the cytoskeleton group (Z); meanwhile, all strains had one gene of the RNA processing and modification group (A). The largest COG category in this analysis comprised the amino acid transport and metabolism group (E) with an average of 369 genes, followed by the carbohydrate transport and metabolism group (G) with an average of 331 genes,



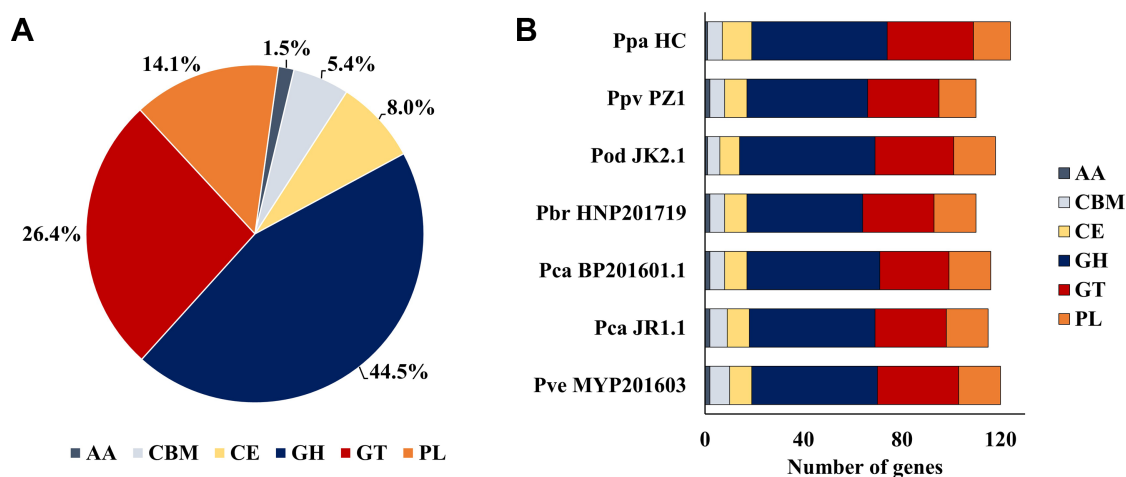
**Fig. 3.** Comparison of coding sequences using cluster of orthologous group classification with seven strains. Ppa, *Pectobacterium parmentieri*; Ppv, *P. parvum*; Pod, *P. odoriferum*; Pbr, *P. brasiliense*; Pca, *P. carotovorum*; Pve, *P. versatile*.

the inorganic ion transport and metabolism group (P) with an average of 318 genes, and the transcription group (K) with an average of 314 genes. Pca, JR1.1, and BP201601.1 showed similar distribution of functional genes. The number of genes showed a deviation of more than 10% among seven strains in the four COGs including the lipid transport and metabolism group (I); the replication, recombination, and repair group (L); the secondary metabolites biosynthesis, transport, and catabolism group (Q); and the intracellular trafficking, secretion, and vesicular transport group (U). The cell cycle control, cell division, chromosome partitioning group (D); the coenzyme transport and metabolism group (H); and the translation, ribosomal structure, and biogenesis group (J) showed a similar distribution of functional genes among the seven strains. MYP201603 had more functional genes in 11 groups. *P. aroidearum* L6 was different in three COGs groups including the defense mechanisms group (V), the cell motility group (N), and the signal transduction mechanisms group (T) (Xu et al., 2021). The Pba strain SX309 differed from HNP201719 in the N group and the T group (Li et al., 2019).

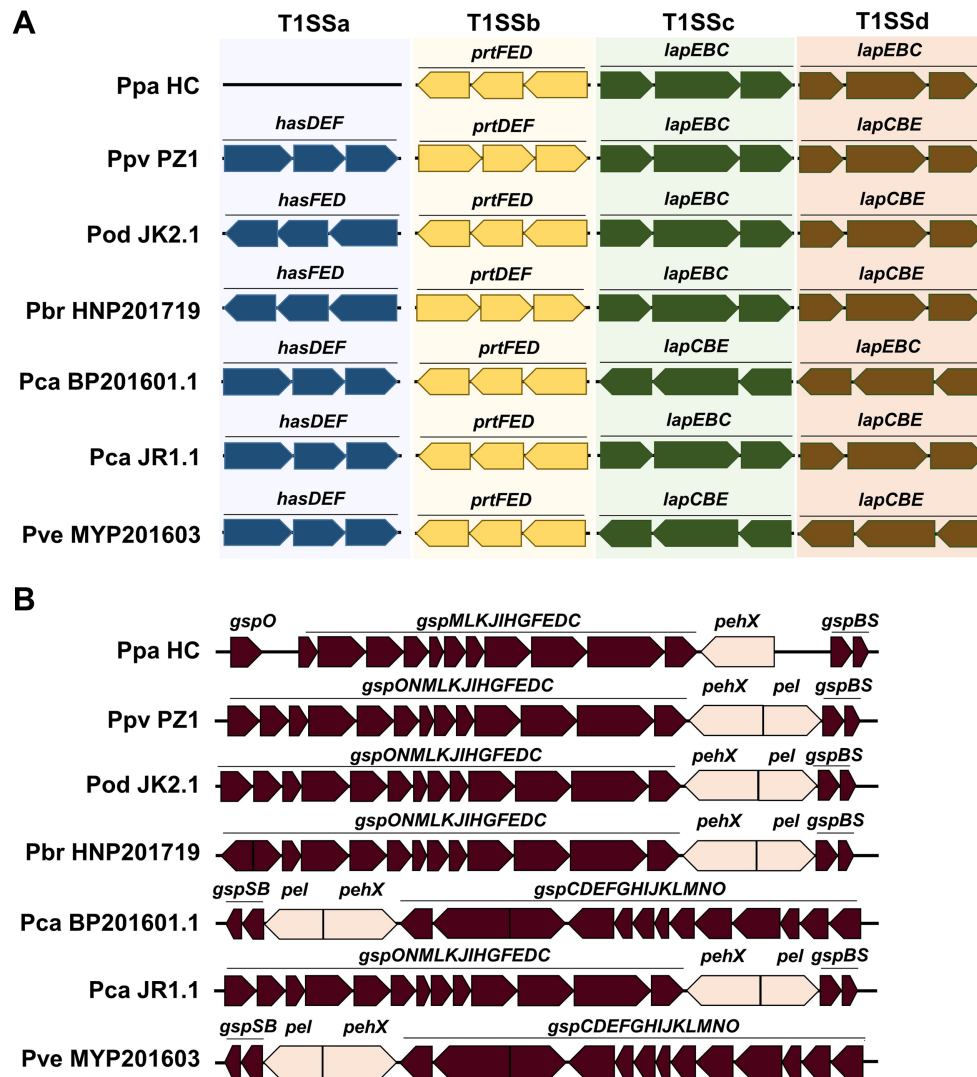
**Comparison of enzymes distributed in the genomes with the analysis of CAZymes and PCWDEs.** CAZyme analysis was conducted with seven strains to identify various enzymes in *Pectobacterium*. All strains showed similar distributions in the CAZymes class (Fig. 4). The average of glycoside hydrolases (GHs) was 44.5%, glycosyltransferases (GTs) was 26.4%, polysaccharide lyases (PLs) was 14.1%, carbohydrate esterases (CEs) was 8.0%,

carbohydrate-binding modules was 5.4%, and auxiliary activity components (AAs) was 1.5% (Fig. 4A). HNP201719 and PZ1 showed the lowest total number of CAZymes, whereas HC demonstrated the highest number. The highest number of two CAZyme classes, namely, CEs and GTs, was identified in HC. HC and JK2.1 had the highest number of GHs. Known PCWDEs were identified in the genomic sequence of the isolated strains (Supplementary Table 3). This analysis focused on secreted PCWDEs with signal peptides, including pectin-degrading enzymes, cellulases, and proteases. The presence of pectin-degrading enzymes and cellulase was similar among five strains, JR1.1, BP201601.1, JK2.1, HNP201719, and MY201603. HC strains lacked at least one enzyme in the three categories of enzymes compared to other strains. JK2.1, HNP201719, and MYP201603 showed more proteases than JR1.1 and BP201601.1 of Pca.

He et al. (2021) conducted CAZyme analysis with Pca strain PC1, grouped in Par by Li et al. (2018). He et al. (2021) identified 144 CAZymes, including 57 GHs, 35 GTs, 26 CEs, 15 PLs, and 8 AAs. The number of AAs and CEs in PC1 was higher than 1-2 in AAs and 8-12 in CEs in this study. The number of PLs was conserved in the *Pectobacterium* species. Several enzymes with signal peptides in the CAZymes analysis were identified in PCWDEs (Supplementary Table 3). These enzymes in the CAZyme category were considered an important tool for mechanisms underlying pathogenesis (He et al., 2021). Three main peptidase families, serine, metallo-, and cysteine super-families, were identified at more than 80% of se-



**Fig. 4.** Distribution of carbohydrate-active enzymes (CAZymes) in *Pectobacterium* species. (A) The average of CAZymes was calculated as percentages among the seven strains. (B) Each CAZymes distribution was compared among the seven *Pectobacterium* strains. AA, auxiliary activities; CBM, carbohydrate-binding module; CE, carbohydrate esterase; GH, glycoside hydrolase; GT, glycosyltransferase; PL, polysaccharide lyase; Ppa, *Pectobacterium parmentieri*; Ppv, *P. parvum*; Pod, *P. odoriferum*; Pbr, *P. brasiliense*; Pca, *P. carotovorum*; Pve, *P. versatile*.

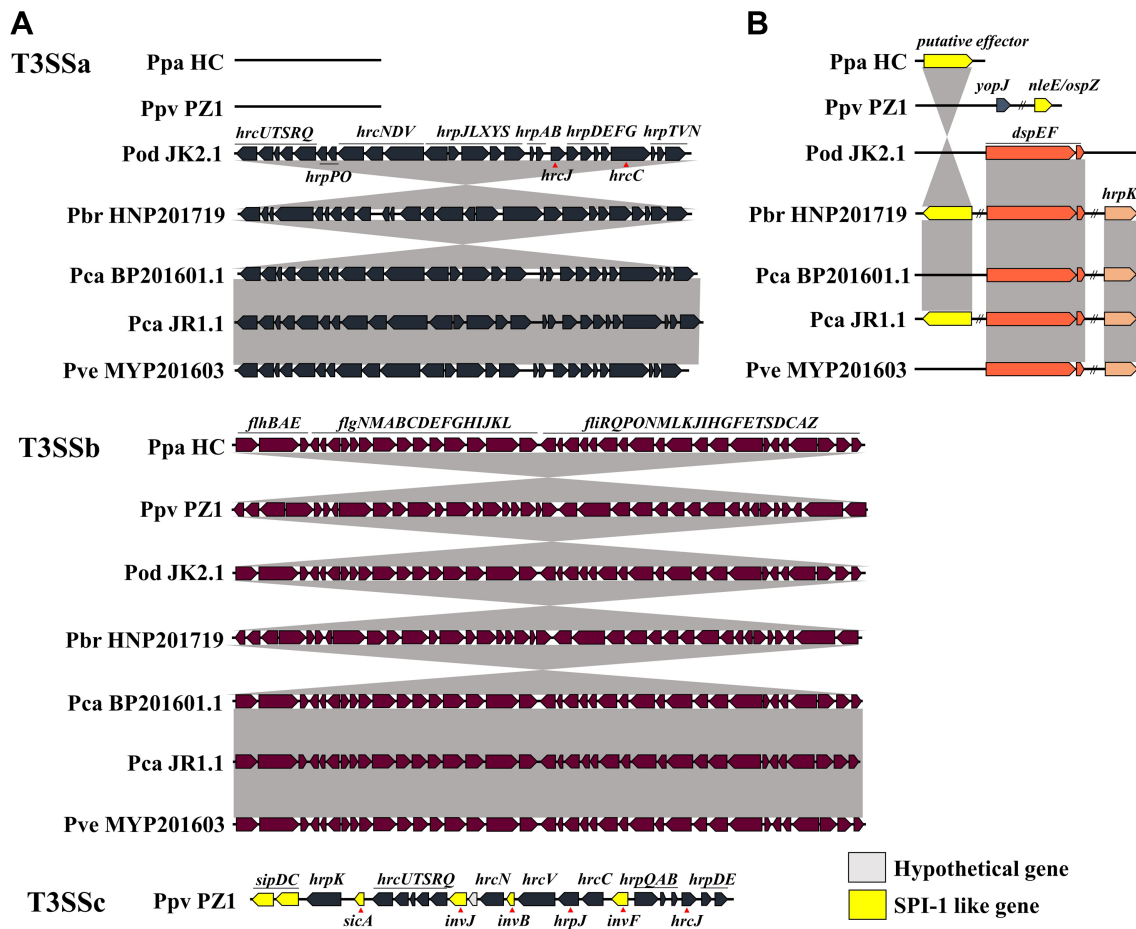


**Fig. 5.** Physical map of type I secretion system (T1SS) and type II secretion system (T2SS) in *Pectobacterium* strains. (A) a-d represent different gene clusters of T1SS. (B) T2SS was compared among seven strains. Ppa, *Pectobacterium parmentieri*; Ppv, *P. parvum*; Pod, *P. odoriferum*; Pbr, *P. brasiliense*; Pca, *P. carotovorum*; Pve, *P. versatile*.

creted peptidase genes of 2,191 bacterial genomes, and the metallopeptidase family was more diverse than the serine-peptidase and cysteine-peptidase families (Nguyen et al., 2019). Figaj et al. (2019) referred to *Pectobacterium* as a phytopathogen with the best-characterized metalloproteases such as Prt1 and PrtW. In the extracellular protease of this study, four serine protease genes were identified along with eight metalloproteases, consistent with the results of Li et al. (2019).

**Comparison of genetic variation in T1SS, T2SS, and T3SS among *Pectobacterium* strains.** To understand the pathogenicity of *Pectobacterium* strains, we compared the gene clusters of T1SS, T2SS, and T3SS among the strains

and identified three subtypes of T1SS and two copies of the *lapBCE* subtype in this collection (Fig. 5A). *Pectobacterium* strains had four T1SS gene clusters, except for the HC strain of Ppa. T1SSa was not detected in the HC group. The T2SS gene cluster typically comprises 15 structural genes linked to *pel* and *pehX* in *Pectobacterium* strains. T2SS was highly conserved in the genomes of the six strains, but HC did not show *gspN* and *pel* (Fig. 5B). Six Ppa strains, including type strains, also did not demonstrate T1SSa similar to HC, and three subsets have been identified in *Pectobacterium* species (Li et al., 2018). In other studies, *P. aroidearum* strain L6 showed 14 genes, and *P. aroidearum* strain PC1 demonstrated one gene (He et al., 2021; Xu et al., 2021). Further study on T1SS and protease secretion is



**Fig. 6.** Open reading frame maps of structural gene clusters and effector genes of type III secretion system (T3SS) in South Korean *Pectobacterium* collection. The double-slash indicates long-distance in the genome. (A) Different gene clusters in a genome sequence are indicated as a-c. (B) *dspEF*, effector-chaperone pair of *Pectobacterium*, putative effectors including *hrpK*, and effectors of animal pathogen, *yopJ* and *nleE/ospZ*, were identified. Ppa, *Pectobacterium parmentieri*; Ppv, *P. parvum*; Pod, *P. odoriferum*; Pbr, *P. brasiliense*; Pca, *P. carotovorum*; Pve, *P. versatile*.

required since Ppa strains, including HC with three T1SS, did not show protease activity in the medium (Jee et al., 2020). T2SS was identified to be highly conserved in this study, similar to other studies (Li et al., 2018, 2019). Xu et al. (2021) reported 23 genes with loss of *gspS* in the T2SS of strain L6. This conservation could be attributed to the importance of T2SS, which secretes extracellular PCWDEs used as the main weapon by *Pectobacterium*.

Three subsets of T3SS were also identified in the *Pectobacterium* strains (Fig. 6A). The gene cluster of T3SSa was conserved in five strains, namely JR1.1, BP201601.1, JK2.1, HNP201719, and MYP201603. All strains had T3SSb, which encodes flagellar components and was highly conserved. Except for the flagellar system, HC did not demonstrate T3SS, and PZ1 showed T3SSc, an alternative to T3SSa. Several SPI-1-related genes were identified in T3SSc. Fig. 6B shows a comparison of the effectors of

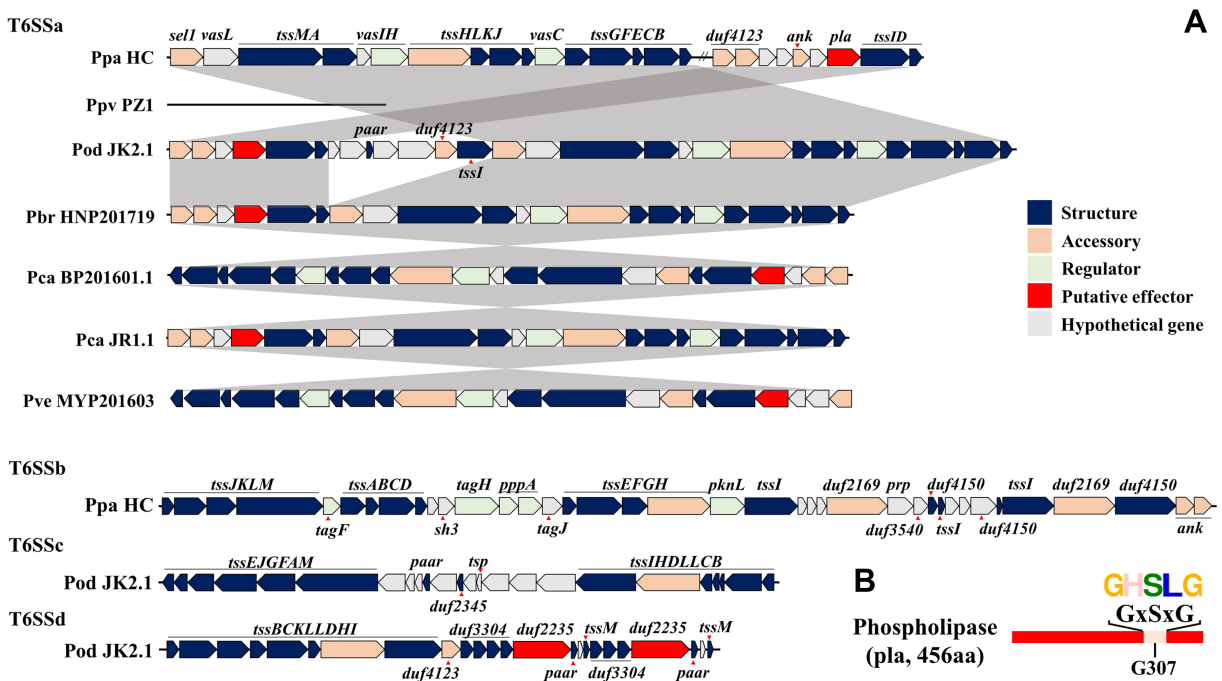
T3SS. Effector genes of *Pectobacterium*, such as *dspE*, *dspF*, and *hrpK*, were identified in all strains except for HC and PZ1. Owing to the query coverage of 16%, *hrpK* of JK2.1 is not shown in Fig. 6B. The putative T3SS effector, *srfC*-like gene, was identified in JR1.1, HNP201719, and HC. PZ1 demonstrated cysteine protease gene *yopJ* and cysteine methyltransferase gene *ospZ*. The putative effector *srfC* has been observed from plant to animal bacterial pathogens (Pallen et al., 2005). DspE is known to induce plant cell death in *Pectobacterium*, but the role of two putative effectors, *hrpK* and *srfC*-like genes, remains unknown (Arizala and Arif, 2019). Loss of T3SSa was observed in six Ppa strains and the Ppo strain NCPPB3395, reclassified as Ppv (Li et al., 2018; Pasanen et al., 2020). Pasanen et al. (2020) identified SPI-1-like T3SS in four of the 18 strains of *P. parvum*. We identified T3SS genes and the effector, namely *yopJ* and *ospZ*, of human pathogens such as *Sal-*



*monella*, *Shigella*, *Yersinia*, and *Escherichia* in PZ1. *sipDC* is known to participate in the formation of the needle complex in SPI-1 T3SS and aids in pathogen internalization and secretion of effectors (Lou et al., 2019). The gene *invJ* controls the length of the needle segment, and *invF* regulates the outer proteins (Sops) of *Salmonella*. *invB* encodes a chaperone for SopA. The original function of the *yopJ* effector is to mediate immunosuppressive reactions, which was identified in *Yersinia* species. The *nleE/ospZ* effector was identified in enteropathogenic *Escherichia* spp., *Shigella* spp., and *Salmonella* spp. *nleE/ospZ* effector of PZ1 showed approximately 67% homology with that of *Salmonella enterica*, and *yopJ* effector in PZ1 showed 48% homology with that of *Yersinia* species in BLAST. YopJ family effectors are widely distributed in diverse bacteria and are homologous to the HopZ effector of *Pseudomonas syringae* (Ma et al., 2006).

**Comparison of T6SS and putative effector genes revealed in the genome of *Pectobacterium* strains.** We identified 13 structural genes, six regulator genes, and five accessory genes in the T6SS gene cluster of seven *Pec-*

*tobacterium* species (Fig. 7A). Structural genes included *tssA-G*, *tssI-M*, and *paar*. Regulator genes included *vasC*, *vasH*, *tagF*, *tagH*, serine/threonine-protein kinase (*pknL*), and serine/threonine phosphatase (*pppA*). Accessory genes comprised putative chaperone genes, including *tssH*, *duf2169*, and *duf4123*, and putative immune genes, including sell-like repeat (*sell*) and ankyrin repeat domain (*ank*). We identified T6SSa, a conserved subset, in four strains, JR1.1, BP201601.1, HNP201719, and MYP201603. We also observed the loss of T6SS in PZ1. HC showed a longer distance between the needle complex and other components of T6SSa and demonstrated a higher number of regulatory genes and accessory genes in T6SSb compared to other strains. JK2.1 showed three subsets of T6SS and an incomplete composition of structural genes in T6SSc and T6SSd. No *tssK* was observed in T6SSc, and *tssA*, *tssE*, *tssF*, and *tssJ* were absent in T6SSd. In the T6SSa cluster, the phospholipase A gene (*pla*) was identified between *tssID* and *duf4123* for all strains. *tssID* represents valine-glycine repeat encoding gene (*vgrG*) and hemolysin-coregulated protein-encoding gene (*hcp*), known as the core components of the needle complex (Cascales and



**Fig. 7.** Open reading frame maps of type VI secretion system (T6SS) and motif of putative effector in South Korean *Pectobacterium* collection. Structural genes included baseplate complex, membrane complex, and tail tube/sheath complex in T6SS. Accessory genes included chaperons and immune genes. Double slash refers to long distances in the genome. (A) a-d represent different gene clusters of the T6SS. (B) Putative effector, *pla*, had the typical motif of phospholipase A. Ppa, *Pectobacterium parmentieri*; Ppv, *P. parvum*; Pod, *P. odoriferum*; Pbr, *P. brasiliense*; Pca, *P. carotovorum*; Pve, *P. versatile*; *sell*, sell-like repeat; *ank*, ankyrin repeat domain; *sh3*, bacterial src homology 3 domain; *pppA*, serine-threonine phosphatase; *pknL*, a serine/threonine-protein kinase; *prp*, pentapeptide repeat; *tsp*, transposase.

Cambillau, 2012) and effectors in T6SS of *Pectobacterium* species (Mattinen et al., 2007).

The DUF4123 protein family has been described as an effector-specific chaperone or adapter. As one of the conserved catalytic motifs of antibacterial phospholipase effectors (Russell et al., 2013), the GxSxG motif is located on G307A in the phospholipase A protein (Fig. 7B). This PLA showed a sequence length of 456 aa and was highly conserved among *Pectobacterium* strains. JK2.1 showed two DUF2235 domain-containing genes with the predicted protein of 805 aa and 813 aa in T6SSd. In JK2.1, this domain was classified as an alpha/beta hydrolase domain and demonstrated a GxSxG motif on G244A and G246A. T6SS is a bacterial machine used to inject an effector to manipulate the host and communicate or fight other bacteria. This system has 13 core components encoded within the same gene cluster (Bernal et al., 2018). T6SS resembles a phage tail and consists of four structures: a contractile sheath constituted by TssB and TssC, a baseplate structure constituted by TssA, TssE, TssF, TssG, and TssK, an integral membrane complex constituted by TssL, TssM, and TssJ, and the needle complex as an inner tube constituted by VgrG (TssI), Hcp (TssD), and PAAR (Bernal et al., 2018).

A chaperone is required for the interaction or secretion of effector and VgrG/PAAR structures. Immune proteins are required for self-protection against effectors in attacking cells.

All strains showed T6SS except for PZ1 as Ppv, which was identified in *P. parvum* strain NCPPB3395. Many T6SS effectors show antibacterial activity, such as peptidoglycan amidase, peptidoglycan glycoside hydrolase, lipase/phospholipase, pore-forming protein, DNase, and NAD(P)+ glycohydrolase (Coulthurst, 2019). *pla* is located in the T6SS cluster of the genome of six South Korean *Pectobacterium* strains and shows the known motif of phospholipase A effector, type VI lipase effectors (Tle) 1-4. *duf2235*, identified in T6SSd, belongs to the Tle1 family, and is known to possess phospholipase A1 and phospholipase A2 activity (Jana and Salomon, 2019). The *hcp* genes are known along with three families as chaperones of effectors, including *duf4123*, *duf1795*, and *duf2169*, usually identified upstream of their cognate effectors (Jana and Salomon, 2019). This finding additionally supports the role of *pla* and *duf2235* as effectors. To date, only Hcp and VgrG are considered effectors in *Pectobacterium*, and this study is the first report showing phospholipase as a putative

Domain organization	No. <i>hcp/vgrG/paar</i> islands						
	Ppa	Ppv	Pod	Pbr	Pca	Pca	Pve
	HC	PZ1	JK2.1	HNP201719	BP201601.1	JR1.1	MYP201603
Hcp	3	0	10	5	6	8	5
VgrG	1	0	0	0	0	1	0
Hcp - VgrG	2	1	1	1	1	1	2
VgrG — PAAR	1	0	0	0	0	0	0
VgrG — DUF4123	0	0	0	1	0	0	0
VgrG — DUF4123 — DUF2235	0	0	0	0	1	0	0
VgrG — DUF4123 — Extra domain — PAAR	1	0	1	0	1	2	1
VgrG — DUF4123 — Alpha/beta hydrolase — Extra domain — PAAR	1	0	0	0	0	0	0
VgrG — VRR-NUC — DUF3396 — PAAR	1	0	0	0	0	0	0
Hcp - VgrG — DUF4123 — Extra domain — PAAR	1	0	0	0	0	0	0
VgrG — DUF4123 — Extra domain — DUF2235 — PAAR	0	0	0	0	0	0	1
VgrG — DUF4123 — DUF2235 — DUF3396 — PAAR	0	0	0	0	0	0	1

**Fig. 8.** Distribution of *hcp/vgrG/paar* islands in the genome of seven *Pectobacterium* strains. The blue box indicates the basic component of *hcp/vgrG/paar* island; yellow box, putative chaperon domain; red box, putative effector domain; green box, putative immune domain. Ppa, *P. parmentieri*; Ppv, *P. parvum*; Pod, *P. odoriferum*; Pbr, *P. brasiliense*; Pca, *Pectobacterium carotovorum*; Pve, *P. versatile*.

effector of T6SS in *Pectobacterium* species.

Two to 14 *hcp/vgrG/paar* islands were identified for each *Pectobacterium* strain (Fig. 8). Single *hcp* varied from 0 to 10 for each strain, but a copy of single *vgrG* was observed in HC and JR1.1. A single *paar* was not identified in any of the strains. The *hcp-vgrG* set was identified with one or two copies in all strains, including PZ1, which did not demonstrate T6SS. *P. wasabiae* SCC3193 had 26 genes for *hcp* or *vgrG*, and *P. wasabiae* WPP163 showed 18 genes for *hcp* or *vgrG* (Nykyri et al., 2012). This is a higher number than that reported here, although Nykyri et al. (2012) did not report *hcp/vgrG/paar* islands. Several domains presumed to have specific functions were found in the *hcp/vgrG/paar* islands of seven strains, including *duf4123*, *duf2235*, alpha/beta hydrolase, *duf2931*, virus-type replication-repair nuclease domain (*vrr-nuc*), and *duf3396*. *duf4123* could be considered a putative chaperone, and *duf2235* and alpha/beta hydrolase could be considered as an effector with phospholipase activity. *duf2931* was suggested as putative immunity proteins on downstream of Tle family effector in *Aeromonas hydrophila* (Ma et al., 2020). Wang et al. (2021) identified TseV-TsiV as a putative effector-immune protein pair, which showed a VRR-NUC domain and DUF3396. VRR-NUC and DUF3396 remain unknown in the *Pectobacterium* genus, and ours is the first report to demonstrate the presence of this pair in *Pectobacterium*.

Our results showed that seven representative strains were classified as six species, Pbr, Pca, Pod, Ppa, Ppv, and Pve, with the ANI, *isDDH*, and phylogenetic tree. The number of PCWDEs was 30-32 in all strains except for HC in which it was 26. Four subtypes of T1SS were identified in six strains. Three subsets were identified in T3SS, and T3SSb was conserved in seven strains. HC showed three subsets of T1SS with the loss of T1SSa and a subset of T3SSb. SPI-1 like T3SS structural gene and effectors were identified in T3SSc of PZ1. Two putative effectors including *hrpK* were identified in seven *Pectobacterium* strains; however, their function was unknown. In T6SS, JR1.1, BP201601.1, HNP201719, and MYP201603 demonstrated a subset of T6SSa, while PZ1 did not include any subset. HC had two subsets including T6SSa and T6SSb, and JK2.1 showed three subsets of T6SS. T6SSa cluster genes harbor a *pla* gene located among *tssID* and *duf4123* genes in six *Pectobacterium* strains. Phospholipase is a representative effector of T6SS with a unique motif, GxSxG, identified in the *pla* of six *Pectobacterium* strains. Considering this, the *pla* gene is a putative effector of T6SS in *Pectobacterium*. Although Pasanen et al. (2020) referred to the neighboring lipase in T6SS of Ppo, the authors did not identify the lipase as an effector. Important domains

were identified in the *hcp/vgrG/paar* islands, including the *duf2235* gene and alpha/beta hydrolase, which potentially has phospholipase activity in BP201601.1, HC, and MYP201603, and the *vrr-nuc* gene, which was referred to as a putative effector with nuclease activity in HC. This report identified the putative effectors with *pla* gene, *duf2235*, and *vrr-nuc* domains in *Pectobacterium* strains. Nevertheless, further studies are warranted to unravel their specific functions in *Pectobacterium* strains.

## Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

## Acknowledgments

This research was supported by the Rural Development Administration (RDA) fund PJ01486602.

## Electronic Supplementary Material

Supplementary materials are available at The Plant Pathology Journal website (<http://www.ppjonline.org/>).

## References

- Arizala, D. and Arif, M. 2019. Genome-wide analyses revealed remarkable heterogeneity in pathogenicity determinants, antimicrobial compounds, and CRISPR-Cas systems of complex phytopathogenic genus *Pectobacterium*. *Pathogens* 8:247.
- Bailey, T. L., Johnson, J., Grant, C. E. and Noble, W. S. 2015. The MEME suite. *Nucleic Acids Res.* 43:W39-W49.
- Bernal, P., Llamas, M. A. and Filloux, A. 2018. Type VI secretion systems in plant-associated bacteria. *Environ. Microbiol.* 20:1-15.
- Cascales, E. and Cambillau, C. 2012. Structural biology of type VI secretion systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367:1102-1111.
- Coulthurst, S. 2019. The Type VI secretion system: a versatile bacterial weapon. *Microbiology* 165:503-515.
- Darling, A. C. E., Mau, B., Blattner, F. R. and Perna, N. T. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394-1403.
- Figaj, D., Ambroziak, P., Przepiora, T. and Skorko-Glonek, J. 2019. The role of proteases in the virulence of plant pathogenic bacteria. *Int. J. Mol. Sci.* 20:672.
- Green, E. R. and Meccas, J. 2016. Bacterial secretion systems: an overview. *Microbiol. Spectr.* 4:10.1128/microbiolspec.VMBF-0012-2015.
- Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., Li, W., Chitsaz, F., Derbyshire, M. K., Gon-

- zales, N. R., Gwadz, M., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Zheng, C., Thibaud-Nissen, F., Geer, L. Y., Marchler-Bauer, A. and Pruitt, K. D. 2018. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* 46:D851-D860.
- He, X., Lu, T. and Zhou, X. 2021. Whole genome sequencing and comparative genomics analysis of *Pectobacterium carotovorum* identifies key pathogenic genes. *Mol. Phylogenet. Evol.* 162:107114.
- Jana, B. and Salomon, D. 2019. Type VI secretion system: a modular toolkit for bacterial dominance. *Future Microbiol.* 14:1451-1463.
- Jee, S., Choi, J.-G., Lee, Y.-G., Kwon, M., Hwang, I. and Heu, S. 2020. Distribution of *Pectobacterium* species isolated in South Korea and comparison of temperature effects on pathogenicity. *Plant Pathol. J.* 36:346-354.
- Lee, D. H., Kim, J.-B., Lim, J.-A., Han, S.-W. and Heu, S. 2014. Genetic diversity of *Pectobacterium carotovorum* subsp. *brasiliensis* isolated in Korea. *Plant Pathol. J.* 30:117-124.
- Lee, D. H., Lim, J.-A., Koh, Y.-J., Heu, S. and Roh, E. 2017. The draft genome sequence of *Pectobacterium carotovorum* subsp. *actinidiae* KKH3 that infects kiwi plant and potential bioconversion applications. *Korean J. Microbiol.* 53:323-325.
- Li, L., Yuan, L., Shi, Y., Xie, X., Chai, A., Wang, Q. and Li, B. 2019. Comparative genomic analysis of *Pectobacterium carotovorum* subsp. *brasiliense* SX309 provides novel insights into its genetic and phenotypic features. *BMC Genomics* 20:486.
- Li, W., O'Neill, K. R., Haft, D. H., DiCuccio, M., Chetvernin, V., Badretin, A., Coulouris, G., Chitsaz, F., Derbyshire, M. K., Durkin, A. S., Gonzales, N. R., Gwadz, M., Lanczycki, C. J., Song, J. S., Thanki, N., Wang, J., Yamashita, R. A., Yang, M., Zheng, C., Marchler-Bauer, A. and Thibaud-Nissen, F. 2021. RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res.* 49:D1020-D1028.
- Li, X., Ma, Y., Liang, S., Tian, Y., Yin, S., Xie, S. and Xie, H. 2018. Comparative genomics of 84 *Pectobacterium* genomes reveals the variations related to a pathogenic lifestyle. *BMC Genomics* 19:889.
- Lou, L., Zhang, P., Piao, R. and Wang, Y. 2019. *Salmonella* pathogenicity island 1 (SPI-1) and its complex regulatory network. *Front. Cell. Infect. Microbiol.* 9:270.
- Ma, S., Dong, Y., Wang, N., Liu, J., Lu, C. and Liu, Y. 2020. Identification of a new effector-immunity pair of *Aeromonas hydrophila* type VI secretion system. *Vet. Res.* 51:71.
- Ma, W., Dong, F. F. T., Stavrinides, J. and Guttman, D. S. 2006. Type III effector diversification via both pathoadaptation and horizontal transfer in response to a coevolutionary arms race. *PLoS Genet.* 2:e209.
- Mattinen, L., Nissinen, R., Riipi, T., Kalkkinen, N. and Pirhonen, M. 2007. Host-extract induced changes in the secretome of the plant pathogenic bacterium *Pectobacterium atrosepticum*. *Proteomics* 7:3527-3537.
- Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L. and Göker, M. 2021. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res.* 50:D801-D807.
- Moussa, H. B., Pédrón, J., Bertrand, C., Hecquet, A. and Barny, M.-A. 2021. *Pectobacterium quasiquaticum* sp. nov., isolated from waterways. *Int. J. Syst. Evol. Microbiol.* 71:005042.
- Nguyen, T. T. H., Myrold, D. D. and Mueller, R. S. 2019. Distributions of extracellular peptidases across prokaryotic genomes reflect phylogeny and habitat. *Front. Microbiol.* 10:413.
- Nykyri, J., Niemi, O., Koskinen, P., Nokso-Koivisto, J., Pasanen, M., Broberg, M., Plyusnin, I., Törönen, P., Holm, L., Pirhonen, M. and Palva, E. T. 2012. Revised phylogeny and novel horizontally acquired virulence determinants of the model soft rot phytopathogen *Pectobacterium wasabiae* SCC3193. *PLoS Pathog.* 8:e1003013.
- Pallen, M. J., Beatson, S. A. and Bailey, C. M. 2005. Bioinformatics, genomics and evolution of non-flagellar type-III secretion systems: a Darwinian perspective. *FEMS Microbiol. Rev.* 29:201-229.
- Park, D. H., Kim, J. S., Lee, H. G., Hahm, Y. I. and Lim, C. K. 1999. Black leg of potato plants by *Erwinia carotovora* subsp. *atroseptica*. *Plant Dis. Agric.* 5:64-66.
- Park, T.-H., Choi, B.-S., Choi, A.-Y., Choi, I.-Y., Heu, S. and Park, B.-S. 2012. Genome sequence of *Pectobacterium carotovorum* subsp. *carotovorum* strain PCC21, a pathogen causing soft rot in Chinese cabbage. *J. Bacteriol.* 194:6345-6346.
- Pasanen, M., Waleron, M., Schott, T., Cleenwerck, I., Misztak, A., Waleron, K., Pritchard, L., Bakr, R., Degefu, Y., van der Wolf, J., Vandamme, P. and Pirhonen, M. 2020. *Pectobacterium parvum* sp. nov., having a *Salmonella* SPI-1-like type III secretion system and low virulence. *Int. J. Syst. Evol. Microbiol.* 70:2440-2448.
- Portier, P., Pédrón, J., Taghouti, G., Dutrieux, C. and Barny, M. A. 2020. Updated taxonomy of *Pectobacterium* genus in the CIRM-CFBP bacterial collection: when newly described species reveal “Old” endemic population. *Microorganisms* 8:1441.
- Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., Gabaldón, T., Rattei, T., Creevey, C., Kuhn, M., Jensen, L. J., von Mering, C. and Bork, P. 2014. eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.* 42:D231-D239.
- Russell, A. B., LeRoux, M., Hathazi, K., Agnello, D. M., Ishikawa, T., Wiggins, P. A., Wai, S. N. and Mougous, J. D. 2013. Diverse type VI secretion phospholipases are functionally plastic antibacterial effectors. *Nature* 496:508-512.
- Tatusova, T., DiCuccio, M., Badretin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K. D., Borodovsky, M. and Ostell, J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44:6614-6624.
- Visnovsky, S. B., Panda, P., Taylor, R. and Pitman, A. R. 2017. Draft genome sequences of *Pectobacterium carotovorum*



- subsp. *actinidiae* ICMP 19971 and ICMP 19972, two strains isolated from *Actinidia chinensis* with symptoms of summer canker in South Korea. *Genome Announc.* 5:e00104-e00117.
- Wang, S., Geng, Z., Zhang, H., She, Z. and Dong, Y. 2021. The *Pseudomonas aeruginosa* PAAR2 cluster encodes a putative VRR-NUC domain-containing effector. *FEBS J.* 288:5755-5767.
- Xu, P., Wang, H., Qin, C., Li, Z., Lin, C., Liu, W. and Miao, W. 2021. Analysis of the taxonomy and pathogenic factors of *Pectobacterium aroidearum* L6 using whole-genome sequencing and comparative genomics. *Front. Microbiol.* 12:679102.
- Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y. and Yin, Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46:W95-W101.