As a step towards discovering novel pathogenesis-related proteins, we performed a genome scale computational identification and characterization of secreted and transmembrane (TM) proteins, which are mainly responsible for bacteria-host interactions and interactions with other bacteria, in the genomes of six representative *Burkholderia* species. The species comprised plant pathogens (*B. glumae* BGR1, *B. gladioli* BSR3), human pathogens (*B. pseudomallei* K96243, *B. cepacia* LO6), and plant-growth promoting endophytes (*Burkholderia* sp. KJ006, *B. phytofirmans* PsJN). The proportions of putative classically secreted proteins (CSPs) and TM proteins among the species were relatively high, up to approximately 20%. Lower proportions of putative type 3 non-classically secreted proteins (T3NCSPs) (~10%) and unclassified non-classically secreted proteins (NCSPs) (~5%) were observed. The numbers of TM proteins among the three clusters (plant pathogens, human pathogens, and endophytes) were different, while the distribution of these proteins according to the number of TM domains was conserved in which TM proteins possessing 1, 2, 4, or 12 TM domains were the dominant groups in all species.

NCSPs. Furthermore, we assigned the complete sets of the conserved and unique NCSP candidates of the collected *Burkholderia* species using sequence similarity searching. This study could provide new insights into the relationship among plant-pathogenic, human-pathogenic, and endophytic bacteria.

**Keywords**: *Burkholderia*, endophyte, human pathogen, plant pathogen, secreted protein

**Handling Associate Editor**: Oh, Chang-Sik

Secreted and transmembrane (TM) proteins are crucial agents that initiate communication between bacteria and the outside environment, as well as mediating infection of host cells or other bacterial competitors, participating in both harmful and beneficial interactions (Collmer, 1998; Costa et al., 2015; Tseng et al., 2009). In gram-negative bacteria, the main determinant factors of pathogenesis are the effector proteins, which are usually secreted through non-classical pathways of the type I, III, IV, and VI secretion systems (named as T1SS, T3SS, T4SS, and T6SS, respectively) (Büttner and He, 2009; Feng and Zhou, 2012; Schell et al., 2007). When the effectors encounter plant or human cells, they are capable of promoting virulence through breaking up and repressing the cells’ immune signals. These effectors are typically categorized as non-classically secreted proteins (NCSPs), which have no signal peptides, and contain uncommon or diverse patterns of the amino acids in their corresponding sequence regions (Arnold et al., 2009; Bendtsen et al., 2004; Kampenusa and Zikmanis, 2010). By contrast, the proteins secreted through the general secretion (Sec) or twin-arginine translocation (Tat) pathways of the type II, V, VII secretion systems (termed as T2SS, T5SS, and T7SS, respectively), and sometimes the T4SS, are categorized as classically
secreted proteins (CSPs) (Nielsen and Krogh, 1998; Saier, 2006; Tseng et al., 2009). These proteins utilize Sec or Tat signal peptides to penetrate their inner cell membrane via the Sec or Tat translocons, respectively. Signal peptides are normally found at the N-terminus, greater than 11 residues, started by a positively charged n-region, followed by a core hydrophobic region, and a c-region (Nielsen and Krogh, 1998; Petersen et al., 2011). Besides secreted proteins, TM proteins also perform various functions vital to the survival of microorganisms, and are involved in the initial microbe-host interaction. TM proteins usually represent a high fraction of the total proteins of bacterial genomes (Chiba et al., 2008; Engel and Gaub, 2008; Saier, 2006).

The *Burkholderia* genus includes over 60 species, which are found in a variety of ecological niches, including humid areas and industrial zones (Estrada-de los Santos et al., 2013; Weisskopf et al., 2011). According to the phylogenetic analyses, this genus can be divided into two large clusters, in spite of its wide range, including the cluster of plant or human pathogens, and the cluster of plant-associated species pathogens (Estrada-de los Santos et al., 2013; Weisskopf et al., 2011). Of the plant pathogens, two species *B. glumae* and *B. gladioli* are emergent agents that cause serious diseases on rice, such as seedling blight, panicle blight, grain rot, and sheath rot, resulting in heavy yield losses in many countries worldwide (Harn et al., 2011; Lee et al., 2016; Nandakumar et al., 2009; Ura et al., 2006). While *B. pseudomallei* and *B. cepacia* have been well-known as important representative species of human pathogens because they are opportunistic pathogens that are common agents of hospital-associated infections, which act by repressing the immune system of animals and humans, causing melioidosis (by *B. pseudomallei*) and cystic fibrosis disease (by *B. cepacia*) (Govan and Deretic, 1996; Wiersinga et al., 2006). Of the latter cluster, some recently discovered *Burkholderia* species are non-pathogenic bacteria termed endophytes, which possess the ability to not only promote the growth and development of plants by enhancing their adaption to environmental changes, but also protect them from other pathogenic bacteria (Reinhold-Hurek and Hurek, 2011; Santoyo et al., 2016). The genomes of many endophytes have been sequenced recently; however, only a small number of them have been reported two species of *B. phytofirmans* PsJN and *Burkholderia* strain KJ006 (Mitter et al., 2013; Santoyo et al., 2016). The *aaiA* gene-producing *Burkholderia* sp. KJ006 can attenuate the mechanisms of quorum sensing and virulence of *B. glumae*; therefore, *Burkholderia* sp. KJ006 might represent a promising biocontrol agent to repress the plant pathogens (Cho et al., 2007). For these reasons, six *Burkholderia* species of *B. glumae*, *B. gladioli*, *B. pseudomallei*, *B. cepacia*, *B. phytofirmans* PsJN, and *Burkholderia* sp. KJ006 have been chosen as representatives for our comparative analyses among plant pathogens, human pathogens, and endophytes.

Increasing the productivity of crops, together with protecting the plants from pathogens, are important tasks in the field of plant science; however, they are still challenges to researchers because of the quick and complicated evolution of virulence-transferring secretion systems of bacteria, as well as climate changes leading to unhealthy environments for plants (Costa et al., 2015; Naughton et al., 2016; Park et al., 2014). The protection or damage mediated by endophytic and pathogenic bacteria on host cell, respectively, have been observed phenotypically; however, the details of the molecular interaction mechanisms, especially the proteins directly involved in host-bacterial interactions are still not fully understood or are often analyzed separately.

In this study, we performed the computational identification and comparative analysis of putative secreted proteins and TM proteins on genome-scale in six different *Burkholderia* species. The data allowed us to study the relationship between *Burkholderia* plant pathogenic and human pathogenic or plant-growth promoting bacterial species *Burkholderia*, to gain a better understanding of their interaction mechanisms, and to identify further novel pathogenesis-related proteins. The protein sequence data collected for analysis were for six *Burkholderia* genomes representing plant pathogens, human pathogens, and the plant-growth promoting endophytes.

**Materials and Methods**

**Bacterial strains.** The representative *Burkholderia* strains analyzed in this study comprised two plant pathogens (*B. glumae* BGR1, *B. gladioli* BSR3), two human pathogens (*B. pseudomallei* K96243, *B. cepacia* LO6), and two non-pathogenic endophytes (*B. phytofirmans* PsJN, *Burkholderia* sp. KJ006). The whole genome sequences of these strains were extracted from the RefSeq database at the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/). All duplicated proteins in the downloaded datasets were removed. Information concerning the secretion systems existing in each *Burkholderia* strain was also examined, based on the available annotations of individual proteins and bacterial secretion systems on Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (http://www.genome.jp/kegg/pathway.html).

**Prediction and characterization of proteins.** Three pro-
grams, SignalP (http://www.cbs.dtu.dk/services/SignalP/), TatP (http://www.cbs.dtu.dk/services/TatP/), and LipoP (http://www.cbs.dtu.dk/services/LipoP/) (Bendtsen et al., 2005; Juncker et al., 2003; Petersen et al., 2011), with default cutoff values, were used to predict signal peptides in the bacterial proteins. Secreted proteins that do not contain any clear signal peptide at their N-terminus are termed the NCSPs, which are mostly secreted through the one-step non-conventional secretion pathways such as T1SS, T3SS, T4SS, and T6SS. To identify these proteins, we utilized the SecretomeP server, which is used widely to predict secreted proteins, especially the NCSPs. with a S-score ≥ 0.5 (http://www.cbs.dtu.dk/services/SecretomeP/) (Bendtsen et al., 2004). In addition, the EffectiveT3 (Arnold et al., 2009) and T3SS_prediction (Löwer and Schneider, 2009) servers were used to recognize potential type 3 effectors, which are NCSPs that are possibly secreted through the T3SS pathway, and are termed as type 3 non-classically secreted proteins (T3NCSPs). The TMHMM program (http://www.cbs.dtu.dk/services/TMHMM/), which is considered as the best algorithm to recognize TM proteins with a hidden Markov model (HMM), through performing the prediction of TM domains and topology of proteins, was used to eliminate the potential TM proteins among the putative secreted proteins and to recognize potential TM proteins in the rest of the genomes (Krogh et al., 2001; Möller et al., 2001). All predicted secreted proteins that possessed at least two potential TM domains were moved into the TM protein group; those that contained one potential TM domain remained in the secreted proteins group if the location of the TM domain overlapped with or was next to the position of the predicted signal peptide, otherwise they were also moved into the TM protein group. The distribution of TM proteins according to number of TM domains was also established.

Sequence analysis. The orthologous proteins among six genomes of the Burkholderia species were extracted by scanning BLASTp search (https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the non-redundant proteins of six genomes in which the coverage and identity were set up to 50% and the E-value was set up to E-10 to obtain the orthologs with high confidence. The distributions of orthologs in various groups of predicted secreted and TM proteins were examined. The protein sizes of putative secreted proteins in various groups were analyzed and compared among the six Burkholderia strains. The DeGO prediction server (http://supfam.org/SUPERFAMILY/deGO/) was utilized to assign the Gene Ontology (GO) functions of the secreted proteins, according to three categories of molecular function (MF), cellular component (CC), and biological process (BP) (Fang and Gough, 2013). The BLASTp server was also used to infer the conserved and unique putative NCSPs among the six representative Burkholderia species, in which two proteins were considered as homologs if they shared a statistically significant similarity with an identity ≥ 30% and an E-value < E-5 (Altschul et al., 1990).

Results and Discussion

General characteristics of the genomes. The genome information of the six Burkholderia strains used in this study is shown in Table 1. Two strains of plant pathogens comprised B. glumae BGR1 isolated from grain and B. gladioli BSR3 isolated from the sheath parts of rice in Korea (Lim et al., 2009; Seo et al., 2011). The B. cepacia LO6 strain was recovered from an infected cystic fibrosis patient, and belongs to the B. cepacia genomovar VI, a new member of the B. cepacia complex (Belcaid et al., 2015). B. pseudomallei K96243 was isolated from a melioidosis patient in Thailand in 1996 and its genome was sequenced completely in 2004 (Holden et al., 2004). Burkholderia sp. KJ006 is an endophytic bacterium of rice with antifungal activity (Kwak et al., 2012) and B. phytofirmans PsJN can produce the beneficial effects on plants such as tomato, potato, cucumber, and grape (Mitter et al., 2013; Weilharter et al., 2011).

The possible secretion systems existing in the target Burkholderia strains were investigated according to the available information of bacterial secretion systems at the

| Organism                  | Size (Mb) | GC (%) | Host/isolation source                  | Geographic location | Reference       |
|---------------------------|-----------|--------|----------------------------------------|---------------------|-----------------|
| B. glumae BGR1            | 7.28      | 67.93  | Rice                                   | Korea               | Lim et al., 2009|
| B. gladioli BSR3          | 9.05      | 67.41  | Rice                                   | Korea               | Seo et al., 2011|
| B. pseudomallei K96243    | 7.25      | 68.05  | Human melioidosis                      | Thailand            | Holden et al., 2004|
| B. cepacia LO6            | 6.42      | 67.00  | Pulmonary/cystic fibrosis              | Thailand            | Belcaid et al., 2015|
| B. phytofirmans PsJN      | 8.21      | 62.32  | Surface-sterilized onion root          | Canada              | Weilharter et al., 2011|
| Burkholderia sp. KJ006    | 6.63      | 67.19  | Rice root                              | Korea               | Kwak et al., 2012|
KEGG PATHWAY database (Kanehisa et al., 2006) and in the literature (Holden et al., 2004; Mitter et al., 2013; Seo et al., 2015). The information concerning the secretion systems in *B. cepacia* LO6 was derived based on the annotations of the individual proteins in its whole genome (Table 2). Besides the conventional secretion pathway T2SS, the Tat and Sec translocons are usually found in most gram-negative bacteria. The non-conventional secretion pathways of T3SS and T6SS, which have been reported to transport the virulence determinants of bacterial pathogens into the host cell (Alfano and Collmer, 2004; Schell et al., 2007), were also found in all six collected bacterial strains. Among these bacteria, the *B. pseudomallei* strain might have the most complicated secretion systems observed so far: three T3SS clusters and six evolutionally distinct T6SSs, which are potentially involved in intra-macrophage growth, have been discovered in this species (Schell et al., 2007; Shalom et al., 2007). In addition, some studies have shown that T6SS appears commonly and T3SS is extremely rare in most of the identified endophytes (Reinhold-Hurek and Hurek, 2011; Xia et al., 2015); however, according to our examination, T3SS is present in both rice endophytes, *B. phytofirmans* PsJN and *Burkholderia* sp. KJ006. This suggested that T3SS might play a significant role in promoting plant growth, as well as in interactions with other pathogenic or endophytic bacteria in the same host.

**Computational identification of secreted and TM proteins.** The total numbers of putative secreted proteins were determined from the full datasets of the non-redundant proteins of the six *Burkholderia* genomes and were classified into various groups (Table 3). The CSP group comprised secreted proteins that contained a signal peptide predicted by at least one of the three servers (SignalP, TatP, and LipoP). The average proportion of putative CSPs obtained from the six genomes was significantly high (20.7 ± 0.8%). The two plant pathogens had relatively similar proportions of putative CSPs (19.7% for *B. glumae* BGR1; 20.3% for *B. gladioli* BSR3), despite of their different genome sizes. The two human pathogens had the highest percentages of CSPs (21.9% for *B. cepacia* LO6 and 21.3% for *B. pseudomallei* K96243). The two remaining endophytes, *B. phytofirmans* PsJN and *Burkholderia* sp. KJ006, had intermediate levels of CSPs, at 20.1% and 20.8%, respectively. In addition, among the predicted CSPs, the proteins recognized only by the TatP program (CSP1-T) were dominant (7.8 ± 0.7%) compared

| Secretion pathway | *B. glumae* BGR1 | *B. gladioli* BSR3 | *B. pseudomallei* K96243 | *B. cepacia* LO6 | *B. phytofirmans* PsJN | *Burkholderia* sp. KJ006 |
|-------------------|------------------|---------------------|-------------------------|------------------|------------------------|-------------------------|
| T1SS              | Yes              | Unclear             | Unclear                 | Yes              | Unclear                | Unclear                |
| T2SS              | Yes              | Yes                 | Yes                     | Yes              | Yes                    | Yes                    |
| T3SS              | Yes              | Yes                 | Yes                     | Yes              | Yes                    | Yes                    |
| T4SS              | Yes              | Unclear             | Unclear                 | Yes              | Yes                    | Yes                    |
| T5SS              | Unclear          | Unclear             | Unclear                 | Unclear          | Unclear                | Unclear                |
| T6SS              | Yes              | Yes                 | Yes                     | Yes              | Yes                    | Yes                    |
| Tat translocon    | Yes              | Yes                 | Yes                     | Unclear          | Yes                    | Yes                    |
| Sec translocon    | Yes              | Yes                 | Yes                     | Yes              | Yes                    | Yes                    |

Tat, twin-arginine translocation; Sec, general secretory.

| Group                | *B. glumae* BGR1 | *B. gladioli* BSR3 | *B. pseudomallei* K96243 | *B. cepacia* LO6 | *B. phytofirmans* PsJN | *Burkholderia* sp. KJ006 |
|----------------------|------------------|---------------------|-------------------------|------------------|------------------------|-------------------------|
| CSP                  | 1,161 (19.7)     | 1,492 (20.3)        | 1,222 (21.3)            | 1,213 (21.9)     | 1,426 (20.1)           | 1,165 (20.8)            |
| T3NCSP               | 660 (11.2)       | 795 (10.8)          | 574 (10.0)              | 520 (9.4)        | 804 (11.3)             | 535 (9.6)              |
| Unclassified NCSP    | 314 (5.3)        | 292 (4.0)           | 268 (4.7)               | 256 (4.6)        | 346 (4.9)              | 244 (4.4)              |
| TM protein           | 1,075 (18.3)     | 1,331 (18.1)        | 1,112 (19.4)            | 1,051 (19.0)     | 1,415 (20.0)           | 1,126 (20.2)            |

Values are presented as number (%). % values refer to the fraction of proteins in each group compared with total non-redundant proteins in the whole genome.

TM, transmembrane; CSP, classically secreted protein; T3NCSP, type 3 non-classically secreted protein; NCSP, non-classically secreted protein.
with those recognized only by either SignalP (0.9 ± 0.2%) or Lipop (1.7 ± 0.1%). The CSP1-T levels were lowest in the endophytes (6.9% for *B. phytofirmans* PsJN; 7.6% for *Burkholderia* sp. KJ006), followed by the plant pathogens (7.3% for *B. gladioli* BSR3; 8.1% for *B. glumae* BGR1), and were highest in the human pathogens (8.5% for *B. cepacia* LO6; 8.6% for *B. pseudomallei* K96243) (Table 4). The high fractions of putative Tat proteins in all six *Burkholderia* species suggested that the Tat translocation system could be essential, and that the Tat secreted proteins might play an important role in the interactions and pathogenic processes of these bacteria.

The T3NCSPs, which are NCSPs that are possibly secreted through the non-conventional T3SS pathway, comprised proteins that were predicted to be potential type 3 effectors by either the EffectiveT3 or the T3SS_prediction servers. The fractions of putative T3NCSPs recognized by combining results of both prediction servers in all strains were significantly high (10.4 ± 0.8%), in which the two plant pathogens, *B. glumae* BGR1 and *B. gladioli* BSR3, had similar percentages at 11.2% and 10.8%, respectively; the endophyte *B. phytofirmans* PsJN had the highest percentage (11.3%), and the human pathogen *B. cepacia* LO6 had the lowest percentage (9.4%) (Table 3). It is notable that the average proportion of T3NCSPs was approximately two times the average fraction of the predicted remaining NCSPs (4.6 ± 0.5%), which were predicted to be secreted proteins only by the SecretomeP program. Their specific secretion systems are unknown and they are thus named as the unclassified NCSP group (Table 3). For these these NCSP proteins, *B. glumae* BGR1 had the highest proportion (5.3%), while *B. gladioli* BSR3 had the lowest proportion (4%); the two human pathogens had similar proportions (4.7% for *B. pseudomallei* K96243; 4.6% for *B. cepacia* LO6). The two endophytes, *B. phytofirmans* PsJN and *Burkholderia* sp. KJ006, had proportions of 4.9% and 4.4%, respectively.

The six *Burkholderia* genomes also received the considerable proportions of TM proteins (19.1 ± 0.9%) recognized out of fully non-redundant proteins in whole genomes (Table 3). Remarkably, the percentages of TM proteins within the plant pathogens, human pathogens, and endophytes were quite similar, in which the two plant pathogens had the lowest percentages (18.3% for *B. glumae* BGR1; 18.1% for *B. gladioli* BSR3), two human pathogens had higher percentages (19.4% for *B. pseudomallei* K96243; 19.0% for *B. cepacia* LO6), and the two endophytes had the highest percentages (20.0% for *B. phytofirmans* PsJN; 20.2% for *Burkholderia* sp. KJ006). However, these TM protein proportions (18.1–20.2%) were clearly lower than ones predicted in many other bacterial genomes (24–31%) (Stevens and Arkin, 2000). The explanation plausible for this could come from moving collectively some proteins predicted by TMHMM program into CSP group. These proteins were recognized to have a TM domain and a signal peptide, simultaneously, but have the TM domain overlapping with or standing next to the signal peptide at N-terminal sequence parts. In particular, these proportions of such these proteins ranged from 3.6% to 4.4% for six genomes.

In addition, to get a more comprehensive insight about gene contents of *Burkholderia* species, the orthologs which were considered as core proteins in this study were carried out of six genomes with around 2,448 proteins. *B. cepacia* LO6 and *Burkholderia* sp. KJ006 received the highest ratios of core protein in their genomes of 44.2% and 43.8%, respectively. Next the *B. glumae* BGR1 got 41.5% and *B. pseudomallei* got 42.5% of core proteins in their genomes. Finally, two species *B. gladioli* BSR3 and *B. phytofirmans* PsJN gained the lowest proportions of 33.2% and 34.5%, due to large sizes of their genomes. The distributions of the core proteins in groups of CSPs,

Table 4. Classically secreted proteins (CSPs) predicted in the *Burkholderia* strains

| Organism       | B. glumae BGR1 | B. gladioli BSR3 | B. pseudomallei K96243 | B. cepacia LO6 | B. phytofirmans PsJN | Burkholderia sp. KJ006 |
|----------------|----------------|------------------|-------------------------|---------------|----------------------|------------------------|
| CSP3           | 217 (4.0)      | 297 (4.0)        | 243 (4.2)               | 209 (3.8)     | 260 (3.7)            | 204 (3.7)              |
| CSP2           | 320 (5.4)      | 480 (6.5)        | 346 (6.0)               | 366 (6.6)     | 489 (6.9)            | 380 (6.8)              |
| CSP1-S         | 49 (0.8)       | 58 (0.8)         | 50 (0.9)                | 73 (1.3)      | 66 (0.9)             | 51 (0.9)               |
| CSP1-T         | 477 (8.1)      | 538 (7.3)        | 493 (8.6)               | 469 (8.5)     | 487 (6.9)            | 423 (7.6)              |
| CSP1-L         | 98 (1.7)       | 119 (1.6)        | 90 (1.6)                | 96 (1.7)      | 124 (1.7)            | 107 (1.9)              |
| Total CSPs     | 1,161 (19.7)   | 1,492 (20.3)     | 1,222 (21.3)            | 1,213 (21.9)  | 1,426 (20.1)         | 1,165 (20.8)           |

Values are presented as number (%). CSP3 and CSP2 are the CSPs that were predicted to contain a signal peptide by all three and two prediction tools, respectively. CSP1-S, CSP1-T, and CSP1-L refer to the CSPs that were predicted to have a signal peptide by only one of three servers (SignalP, TatP, or Lipop) respectively.
Secreted and Transmembrane Proteins in *Burkholderia*

T3NCSPs, unclassified NCSPs, and TM proteins of six *Burkholderia* species were found to be quite similar in overall (Supplementary Fig. 1). The ratios of core proteins in *B. gladioli* BSR3 and *B. phytofirmans* PsJN were still lower in all groups and approximately equal in two groups of CSP and TM protein, but unequal in two groups of T3NCSP and unclassified NCSP.

The distribution of TM proteins. The TM domains appearing in each TM protein were identified based on the TMHMM program, except some TM proteins recognized by only the LipoP program without specific TM domain number. We found that the distribution of TM proteins based on the number of TM domains in all six collected *Burkholderia* species was somewhat conserved, in spite of their different lifestyles (Fig. 1). The fractions of TM proteins that possess one or two TM domains were the highest, especially for *B. glumae* BGR1. Additionally, the proportions of TM proteins that contained four, five, six, or twelve TM domains were also higher than the other TM proteins: *B. cepacia* LO6 had the highest proportion

Fig. 1. The distribution of transmembrane (TM) proteins according to the number of TM domains in six *Burkholderia* strains. The vertical axis shows the number of TM proteins possessing the corresponding number of TM domains on the horizontal axis.

Fig. 2. Length distribution of three secreted protein groups consisting of putative classically secreted proteins (CSPs), type 3 non-classically secreted proteins (T3NCSPs), and unclassified non-classically secreted proteins (NCSPs) obtained from six *Burkholderia* strains. The vertical axis shows the number of secreted proteins with the corresponding length scale on the horizontal axis in each group.
of TM proteins that possessed 12 TM domains.

**Length distribution of secreted proteins.** Fig. 2 shows the distribution of putative secreted proteins in the six *Burkholderia* strains according to the lengths of various groups, including CSPs, T3NCSPs, and unclassified NC-SPs. Conservation of length distributions was observed in each group of secreted proteins, although the proteins belonging to different groups of plant pathogens, human pathogens, and endophytes varied in length. The largest unclassified NCSP group had lengths varying from 100 to 200 residues (around 35%), the second group varied from 0 to 100 residues (around 20–25%), and the third group varied from 201 to 300 residues (around 15%). Among the T3NCSP proteins, the largest group varied from 200 to 300 residues (around 35%), and lower proportions were observed for the groups of 100 to 200 residues, and from 300 to 400 residues (around 20%). Finally, in the CSP group, the numbers of proteins whose lengths varied from 101 to 200, from 201 to 300, and from 302 to 400 residues were the highest (around 20%).

**Functional analysis of putative secreted proteins.** The distributions of GO terms characterizing MF, CC, and BP for the CSPs, T3NCSPs, and unclassified NCSPs in six *Burkholderia* genomes are shown in Fig. 3–5, respectively. The horizontal bars represent the ratios of the secreted proteins assigned each GO term to the total number of putative secreted proteins in the corresponding group, as annotated by the dcGO program at the “general” level (http://supfam.org/SUPERFAMILY/dcGO/).

In the MF assessment, the terms oxidoreductase activity and anion binding were dominant for all three secreted protein groups, while the terms nucleic acid binding, transferase activity, transferring phosphorus-containing group and ligase activity were dominant for the T3NCSPs and unclassified NCSPs. In addition, some functional terms were highly represented in one of three groups: transporter activity, receptor activity, and signal transducer activity terms in the CSP group; small molecule binding and nucleoside phosphate binding terms in the T3NCSP group; and structural molecule activity term in the unclassified NCSP group. In addition, species-specific differences were observed in the MF terms: *B. glumae* BGR1 T3NCSPs were associated predominantly with transferase activity and transferring phosphorus-containing group; its NCSP proteins were associated with nucleic acid binding, enzyme regulator activity and receptor activity terms. In addition, there were less significant associations with receptor activity and transporter activ-

---

**Fig. 3.** Gene Ontology (GO) analysis of putative classically secreted proteins (CSPs). The bars present the ratios (%) of proteins specified in each term to the total CSPs annotated by the dcGO program (at the “general” level) of six *Burkholderia* strains. The GO terms with low percentages were eliminated.
Secreted and Transmembrane Proteins in *Burkholderia*

Fig. 4. Gene Ontology (GO) analysis of putative type 3 non-classically secreted proteins (T3NCSPs). The bars present the ratios (%) of proteins specified in each term to the total T3NCSPs annotated by dcGO program (at the “general” level) of six *Burkholderia* strains. The GO terms with low percentages were eliminated.

Fig. 5. Gene Ontology (GO) analysis of putative unclassified non-classically secreted proteins (NCSPs). The bars present the ratios (%) of proteins specified in each term to the total unclassified NCSPs annotated by dcGO program (at the “general level”) of six *Burkholderia* strains. The GO terms with low percentages were eliminated.
ity terms for the CSP group and with the anion-binding term in the T3NCSP group compared with those of other species. In *B. gladioli* BSR3, the association with transporter activity in the unclassified NCSP group was stronger than that in other species. *B. pseudomallei* K96243 CSP proteins were more significantly associated with oxidoreductase activity, transferase activity, transferring phosphorus-containing group, and nucleic acid binding terms compared with those of other species. *B. cepacia* LO6 CSP proteins showed high associations with transporter activity, receptor activity, and signal transducer activity terms; the T3NCSP group was highly associated with receptor binding; and the NCSP group was highly associated with the structural molecule activity term. *B. phytofirmans* PsJN CSP proteins were highly associated with transporter activity, receptor activity and signal transducer activity terms; the T3NCSP group were highly associated with transporter activity, receptor activity and signal transducer activity terms; the T3NCSP group was highly associated with the small molecule binding and nucleoside phosphate binding terms; and NCSP group was associated with oxidoreductase activity. Finally, *Burkholderia* sp. KJ006 T3NCSP proteins were highly associated with anion binding, nucleoside phosphate binding, and small molecule binding terms; and the unclassified NCSP group was associated with anion binding.

In the CC assessment, the CSP group attained the highest fractions of some CC terms (i.e., plasma membrane part and intrinsic to membrane in *B. phytofirmans* PsJN) up to approximately 45%, which was significantly higher than those of the T3NCSP and unclassified NCSP groups (approximately 30%). All three groups of secreted proteins were significantly associated with the plasma membrane part terms. The CSP group was significantly associated with the intrinsic to membrane and cell projection part terms, while the T3NCSP and unclassified NCSP groups were both associated with the nuclear lumen, mitochondrion, and organelle membrane terms. In the BP assessment, the terms of cellular catabolic process, organic substance catabolic process, and organic acid metabolic process were significantly associated with all three groups of secreted proteins. In addition, we observed that the terms cellular protein modification process and nucleobase-containing small molecule metabolic process were significantly associated with the CSP group, the nucleic acid metabolic process was significantly associated with the unclassified NCSP group, while T3NCSP group was significantly associated with carbohydrate derivative metabolic process, organophosphate metabolic process, and nucleobase-containing small molecule metabolic process terms.

Overall, the distribution of significant GO terms according to groups of secreted proteins was consistent for all collected bacterial strains. However, it was difficult to determine any conservation in the proportions of specific GO terms among six species and even between two species with similar lifestyles (i.e., plant pathogen, human pathogen, and endophyte), especially in the MF and CC assessments. This indicated that the functional features of the secreted proteins in the *Burkholderia* species are most likely species-specific rather than lifestyle-specific, and thus would lead to distinct characteristics in their communication process with the host cells. Another study also demonstrated that although virulent microorganisms and endophytes seem to possess genetically similar weaponry, their expression and regulatory mechanisms are different (López-Fernández et al., 2015; Xu et al., 2014). Further studies of the expressions of specific proteins in each species and their interactions during communication with host cells might explain the differences between these pathogenic and mutualistic bacteria (López-Fernández et al., 2015; Seo et al., 2015).

The conserved and unique type 3 effector candidates. The T3SS effectors play an essential role to many pathogenic bacteria because these virulence genes can typically be inserted directly into host cells via the complicated T3SS, which commonly is constituted by 15 to 25 core genes including *hrp* (hypersensitive response and pathogenicity) and *hrc* (*hrc* conserved) genes (Alfano and Collmer, 2004; Feng and Zhou, 2012). The type III effectors in *Pseudomonas syringae* pathovars, the plant pathogens, have been discovered that can suppress both plant immune system including pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) and effector-triggered immunity (ETI) (Block and Alfano, 2011; Cunnac et al., 2009). Moreover, some studies have reported that T3SS effectors can be virulent for both plant and human hosts (Duarte et al., 2000; He et al., 2004; van Baarlen et al., 2007). To identify the T3SS effector candidates with high confidence, we collected only the proteins that were recognized by both the EffectiveT3 and T3SS_ prediction programs among the previously predicted T3NCSPs (McDermott et al., 2011), and were termed T3NCSP-2 (Table 5). Using sequence similarity searching via BLASTp, the unique and conserved T3SS effectors were obtained from the collected *Burkholderia* species. In particular, the unique effectors had no homologous T3NCSP-2 proteins in the other five species, while the conserved effectors shared statistically significant similarity (E_value < 1e-5, minimal identity 30%) with at least one T3NCSP-2 protein in the five remaining species. The number of conserved and unique T3SS effectors among the six *Burkholderia* species, along with the number of proteins conserved between separate query and subjective species are presented in Table 5. As expected, the two
Secreted and Transmembrane Proteins in *Burkholderia*

plant pathogens *B. glumae* BGR1 and *B. gladioli* BSR3 had the highest number of homologous T3SS effectors (79 for *B. glumae* BGR1 and 75 for *B. gladioli* BSR3). However, the number of homologs within the two human pathogens and within the two endophytes was not emergent compared with their homologs found in the other remaining species. In particular, the human pathogen *B. cepacia* LO6 had a large number of homologous T3SS effectors (81 proteins) compared with the endophyte *Burkholderia* sp. KJ006 (79 proteins), regardless of their different lifestyles. This result was consistent with a study of genome comparisons between a human pathogen and an endophyte strain in which the authors reported that the endophyte strain caused mild virulence in a mouse model test system, while the human pathogen strain possessed the genes relevant for survival inside plants, such as those associated with nitrogen fixation, transport, protection against oxidative agents, and polysaccharide degradation (Fouts et al., 2008).

The detailed list of these conserved putative T3SS effectors is shown in Table 6. These homologous proteins among six *Burkholderia* species have various functions of translation (i.e., 50S ribosomal protein L28), transport (i.e., ABC transporter ATP-binding protein, ATP-binding protein), regulation (i.e., AraC family transcriptional regulator, sigma-54 dependent regulatory protein, Fis family transcriptional regulator), enzymes (i.e., FMN-dependent NADH-azoreductase, cytidylate kinase, alkyl hydroperoxide reductase subunit C, protocatechuate 3,4-dioxygenase subunit beta), and unknown function (i.e., hypothetical proteins). Of the putative conserved type 3 effectors, the alkyl hydroperoxide reductase (peroxiredoxin) protein was found as the extracellular protein regulated by HrpB transcriptional activator (Kang et al., 2008). We found no any homolog between these proteins with the experimented type 3 effectors of T3SEdb database (Tay et al., 2010). It is possible come from the feature that the type 3 effector usually have distant sequences and have no any clear motif or signal peptide in N-terminal or C-terminal sequences to be recognized (Kang et al., 2008; Tay et al., 2010). However, considering that these *Burkholderia* species belong to three various lifestyles, the putative type 3 effectors conserved across in all species may play the role in important common functions.

### Table 5. Conserved and unique type 3 effector candidates in six *Burkholderia* strains

| Query strain | T3NCSP-2 | Conserved proteins* | Proteins conserved in subjective strains | Unique proteins |
|--------------|----------|---------------------|------------------------------------------|-----------------|
|              |          |                     | BGR1 | BSR3 | K96243 | LO6 | PsJN | KJ006 |
| BGR1         | 197      | 11 (5.6)            | -    | 79   | 47     | 58  | 40   | 53    | 87 (44.2) |
| BSR3         | 233      | 12 (5.2)            | 75   | -    | 53     | 51  | 54   | 55    | 110 (47.2) |
| K96243       | 164      | 11 (6.7)            | 46   | 46   | -      | 48  | 35   | 44    | 77 (47.0) |
| LO6          | 141      | 13 (9.2)            | 52   | 51   | 51     | -   | 42   | 81    | 38 (27.0) |
| PsJN         | 228      | 16 (7.0)            | 42   | 54   | 42     | 45  | -    | 50    | 134 (58.8) |
| KJ006        | 149      | 16 (10.7)           | 54   | 59   | 50     | 79  | 46   | -     | 41 (27.5) |

Values are presented as number only or number (%).

*The proteins were conserved in the query strain and in all five remaining studied strains.

The conserved and unique unclassified NCSP candidates. Due to the essential role of NCSPs in term of interactions between bacteria and host (Bendtsen et al., 2004; Schell et al., 2007; Tseng et al., 2009), besides the results of T3NCSP groups we also extracted the unique and conserved proteins among the unclassified NCSP groups of six *Burkholderia* species by using sequence similarity searching via BLASTp. Similarly, the unique proteins had no any homologous protein in the other five species, while the conserved proteins shared statistically significant similarity (E_value < 1e-5, minimal identity 30%) with at least one protein in the five remaining species. The ratios of conserved proteins in these groups were clearly higher than those of T3NCSP groups (Table 7). *B. phytofirmans* PsJN and *B. cepacia* LO6 received the lowest proportions of 14% and 15%, respectively, while *Burkholderia* sp. KJ006 got the proportion up to 23% out of putative unclassified NCSPs. In addition, the number of unique proteins that could be the reason to make the species-specific differences, along with the number of proteins conserved between separate query and subjective species, were also presented in Table 7. Of such the unique proteins, while three strains of *B. glumae* BGR1, *B. gladioli* BSR3, and *B. phytofirmans* PsJN got the close proportions (47%, 46%, 46%), the *Burkholderia* sp. KJ006 had remarkably less ratio (28%).

In conclusion, the results of this study could lead to a better understanding of the general features of secreted and TM proteins, and the relationships among *Burkholderia*...
Table 6. Conserved type 3 effector candidates in six *Burkholderia* strains

| Protein name | Definition | Locus ID |
|--------------|------------|----------|
| **B. glumae** BGR1 | | |
| WP_004186391.1 | MS: 50S ribosomal protein L28 | bglu_1g28680 |
| WP_012734522.1 | FMN-dependent NADH-azoreductase | bglu_1g04330 |
| WP_012734950.1 | Cytidylate kinase | bglu_1g08790 |
| WP_015875283.1 | ABC transporter | bglu_1g12300 |
| WP_015875291.1 | AraC family transcriptional regulator | bglu_1g12380 |
| WP_012734097.1 | Sigma-54-dependent Fis family transcriptional regulator | bglu_1p1190 |
| WP_012733828.1 | Sigma-54-dependent Fis family transcriptional regulator | bglu_1g12300 |
| WP_012734950.1 | Cytidylate kinase | bglu_1g08790 |
| WP_015875283.1 | ABC transporter | bglu_1g12300 |
| WP_015875291.1 | AraC family transcriptional regulator | bglu_1g12380 |
| WP_012734097.1 | Sigma-54-dependent Fis family transcriptional regulator | bglu_1p1190 |
| WP_012733828.1 | Sigma-54-dependent Fis family transcriptional regulator | bglu_1g12300 |
| **B. gladioli** BSR3 | | |
| WP_013696521.1 | FMN-dependent NADH-azoreductase | bgla_1g04700 |
| WP_013696998.1 | Cytidylate kinase | bgla_1g09630 |
| WP_013697336.1 | AraC family transcriptional regulator | bgla_1g13140 |
| WP_013697551.1 | Fis family transcriptional regulator | bgla_1g15580 |
| WP_013698566.1 | AraC family transcriptional regulator | bgla_1g26190 |
| WP_004186391.1 | MS: 50S ribosomal protein L28 | bglu_1g28680 |
| WP_013690079.1 | MS: alkyl hydroperoxide reductase subunit C | bgla_2g13050 |
| WP_013690224.1 | Protocatechuate 3,4-dioxigenase subunit beta | bgla_2g14510 |
| WP_013691065.1 | MS: hypothetical protein | bgla_2g23080 |
| WP_013691078.1 | ABC transporter | bgla_2g23220 |
| WP_013691620.1 | MS: alkyl hydroperoxide reductase subunit C | bgla_2g28810 |
| WP_013691697.1 | ABC transporter | bgla_2g29570 |
| **B. pseudomallei** K96243 | | |
| YP_106977.1 | Fis family transcriptional regulator | BPSL0350 |
| YP_107543.1 | 50S ribosomal protein L28 | BPSL0916 |
| YP_108790.1 | AraC family transcriptional regulator | BPSL2195 |
| YP_109070.1 | Sigma-54-dependent regulatory protein | BPSL2475 |
| YP_109112.1 | Cytidylate kinase | BPSL2516 |
| YP_109327.1 | AraC family transcriptional regulator | BPSL2731 |
| YP_109651.1 | ACP phosphodiesterase | BPSL3056 |
| YP_110488.1 | Putrescine ABC transporter ATP-binding protein | BPSS0466 |
| YP_110513.1 | Alkyl hydroperoxide reductase | BPSS0492 |
| YP_110726.1 | Hypothetical protein | BPSS0712 |
| YP_111309.1 | Protocatechuate 3,4-dioxigenase | BPSS1300 |
| **B. cepacia** LO6 | | |
| WP_006399806.1 | MS: peroxiredoxin | - |
| WP_006764866.1 | MS: cytidylate kinase | - |
| WP_006765287.1 | MS: FMN-dependent NADH-azoreductase | - |
| WP_006765678.1 | MS: hypothetical protein | - |
| WP_006766507.1 | MS: AraC family transcriptional regulator | - |
| WP_006767319.1 | MS: AraC family transcriptional regulator | - |
| WP_035973621.1 | MS: sigma-54-dependent Fis family transcriptional regulator | - |
| WP_035973986.1 | MS: protocatechuate 3,4-dioxigenase subunit beta | - |
| WP_035976167.1 | MS: AraC family transcriptional regulator | - |
| WP_045552239.1 | MS: nitrate ABC transporter ATP-binding protein | - |
| WP_045552403.1 | MS: transporter | - |
| WP_045552421.1 | MS: nitrate ABC transporter ATP-binding protein | - |
| WP_004186391.1 | MS: 50S ribosomal protein L28 | - |
Table 6. Continued

| Protein name | Definition | Locus ID |
|--------------|------------|----------|
| *B. phytofirmans* PsJN | | |
| WP_012426127.1 | Transcriptional regulator | Bphyt_4230 |
| WP_012426304.1 | ABC transporter ATP-binding protein | Bphyt_4412 |
| WP_012426718.1 | ATP-binding protein | Bphyt_4841 |
| WP_012427269.1 | Peroxiredoxin | Bphyt_5403 |
| WP_012428554.1 | Protocatechuate 3,4-dioxygenase subunit beta | Bphyt_6760 |
| WP_012432154.1 | AraC family transcriptional regulator | Bphyt_1112 |
| WP_012432584.1 | AraC family transcriptional regulator | Bphyt_1557 |
| WP_012433284.1 | AraC family transcriptional regulator | Bphyt_2287 |
| WP_012433486.1 | AraC family transcriptional regulator | Bphyt_2493 |
| WP_012433835.1 | Phosphate ABC transporter ATP-binding protein | Bphyt_2855 |
| WP_012433975.1 | Cytidylate kinase | Bphyt_3002 |
| WP_012434141.1 | ABC transporter ATP-binding protein | Bphyt_3178 |
| WP_012434224.1 | Hypothetical protein | Bphyt_3263 |
| WP_012434453.1 | FMN-dependent NADH-azoreductase | Bphyt_3502 |
| WP_041759656.1 | Propionate catabolism operon regulatory protein PrpR | - |

*Burkholderia* sp. KJ006

| Query strain | Unclassified NCSPs | Conserved proteins* | Proteins conserved in subjective strains | Unique proteins |
|--------------|--------------------|---------------------|------------------------------------------|-----------------|
| BGR1         | 314                | 57 (18.2)           | - 123 97 112 91 107                     | 149 (47.5)      |
| BSR3         | 292                | 58 (19.9)           | 126 - 98 97 94 98                       | 133 (45.5)      |
| K96243       | 268                | 55 (20.5)           | 101 101 - 98 101 102                    | 105 (39.2)      |
| LO6          | 256                | 38 (14.8)           | 105 93 100 - - 102 91                   | 93 (36.3)       |
| PsJN         | 346                | 50 (14.5)           | 102 103 113 115 -                       | 103 (46.0)      |
| KJ006        | 244                | 57 (23.4)           | 107 94 103 127 113 -                    | 68 (27.9)       |

Values are presented as number only or number (%).
*The proteins were conserved in the query strain and in all five remaining studied strains.

MS, multispecies.
eria species, which comprised harmful bacteria and bacteria that benefit their plant or human host. Most of the studied bacterial strains possessed the determinant secretion systems for pathogenesis, especially T3SS and T6SS. The numbers of putative CSPs and TM proteins obtained from all species were significantly high, reaching approximately 20%; however, there were lower numbers of putative T3NCSPs (~10%) and unclassified NCSPs (~5%). The proportions of TM proteins among the three groups of plant pathogens, human pathogens, and endophytes were different; however, the distribution of such proteins according to number of TM domains was likely conserved. In addition, we observed conservation in the protein size distribution of the secreted protein groups among the species. There were also species-specific differences in the functional characteristics of the secreted proteins in the various groups (i.e., CSPs, T3NCSPs, unclassified NCSPs), together with distinct features among the groups. Finally, the complete sets of conserved and unique T3SS effector candidates in the selected Burkholderia species were assigned based on sequence similarity searching. To the best of our knowledge, this is the first report of a genome-scale comparative analysis of secreted and TM proteins among plant-pathogenic, human-pathogenic, and plant-growth promoting endophytic bacteria of genome-sequenced Burkholderia species.

Acknowledgments

This research was supported by a grant from the Strategic initiative for Microbiomes in Agriculture and Food, Ministry of Agriculture, Food and Rural Affairs, Republic of Korea (No. 916009-2).

References

Alfano, J. R. and Collmer, A. 2004. Type III secretion system effector proteins: double agents in bacterial disease and plant defense. Annu. Rev. Phytopathol. 42:385-414.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Arnold, R., Brandmaier, S., Kleine, F., Tischler, P., Heinz, E., Behrens, S., Niinkoski, A., Mewes, H. W., Horn, M. and Rattei, T. 2009. Sequence-based prediction of type III secreted proteins. PLoS Pathog. 5:e1000376.

Belcaid, M., Kang, Y., Tuanyok, A. and Hoang, T. T. 2015. Complete genome sequence of Burkholderia cepacia strain LO6. Genome Announc. 3:e00587-15.

Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G. and Brunak, S. 2004. Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng. Des. Sel. 17:349-356.

Bendtsen, J. D., Nielsen, H., Widdick, D., Palmer, T. and Brunak, S. 2005. Prediction of twin-arginine signal peptides. BMC Bioinformatics 6:167.

Block, A. and Alfano, J. R. 2011. Plant targets for Pseudomonas syringae type III effectors: virulence targets or guarded decoys? Curr. Opin. Microbiol. 14:39-46.

Böttner, D. and He, S. Y. 2009. Type III protein secretion in plant pathogenic bacteria. Plant Physiol. 150:1656-1664.

Chiba, H., Osanai, M., Murata, M., Kojima, T. and Sawada, N. 2008. Transmembrane proteins of tight junctions. Biochim. Biophys. Acta 1778:588-600.

Cho, H. S., Park, S. Y., Ryu, C. M., Kim, J. F., Kim, J. G. and Park, S. H. 2007. Interference of quorum sensing and virulence of the rice pathogen Burkholderia glumae by an engineered endophytic bacterium. FEMS Microbiol. Ecol. 60:14-23.

Collmer, A. 1998. Determinants of pathogenicity and avirulence in plant pathogenic bacteria. Curr. Opin. Plant Biol. 1:329-335.

Costa, T. R., Felisberto-Rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M. and Waksman, G. 2015. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nat. Rev. Microbiol. 13:343-359.

Cunnac, S., Lindeberg, M. and Collmer, A. 2009. Pseudomonas syringae type III secretion system effectors: repertoires in search of functions. Curr. Opin. Microbiol. 12:53-60.

Duarte, X., Anderson, C. T., Grimson, M., Barabote, R. D., Strauss, R. E., Gollahon, L. S. and San Francisco, M. J. 2000. Erwinia chrysanthemi strains cause death of human gastrointestinal cells in culture and express an intimin-like protein. FEMS Microbiol. Lett. 190:81-86.

Engel, A. and Gaub, H. E. 2008. Structure and mechanics of membrane proteins. Annu. Rev. Biochem. 77:127-148.

Estrada-de los Santos, P., Vinuesa, P., Martinez-Aguilar, L., Hirsch, A. M. and Caballero-Mellado, J. 2013. Phylogenetic analysis of Burkholderia species by multilocus sequence analysis. Curr. Microbiol. 67:51-60.

Fang, H. and Gough, J. 2013. DeGO: database of domain-centric ontologies on functions, phenotypes, diseases and more. Nucleic Acids Res. 41:D536-D544.

Feng, F. and Zhou, J. M. 2012. Plant-bacterial pathogen interactions mediated by type III effectors. Curr. Opin. Plant Biol. 15:469-476.

Fouts, D. E., Tyler, H. L., DeBoy, R. T., Daugherty, S., Ren, Q., Badger, J. H., Durkin, A. S., Huot, H., Shrivastava, S., Kohari, S., Dodson, R. J., Mohamoud, Y., Khouri, H., Roesch, L. F., Krogfelt, K. A., Struve, C., Tripplett, E. W. and Methé, B. A. 2008. Complete genome sequence of the N2-fixing broad host range endophyte Klebsiella pneumoniae 342 and virulence predictions verified in mice. PLoS Genet. 4:e1000141.

Govan, J. R. and Deretic, V. 1996. Microbial pathogenesis in cystic fibrosis: mucoid Pseudomonas aeruginosa and Burkholderia cepacia. Microbiol. Rev. 60:539-574.

Ham, J. H., Melanson, R. A. and Rush, M. C. 2011. Burkholderia glumae: next major pathogen of rice? Mol. Plant Pathol. 12:329-339.

He, S. Y., Nomura, K. and Whittam, T. S. 2004. Type III protein
secretion mechanism in mammalian and plant pathogens. *Biochim. Biophys. Acta* 1694:181-206.

Holden, M. T., Titball, R. W., Peacock, S. J., Cerdeño-Tárraga, A. M., Atkins, T., Crossman, L. C., Pitt, T., Churcher, C., Mungall, K., Bentley, S. D., Sebaiyah, M., Thomson, N. R., Bason, N., Beacham, I. R., Brooks, K., Brown, K. A., Brown, N. F., Challis, G. L., Cherevach, I., Chillingworth, T., Cronin, A., Crossett, B., Davis, P., DeShazer, D., Feltwell, T., Fraser, A., Hance, Z., Hauser, H., Holroyd, S., Jagels, K., Keith, K. E., Maddison, M., Moule, S., Price, C., Quail, M. A., Rabinowitsch, E., Rutherford, K., Sanders, M., Simmonds, M., Songsvilai, S., Stevens, K., Tumapa, S., Vesaratchavest, M., Whitehead, S., Yeats, C., Barrell, B. G., Oyston, P. C. and Parkhill, J. 2004. Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc. Natl. Acad. Sci. U. S. A.* 101:14240-14245.

Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H. and Krogh, A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* 12: 1652-1662.

Kampenusa, I. and Zikmanis, P. 2010. Distinguishable codon usage and amino acid composition patterns among substrates of leaderless secretory pathways from proteobacteria. *Appl. Microbiol. Biotechnol.* 86:285-293.

Kanehisa, M., Goto, S., Kawashima, M., Araki, M. and Hirakawa, M. 2006. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 34:D354-D357.

Kang, Y., Kim, J., Kim, S., Kim, H., Lim, J. Y., Kim, M., Kwak, J., Moon, J. S. and Hwang, I. 2008. Proteomic analysis of the proteins regulated by HrpB from the plant pathogenic bacterium *Burkholderia glumae*. *Proteomics* 8:106-121.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305:567-580.

Kwak, M. J., Song, J. Y., Kim, S. Y., Jeong, H., Kang, S. G., Kim, B. K., Kwon, S. K., Lee, C. H., Yu, D. S., Park, S. H. and Kim, J. F. 2012. Complete genome sequence of the endophytic bacterium *Burkholderia gladioli* sp. strain KJ006. *J. Bacteriol.* 194:4432-4433.

Lee, J., Park, J., Kim, S., Park, I. and Seo, Y. S. 2016. Differential regulation of toxofilin production and its role in the enhanced virulence of *Burkholderia gladioli*. *Mol. Plant Pathol.* 17:65-76.

Lim, J., Lee, T. H., Nahm, B. H., Choi, Y. D., Kim, M. and Hwang, I. 2009. Complete genome sequence of *Burkholderia glumae* BGR1. *J. Bacteriol.* 191:3758-3759.

López-Fernández, S., Sonago, P., Moretto, M., Pancher, M., Engelen, K., Pertot, I. and Campisano, A. 2015. Whole-genome comparative analysis of virulence genes unveils similarities and differences between endophytes and other symbiotic bacteria. *Front. Microbiol.* 6:419.

Löwer, M. and Schneider, G. 2009. Prediction of type III secretion signals in genomes of Gram-negative bacteria. *PLoS One* 4:e5917.

McDermott, J. E., Corrigan, A., Peterson, E., Oehmen, C., Niemann, G., Cambron, E. D., Sharp, D., Adkins, J. N., Samudrala, R. and Heffron, F. 2011. Computational prediction of type III and IV secreted effectors in Gram-negative bacteria. *Infect. Immun.* 79:23-32.

Mitter, B., Petric, A., Shin, M. W., Chain, P. S., Hauberg-Lotte, L., Reinhold-Hurek, B., Nowak, J. and Sessitsch, A. 2013. Comparative genome analysis of *Burkholderia phytofirmans* PsJN reveals a wide spectrum of endophytic lifestyles based on interaction strategies with host plants. *Front. Plant Sci.* 4: 120.

Möller, S., Croning, M. D. and Apweiler, R. 2001. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 17:464-465.

Nandakumar, R., Shahjahan, A. K. M., Yuan, X. L., Dickstein, E. R., Groth, D. E., Clark, C. A., Cartwright, R. D. and Rush, M. C. 2009. *Burkholderia glumae* and *B. gladioli* cause bacterial panicle blight in rice in the southern United States. *Plant Dis.* 93:896-905.

Naughton, L. M., An, S. Q., Hwang, I., Chou, S. H., He, Y. Q., Tang, J. L., Ryan, R. P. and Dow, J. M. 2016. Functional and genomic insights into the pathogenesis of *Burkholderia* species to rice. *Environ. Microbiol.* 18:780-790.

Nielsen, H. and Krogh, A. 1998. Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 6:122-130.

Park, S., Seo, Y. S. and Hegeman, A. D. 2014. Plant metabolomics for plant chemical responses to belowground community change by climate change. *J. Plant Biol.* 57:137-149.

Petersen, T. N., Brunak, S., von Heijne, G. and Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8:785-786.

Reinhold-Hurek, B. and Hurek, T. 2011. Living inside plants: bacterial endophytes. *Curr. Opin. Plant Biol.* 14:435-443.

Saier, M. H., Jr. 2006. Protein secretion and membrane insertion systems in gram-negative bacteria. *J. Membr. Biol.* 214:75-90.

Santoyo, G., Moreno-Hagelsieb, G., del Carmen Orozco-Mosqueda, M. and Glick, B. R. 2016. Plant growth-promoting bacterial endophytes. *Microbiol. Res.* 183:92-99.

Schell, M. A., Ulrich, R. L., Ribot, W. J., Brueggemann, E. E., Hines, H. B., Chen, D., Lipscomb, L., Kim, H. S., Mrázek, J., Nierman, W. C. and Deshazer, D. 2007. Type VI secretion is a major virulence determinant in *Burkholderia mallei*. *Mol. Microbiol.* 64:1466-1485.

Seo, Y. S., Lim, J., Choi, B. S., Kim, H., Goo, E., Lee, B., Lim, J. S., Choi, I. Y., Moon, J. S., Kim, J. and Hwang, I. 2011. Complete genome sequence of *Burkholderia gladioli* BSR3. *J. Bacteriol.* 193:3149.

Seo, Y. S., Lim, J. Y., Park, J., Kim, S., Lee, H. H., Cheong, H., Kim, S. M., Moon, J. S. and Hwang, I. 2015. Comparative genome analysis of rice-pathogenic *Burkholderia* provides insight into capacity to adapt to different environments and hosts. *BMC Genomics* 16:349.

Shalom, G., Shaw, J. G. and Thomas, M. S. 2007. *In vivo* expression technology identifies a type VI secretion system locus in *Burkholderia pseudomallei* that is induced upon invasion of macrophages. *Microbiology* 153:2689-2699.
Stevens, T. J. and Arkin, I. T. 2000. Do more complex organisms have a greater proportion of membrane proteins in their genomes? *Proteins* 39:417-420.

Tay, D. M., Govindarajan, K. R., Khan, A. M., Ong, T. Y., Samad, H. M., Soh, W. W., Tong, M., Zhang, F. and Tan, T. W. 2010. T3SEdb: data warehousing of virulence effectors secreted by the bacterial Type III Secretion System. *BMC Bioinformatics* 11 Suppl 7:S4.

Tseng, T. T., Tyler, B. M. and Setubal, J. C. 2009. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol*. 9 Suppl 1:S2.

Ura, H., Furuya, N., Iiyama, K., Hidaka, M., Tsuchiya, K. and Matsuyama, N. 2006. *Burkholderia gladioli* associated with symptoms of bacterial grain rot and leaf-sheath browning of rice plants. *J. Gen. Plant Pathol*. 72:98-103.

van Baarlen, P., van Belkum, A., Summerbell, R. C., Crous, P. W. and Thomma, B. P. 2007. Molecular mechanisms of pathogenicity: how do pathogenic microorganisms develop cross-kingdom host jumps? *FEMS Microbiol. Rev*. 31:239-277.

Weilharter, A., Mitter, B., Shin, M. V., Chain, P. S., Nowak, J. and Sessitsch, A. 2011. Complete genome sequence of the plant growth-promoting endophyte *Burkholderia phytofirmans* strain PsJN. *J. Bacteriol*. 193:3383-3384.

Weisskopf, L., Heller, S. and Eberl, L. 2011. *Burkholderia* species are major inhabitants of white lupin cluster roots. *Appl. Environ. Microbiol*. 77:7715-7720.

Wiersinga, W. J., van der Poll, T., White, N. J., Day, N. P. and Peacock, S. J. 2006. Melioidosis: insights into the pathogenicity of *Burkholderia pseudomallei*. *Nat. Rev. Microbiol*. 4:272-282.

Xia, Y., DeBolt, S., Dreyer, J., Scott, D. and Williams, M. A. 2015. Characterization of culturable bacterial endophytes and their capacity to promote plant growth from plants grown using organic or conventional practices. *Front. Plant Sci*. 6:490.

Xu, X. H., Su, Z. Z., Wang, C., Kubicek, C. P., Feng, X. X., Mao, L. J., Wang, J. Y., Chen, C., Lin, F. C. and Zhang, C. L. 2014. The rice endophyte *Harpophora oryzae* genome reveals evolution from a pathogen to a mutualistic endophyte. *Sci. Rep*. 4:5783.