

Research Article

Comparative Genome Sequence Analysis Reveals the Extent of Diversity and Conservation for Glycan-Associated Proteins in *Burkholderia* spp.

Hui San Ong, Rahmah Mohamed, and Mohd Firdaus-Raih

School of Biosciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia

Correspondence should be addressed to Mohd Firdaus-Raih, firdaus@mfrlab.org

Received 22 February 2012; Accepted 11 June 2012

Academic Editor: H. Heng

Copyright © 2012 Hui San Ong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Members of the *Burkholderia* family occupy diverse ecological niches. In pathogenic family members, glycan-associated proteins are often linked to functions that include virulence, protein conformation maintenance, surface recognition, cell adhesion, and immune system evasion. Comparative analysis of available *Burkholderia* genomes has revealed a core set of 178 glycan-associated proteins shared by all *Burkholderia* of which 68 are homologous to known essential genes. The genome sequence comparisons revealed insights into species-specific gene acquisitions through gene transfers, identified an S-layer protein, and proposed that significantly reactive surface proteins are associated to sugar moieties as a potential means to circumvent host defense mechanisms. The comparative analysis using a curated database of search queries enabled us to gain insights into the extent of conservation and diversity, as well as the possible virulence-associated roles of glycan-associated proteins in members of the *Burkholderia* spp. The curated list of glycan-associated proteins used can also be directed to screen other genomes for glycan-associated homologs.

1. Introduction

Members of the genus *Burkholderia*, with over 30 known species, have a unique ability to occupy diverse ecological niches, ranging from soil to the human respiratory tract [1]. Several strains are known to enhance disease resistance in plants [2] and improve nitrogen fixation [3, 4]. *B. pseudomallei*, *B. mallei* and *B. cenocepacia* are known to be involved in lung infections and are well recognized as pathogens of humans and animals. *B. pseudomallei*, a soil dwelling member of the *Burkholderia* genus is the causative agent for melioidosis and is capable of existing as a latent infection for decades with the longest period reported being 62 years [5]. The current use of antimicrobial therapy to treat melioidosis patients often fails due to intrinsic resistance of these bacteria. Prior to the discovery of an RNA helicase inhibitor toxin [6], the pathogenicity and virulence factors associated to *B. pseudomallei* have been elusive and many reports were inconclusive. Despite this significant progress, much remains to be discovered regarding the virulence of

B. pseudomallei and the *Burkholderia* pathogens in general, including potential roles played by glycan-associated proteins in pathogenesis.

The interactions and association between proteins and carbohydrates play important roles in numerous cellular functions, including processes which are associated with the pathogenicity of many bacterial species. Proteins can interact with carbohydrate molecules covalently (glycoproteins) and non-covalently (protein-carbohydrate complexes). In bacteria, known roles of glycoproteins include recognition of immunodominant proteins during infection [7–9], animal/plant-microbe symbiosis [10], cell-cell interactions [11, 12] and evasion of the immune system [13]. Bacteria and viruses can efficiently adhere to the surface membranes of host cells through carbohydrate binding which then allows host invasion [14, 15]. This process is enhanced by the ability of carbohydrates to generate an array of structurally diverse moieties from a relatively small number of monosaccharide units [16]. Carbohydrates and carbohydrate derivatives are not only associated with the bacterial cellular architecture

(e.g., capsule, lipopolysaccharide) but also function as information carriers [17], potential energy storage, and structural components. The diverse roles of carbohydrates result in the presence of carbohydrate molecules in various pathways responsible for systemic biological function of these bacteria.

In recent years, several genomes of *Burkholderia* species have been successfully sequenced. The completely sequenced *Burkholderia* genomes analyzed in this study are *B. pseudomallei* K96243 [18], *B. mallei* ATCC 23344 [19], *B. thailandensis* E264 [20], *B. xenovorans* LB400 [21], *B. phymatum* STM815 [22], *B. multivorans* ATCC 17616 [23], *B. ambifaria* MC40-6 [24], *B. vietnamiensis* G4 [25], and *B. cenocepacia* J2315 [26]. A number of experimental approaches have been carried out to identify glycan binding proteins present within a bacterial genome. However, the relatively smaller number of glycan-associated proteins identified through experimental methods barely reflects the actual amount of glycan-associated proteins present [27–29].

Here we report the inventorization and comparative analysis of glycan-associated proteins identified from the genomes of *Burkholderia* spp. This list comprises a wide array of biological functions ranging from involvement in cell envelope biogenesis, outer membrane pathway to the involvement in the transcription and translational machinery. In this work, we have compared and contrasted putative glycan-associated proteins by identifying both the orthologs as well as shared and unique glycan-associated proteins present in pathogenic and nonpathogenic *Burkholderia* spp. The selective pressure acting on glycan-associated surface proteins that showed reactivity during infection was also investigated.

2. Materials and Methods

2.1. Collection of Glycan-Associated Proteins in Prokaryotes. An initial target list of known prokaryotic glycan-associated proteins was compiled from the literature and keyword searches of the Uniprot Knowledgebase (Uniprot) [30] and Protein Data Bank (PDB) [31]. For this work, we define glycan-associated proteins as either glycoproteins or other proteins that are known to interact with glycans. The data collected from the literature reviews encompass the protein-carbohydrate complexes obtained through experimental techniques [27, 28] and proteins that have been reported to bind carbohydrates [32–34]. The searches in Uniprot and PDB were executed for the following keywords: glycoprotein, carbohydrate, sialic acid, glycan, bacteria glycoprotein, S-layer, polysaccharide, glucose, galactose, fructose, ribose, cellulose, heptose, trehalose, glucan, xylulose, rhamnose, capsule, mannose, maltose, and arabinose. The target list retrieved from the keyword searches was intensively curated and filtered manually for the proteins that have corresponding reports to be glycoproteins or associated to glycans either through the literature or as associated annotations in the databases searched such as a PDB protein structure with glycan bound.

2.2. Collection of Glycan-Associated Proteins in *Burkholderia* Species. This initial target list was used as queries

for BLASTP [35] searches to extract *Burkholderia* spp. homologs. BLASTP searches were run using the default parameters set by the BLASTP program, with the exception of the *E*-value limited to 0.00001 [36]. The output of the BLASTP programs was divided into two categories: (i) sequences with significant alignment (E -value ≤ 0.00001) and (ii) sequences with *E*-values higher than 0.00001 ($0.00001 < E$ -value ≤ 1) [37, 38]. The biological function of proteins that fall into the second category ($0.00001 < E$ -value ≤ 1) was inspected using pathways presented in the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database [39] and Pfam protein families database [40] searches to identify functional families and domains. If their biological properties were related to glycan, sugar, or carbohydrate binding, they were then regarded as glycan-associated proteins.

2.3. Identification of Orthologous Proteins in *Burkholderia* spp. The protein sequences for nine available *Burkholderia* genomes were downloaded from the National Center for Biotechnology Information (NCBI) website (<http://ftp.ncbi.nih.gov/genomes/Bacteria/>). Ortholog searches for these *Burkholderia* sequences were carried out using Inparanoid 4.1 package [41] with default parameters (overlap threshold = 0.5; confidence = 0.05; Matrix = BLOSUM45). Multiparanoid [42], an extension of the Inparanoid algorithms, was then used to cluster the orthologous groups from the nine *Burkholderia* species into a multispecies ortholog group. These outputs were then compiled into a dataset referred to as the *Burkholderia*_ortholog dataset. The BLASTP searches were carried out first followed by Inparanoid searches because not all complete genome sequence in prokaryote initial target list were available and the input for Inparanoid required the complete set of protein-encoding sequences.

2.4. Comparative Genome Analysis. Comparative analysis between *Burkholderia* spp. genomes was carried out using the Artemis Comparison Tool (ACT) [45] to detect conserved regions in different *Burkholderia* species. Comparison files between each *Burkholderia* species pairs were generated using BLASTN as an input for ACT.

2.5. Characterization of Glycan-Associated Proteins in *B. pseudomallei*. *B. pseudomallei* K96243 was chosen as the reference genome for further analysis due to its importance as a pathogen and the availability of the literature. Potential essential genes present in *B. pseudomallei* K96243 were identified via BLASTP searches [35], to E -value = 0.0001, other parameters at default settings, against the Database of Essential Genes [46]. Proteins with percentage identity above 40% [47, 48] were classified as potential essential genes for *B. pseudomallei*. The glycan-associated proteins identified were also subjected to PSORTb [49] to predict the subcellular location.

2.6. Sequence Analysis of Glycosylation Sites in Surface Proteins That Exhibit Significant Reactivity during Infection. Surface

proteins identified by PSORTb [49] were subjected to glycosylation site prediction using EnsembleGly [50]. The physicochemical properties of these surface proteins were then subjected to a number of analyses, namely, (i) ProtParam [51] to calculate the protein isoelectric point; (ii) ProtScale (<http://web.expasy.org/cgi-bin/protscale/protscale.pl>) to generate hydrophobicity graph using Kyte and Doolittle hydrophobic plot [52]; (iii) NetSurfP to calculate protein surface accessibility [53]; (iv) GOR IV [54] for the prediction of secondary structures; (v) SignalP version 3.0 [55] signal peptide prediction.

2.7. Measuring Selective Pressure on Surface Proteins. To investigate the selective pressure on these proteins, K_a/K_s ratio (nonsynonymous substitution rate/synonymous substitution rate) between *B. pseudomallei* K96243 proteins with its orthologs was calculated using K_a/K_s calculator v1.2 [56] by using the YN model [57]. A K_a/K_s ratio significantly greater than one implies positive selection, less than one implies purifying selection, and a ratio of one indicates neutrality [56].

3. Results and Discussion

3.1. Compilation of Glycan-Associated Proteins in Prokaryotes. Keyword searches against the databases and the literature that were manually curated have successfully identified 1470 glycan-associated proteins across the kingdom Prokaryota. Of these 1470 glycan-associated proteins, 545 were previously reported in the literature while the remaining 925 glycan-associated proteins were identified through data mining of the Uniprot Knowledgebase [30] and the Protein Data Bank [31] as described in Section 2. The use of this curated collection of glycan-associated proteins to probe *Burkholderia* genomes has enabled the identification of conserved glycan-associated proteins as well as species-specific or unique protein sequences.

3.2. Compilation of Glycan-Associated Proteins Present in *Burkholderia* spp. The total number of glycan-associated protein content for each *Burkholderia* species analyzed is presented in Figure 1. From our analysis, the percentage of glycan-associated proteins measured against the number of protein-coding genes is not directly proportional to the total number of protein-coding genes or genome size. However, the reduced number of glycan-associated proteins and the absence of species unique glycan-associated proteins for the *B. mallei* ATCC 23344 proteome are perhaps a reflection of it having the smallest genome size (5.84 Mb) amongst the available genomes. Further investigation into the regions of the missing sequences relative to the *B. pseudomallei* glycan-associated gene clusters showed the presence of transposable genetic elements (Table 1). Transposable genetic elements are known to contribute to bacterial genome variability by causing transposition events [58]. This has formed the basis that *B. mallei* underwent high rates of gene deletion after divergence from a *B. pseudomallei* ancestor as part of the mechanism to remove transposase encoding genes thus

resulting in the inadvertent deletion of the missing glycan-associated gene clusters.

3.3. Conserved Glycan-Associated Proteins in *Burkholderia* spp. The compilation of orthologs for glycan-associated proteins from the nine *Burkholderia* genomes, which was named the *Burkholderia_ortholog* dataset, yielded 178 glycan-associated proteins conserved across *Burkholderia* species, of which 68 are potential essential genes (Supplementary Table 1). The majority of these proteins have been annotated as proteins that are involved in the core regulation machinery such as translation and transcription (*bpss3220*, *bpsl3215*), nucleotide excision repair (*bpss0058*), proteolysis (*bpss1760*), and protein folding (*bpsl2827*).

Our analysis revealed no clear delineation of specific glycoproteins or glycan-associated proteins to only pathogenic members of *Burkholderia* spp. However, we noted the presence of variations in the form of different isomers and species-specific proteins. For example, two different galactosidase isomers were found in pathogenic *Burkholderia*, α -galactosidase (*B. pseudomallei*), and β -galactosidase (*B. ambifaria* MC40-6 and *B. cenocepacia* J2315). These different isomers affect sugar moieties found in the immunogenic capsular polysaccharide.

3.4. Unique Glycan-Associated Proteins in *B. pseudomallei* and *B. mallei*. Our comparative analysis had identified three unique proteins that are specific for *B. pseudomallei* (Table 2) and another six that are present only in *B. pseudomallei* and *B. mallei* (Table 3). Since these unique proteins were identified in well-characterized pathogenic members of the *Burkholderia* spp., two of these proteins were investigated further. These two proteins were chosen mainly because of their functions: (i) an α -galactosidase enzyme (BPSS2081) which is associated to the synthesis of sugar found in the immunogenic polysaccharide [59] and (ii) a putative surface protein (BPSS0796) that consists of a hemagglutinin motif which has been reported as important for bacterial adhesins [60].

3.5. *B. pseudomallei* Putative α -Galactose (*Bpss2081*). The gene *bpss2081* encodes for an α -galactosidase that is needed for the synthesis of the α -galactose found in the immunogenic capsular polysaccharide of *B. pseudomallei* [59]. This sugar is located on the cell surface and may likely participate in host-pathogen interactions.

Comparative analysis revealed a significantly lower G+C content (53.3%) in the flanking gene of *B. pseudomallei* K96243 *bpss2081* compared to the G+C content (68.06%) of its own chromosomes. A previous study by Sim et al. [61] concluded that the Bp-like capsular polysaccharide (Bp-like CPS) was a foreign element recently acquired through horizontal gene transfer when they found that Bp-like CPS exhibited a lower G+C content (59.2%) compared to the *B. thailandensis* gene cluster. Using Sim et al. [61] as a reference, we conducted a comparison for the gene position relative to *bpss2081* in the remaining *Burkholderia* species and discovered that *B. ambifaria* MC40-6 and *B. cenocepacia* J2315

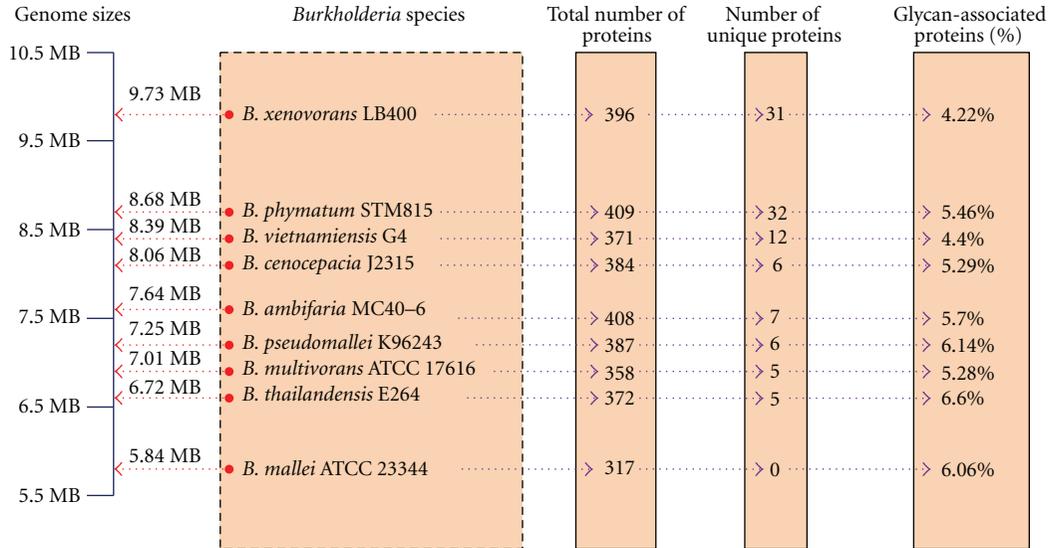


FIGURE 1: The total number of glycan-associated protein content found in nine *Burkholderia* species. The “total number of proteins” refers to total number of glycan-associated protein, and the “number of unique proteins” represent as the unique glycan-associated proteins that are found only in a specific species but does not have orthologs present in other *Burkholderia* spp. The percentages (%) of glycan-associated proteins were obtained by calculating the total number of glycan-associated proteins per the total number of protein coding genes.

TABLE 1: The transposase found in position relatives to glycan associated gene cluster.

Gene clusters in <i>B. pseudomallei</i> K96243	Relative position in <i>B. mallei</i> ATCC 23344
<i>bpss1825-bpss1833</i> (exopolysaccharide biosynthesis-related tyrosine-protein kinase)	Transposase OrfA (<i>bmaa0275</i>), Transposase OrfB (<i>bmaa0276</i>)
<i>bpsl2769-bpsl2790</i> (capsular polysaccharide-related gene cluster)	Transposase OrfA (<i>bma2284</i>), Transposase OrfB (<i>bma2283</i>)
<i>bpsl1830-bpsl1834</i> (ribose synthesis)	Transposase OrfA (<i>bma1204</i>), Transposase OrfB (<i>bma1203</i>)

have another form of isomer, namely, β -galactosidase gene (Figure 2). The genes flanking *bpss2081* for *B. pseudomallei* K96243 was found to have a relatively lower G+C content (54.94%) compared to the *B. ambifaria* MC40-6 gene cluster (69.11%) and the *B. cenocepacia* J2315 gene cluster (65.58%) (Supplementary Table 2). This result is therefore consistent with the possibility that the α -galactosidase gene cluster in *B. pseudomallei* is a foreign element recently acquired through horizontal gene transfer. The magnitude of horizontally acquired genetic information is unexpectedly high in many bacterial pathogens [62, 63] and is by no means unique to *B. pseudomallei*.

Since most gene transfers take place between closely related organisms [64], we extended the investigation of *bpss2081* to the closest homologs outside the *Burkholderia* genus. A BLASTP search revealed that the closest available homologs were from *Leptothrix cholodnii* (GenBank: YP_001792269.1) and *Curvibacter* putative symbiont of *Hydra magnipapillata* (GenBank: CBA30649.1) and belong to the order Burkholderiales. These three species share a common environmental niche; *B. pseudomallei* has been found in water supplies [65–67] while *L. cholodnii* resides

in lowly running and metal-rich aquatic environments [68] and *H. magnipapillata* is a fresh water polyp. Due to similar or shared habitats, it is possible *bpss2081* may have been acquired by *B. pseudomallei* from other bacteria within the same environmental niche.

3.6. *B. pseudomallei* and *B. mallei* Putative Hemagglutinins. BPSS0796 is a surface protein with a hemagglutinin motif that was identified using the Pfam database [40]. Hemagglutinins are known glycoproteins in bacteria [60, 69, 70] and influenza viruses [71, 72]. The protein coded for by the structural gene *bpss0796* is among the four proteins unique to *B. pseudomallei* and *B. mallei*. Analysis of the surrounding genes indicates that the gene cluster *bpss0796-bpss0799* in *B. pseudomallei* K96243 and *bmaa0649-bmaa0653* in *B. mallei* ATCC 23344 do not have orthologs present in *B. thailandensis* (Figure 3). This absence provided further evidence that the corresponding region in *B. thailandensis* E264 had been deleted in comparison to the two pathogenic *Burkholderia* species, *B. pseudomallei* and *B. mallei*. Due to the absence of this gene cluster in *B. thailandensis*, the deletion event is believed to have only happened after the divergence of

TABLE 2: Unique glycan associated proteins found in *B. pseudomallei* K96243.

GeneDB ID	Protein names	Subcellular location
BPSS2081	Putative α -galactosidase	Cytoplasmic
BPSSL1705	Putative membrane protein	Outer membrane/extracellular
BPSS1215	Sugar transporter	Cytoplasmic membrane

TABLE 3: Unique glycan associated proteins found in *B.pseudomallei* K96243 and *B.mallei* ATCC 23344.

<i>B. pseudomallei</i> K96253	Protein names	Subcellular location	Orthologs in <i>B. mallei</i> ATCC 23344
BPSS0254	Putative ribose 5-phosphate isomerase	Cytoplasmic	YP_106366.1
BPSSL2793	Putative D-glycero-d-manno-heptose 1,7-bisphosphate phosphatase	Cytoplasmic	YP_103855.1
BPSSL2795	Putative sedoheptulose 7-phosphate	Cytoplasmic	YP_103857.1
BPSS0796	Putative surface-exposed protein	Outer membrane/extracellular	YP_105401.1
BPSS1727	Putative hemagglutinin related protein	Outer membrane	YP_106315.1
BPSS1439	Putative membrane-anchored cell surface protein	Outer membrane	YP_105520.1



FIGURE 2: Structural comparison of galactosidase genes present in *B. pseudomallei* K96243, *B. ambifaria* MC40-6, and *B. cenocepacia* J2315. The Artemis Comparison tool (ACT) [45] was used to generate the data for this figure.

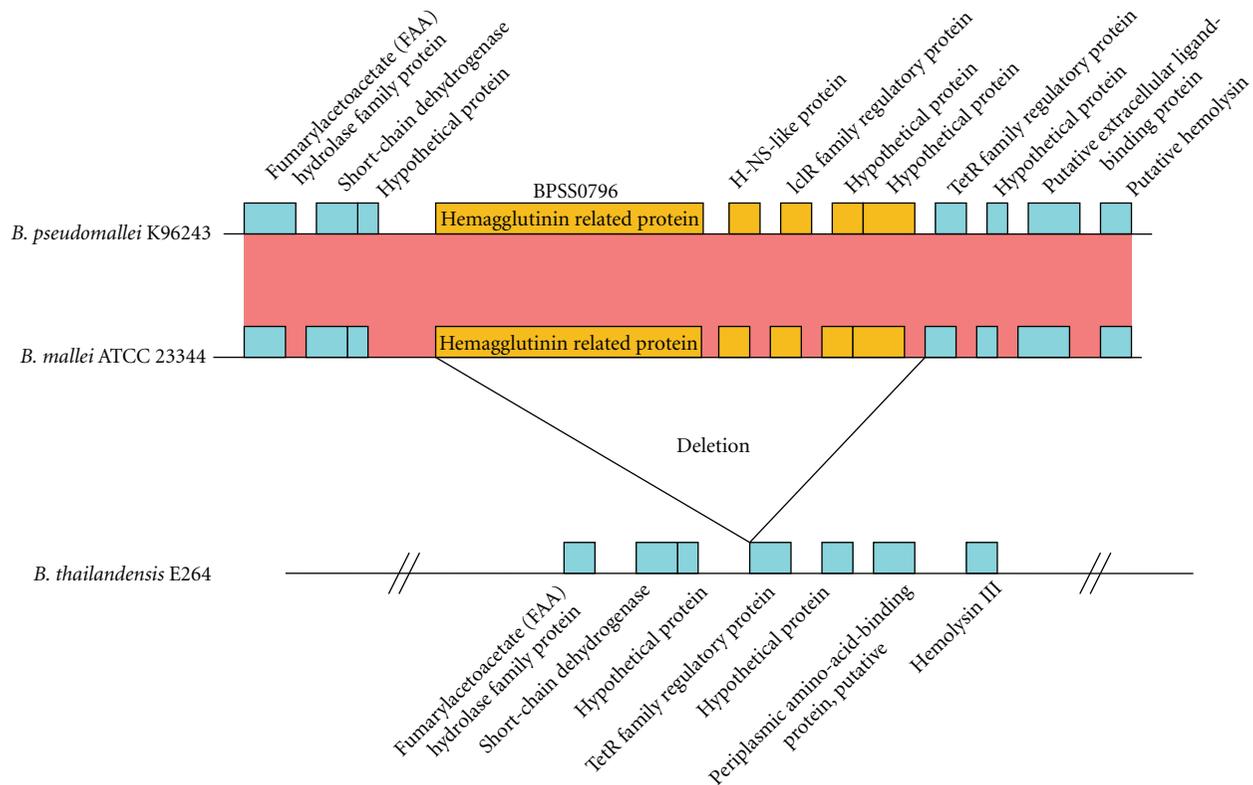


FIGURE 3: Structural comparison of surface-associated gene (*bps0796*) in *B. pseudomallei* K96243, *B. mallei* ATCC 23344, and *B. thailandensis* E264. The genes (orange color boxes) indicate the deleted genes.

B. thailandensis from *B. pseudomallei* and *B. mallei* approximately 47 million years ago [73]. The structural comparison is not shown for the orthologs of *bps0796* that are also absent in the other remaining *Burkholderia* species.

3.7. Significantly Reactive Glycan-Associated Proteins. We were also able to identify 5 serodiagnostic and cross-reactive antigens by cross referencing our data with the studies conducted by Felgner et al. [43] and Tiyawisuttri et al. [44] (Table 4). In the work conducted by Felgner et al. [43], BPSL1705 was found to be a serodiagnostic antigen that exhibited significantly different reactivity levels between the Singapore melioidosis positive and negative cohorts, while BPSL1902, BPSS2053, and BPSS1434 are cross-reactive antigens that are equally reactive in both melioidosis positive and negative cohorts. In another study conducted by Tiyawisuttri et al., the ortholog of BPSS1439 in *B. mallei* exhibited a strong antibody response during experimental glanders. Further analysis revealed one of these five to be an S-layer protein (BPSL1902).

3.8. A *B. pseudomallei* S-Layer Protein (BPSL1902). There has yet to be an experimentally verified S-layer protein identified in the order Burkholderiales. This task is further

complicated by the fact that there are no available overall sequence alignment profiles of the experimentally verified S-layer genes because sequence similarity only exists between S-layer protein genes of related species [74]. We first assigned BPSL1902, which was originally annotated as a putative membrane protein, to be an S-layer protein after analyzing *B. pseudomallei* protein sequences using the PSORTb subcellular location prediction program [49].

BPSL1902 was then subjected to further analyses that revealed similar sequence features to experimentally verified S-layer proteins in other prokaryotes. Most S-layer proteins are in the weakly acidic pH range [75] and this is consistent with BPSL1902 which has a ProtParam [51] predicted isoelectric point of 4.58. The hydrophobicity plot of BPSL1902 showed a similar hydrophobicity profile to known S-layer proteins (Figure 4) where the transmembrane α -helices segments are arranged at the highly hydrophobic N-terminal. It is known to be common for S-layer proteins to have α -helices in the N-terminal region [75] while there were no other consistencies found in the form of secondary structure to other S-layer proteins in the remaining part of the sequences. As with other S-layers, the highly hydrophobic N-terminal α -helices most likely serve as an anchor to the membrane while the remaining hydrophilic regions are likely to protrude outwards from the membrane-attached section.

TABLE 4: Surface glycan associated proteins that showed significant reactivity during infection.

GeneDB ID	Protein names	Subcellular location	Pathogenic description	Reference
BPSL1902	Putative membrane protein	S-Layer	Cross-reactive antigen	Felgner et al. [43]
BPSL1705	Putative membrane protein	Outer membrane/extracellular	Serodiagnostic antigen	Felgner et al. [43]
BPSS2053	Putative cell surface protein	Outer membrane	Cross-reactive antigen	Felgner et al. [43]
BPSS1434	Putative membrane-anchored cell surface protein	Outer membrane/extracellular	Cross-reactive antigen	Felgner et al. [43]
BPSS1439	Putative membrane-anchored cell surface protein	Outer membrane	Strong antibody response ^U	Tiyawitsutsri et al. [44]

Serodiagnostic antigens: antigens that exhibit significantly different reactivity levels between the Singapore melioidosis positive and negative cohorts.

Cross-reactive antigens: antigens that are equally reactive in both melioidosis positive and negative individuals, healthy subjects from endemic or nonendemic areas, or patients with other infections.

^UStrong antibody response: the experiment was carried out using the protein in *B. mallei* which is ortholog for BPSL1439. This protein in *B. mallei* exhibits strong antibody response during experimental glanders.

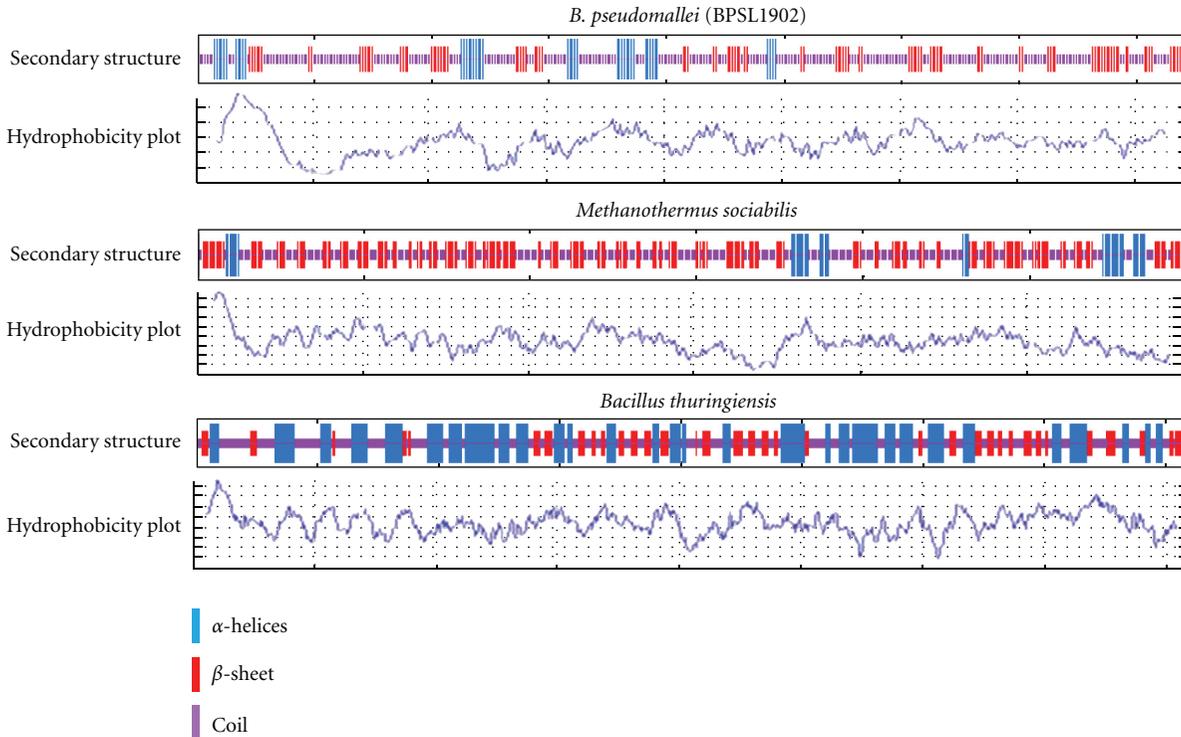


FIGURE 4: Composite diagram of comparisons for hydrophobicity plots and secondary structure predictions of *B. pseudomallei* S-layer protein (BPSL1902) versus *Bacillus thuringiensis ctc* (GeneBank ID: AJ012290) and *Methanothermobacter sociabilis slgA* (GeneBank ID: X58296) S-layer proteins. The program GOR IV [54] was used to generate the secondary structure while ProtScale was used to generate the Kyte and Doolittle plots [52].

The BPSL1902 S-layer was also subjected to EnsembleGly [50] glycosylation sites prediction and was found to potentially be highly O-glycosylated at the serine and threonine residues.

3.9. Features of the Glycosylated Regions. The significantly reactive *B. pseudomallei* surface proteins and their corresponding orthologs (Table 5) were subjected to multiple sequence alignment and EnsembleGly [50] glycosylation sites predictions. The analysis showed that conserved N- and C-terminal regions of protein sequences have a relatively low tendency to be glycosylated compared to the highly variable nonconserved regions. In the nonconserved regions, the propensity for glycosylation is indeed much higher (Figure 5) whereby N- and O-glycosylated residues tend to scatter along protein sequences and displaying inconsistency in the distribution of glycosylated residues. Hence this indicates that the nonconserved region is crucial in generating the versatility of glycosylation patterns as propensity for glycosylation seems to be high in this region. This is mainly believed to be due to the fact that the host immune system evasion mechanisms of pathogens often target the extremely variable and nonconserved portion of proteins [76].

3.10. Occurrence of Glycosylated Serine and Threonine Tandem Repeats That Are Potential Interfaces for Pathogen-Host Cell Interactions. The glycosylation predictions carried out point

to the possibility that BPSS1434, BPSL1705, and BPSS1439 consist of heavily glycosylated serine and threonine tandem repeats (STRs). The amino acid residues in STR regions were shown to be surface exposed and accessible to solvent with a high degree of glycosylation frequency (Figure 6). In this region, high surface accessibility will enable the amino acid with the glycan attached to protrude outwards and towards the host cell. This posture is important to allow for bacteria-host cell interaction because the surface glycans will form the first line of interaction to the mammalian host [77] in order to initiate infection [14, 15].

The presence of large amounts of glycosylated STRs is believed to contribute to host interaction considering that occurrence of tandem repeats (TRs) had been reported to have significant effects on pathogen adherence [78–81] and susceptibility to human diseases [82, 83]. Besides STRs, the presence of hemagglutinin motifs in BPSS1434, BPSS2053, BPSS1439, and BPSL1705 and hemagglutinin motif has been described to contribute to cell adherence and binding purposes. However, we observed that the glycosylation frequency in hemagglutinin motif was lower compared to the glycosylation frequency in the STRs region.

3.11. Selective Pressure Acting on Surface Proteins That Showed Significant Reactivity during Infection. The selective pressure acting on these proteins and their respective orthologs were

TABLE 5: Summary of significantly reactive surface proteins feature in *Burkholderia* spp. and the orthologs.

<i>B. pseudomallei</i> GeneDB ID	Glycosylation types	Additional features	GenBank	Ortholog proteins	Species
BPSS1902	N-linked;	—	YP_102742.1		<i>B. mallei</i> ATCC 23344
	O-linked		YP_443066.1		<i>B. thailandensis</i> E264
			YP_558210.1		<i>B. xenovorans</i> LB400
			YP_001579923.1		<i>B. multivorans</i> ATCC 17616
			YP_001808141.1		<i>B. ambifaria</i> MC40-6
BPSS1434			YP_001119324.1		<i>B. vietnamiensis</i> G4
			YP_002230652.1		<i>B. cenocepacia</i> J2315
	N-linked;	Serine threonine tandem repeats	YP_439154.1		<i>B. thailandensis</i> E264
	O-linked		YP_560399.1		<i>B. xenovorans</i> LB400
			YP_001583453.1		<i>B. multivorans</i> ATCC 17616
BPSS2053			YP_001815920.1		<i>B. ambifaria</i> MC40-6
			YP_001118486.1		<i>B. vietnamiensis</i> G4
			YP_001858800.1		<i>B. phymatum</i> STM815
	N-linked;	—	YP_443237.1		<i>B. thailandensis</i> E264
	O-linked		YP_557393.1		<i>B. xenovorans</i> LB400
BPSS1439			YP_001812417.1		<i>B. ambifaria</i> MC40-6
			YP_001116438.1		<i>B. vietnamiensis</i> G4
			YP_001861848.1		<i>B. phymatum</i> STM815
BPSS1705	N-linked;	Serine threonine tandem repeats	YP_105520.1		<i>B. mallei</i> ATCC 23344
	O-linked		—		—

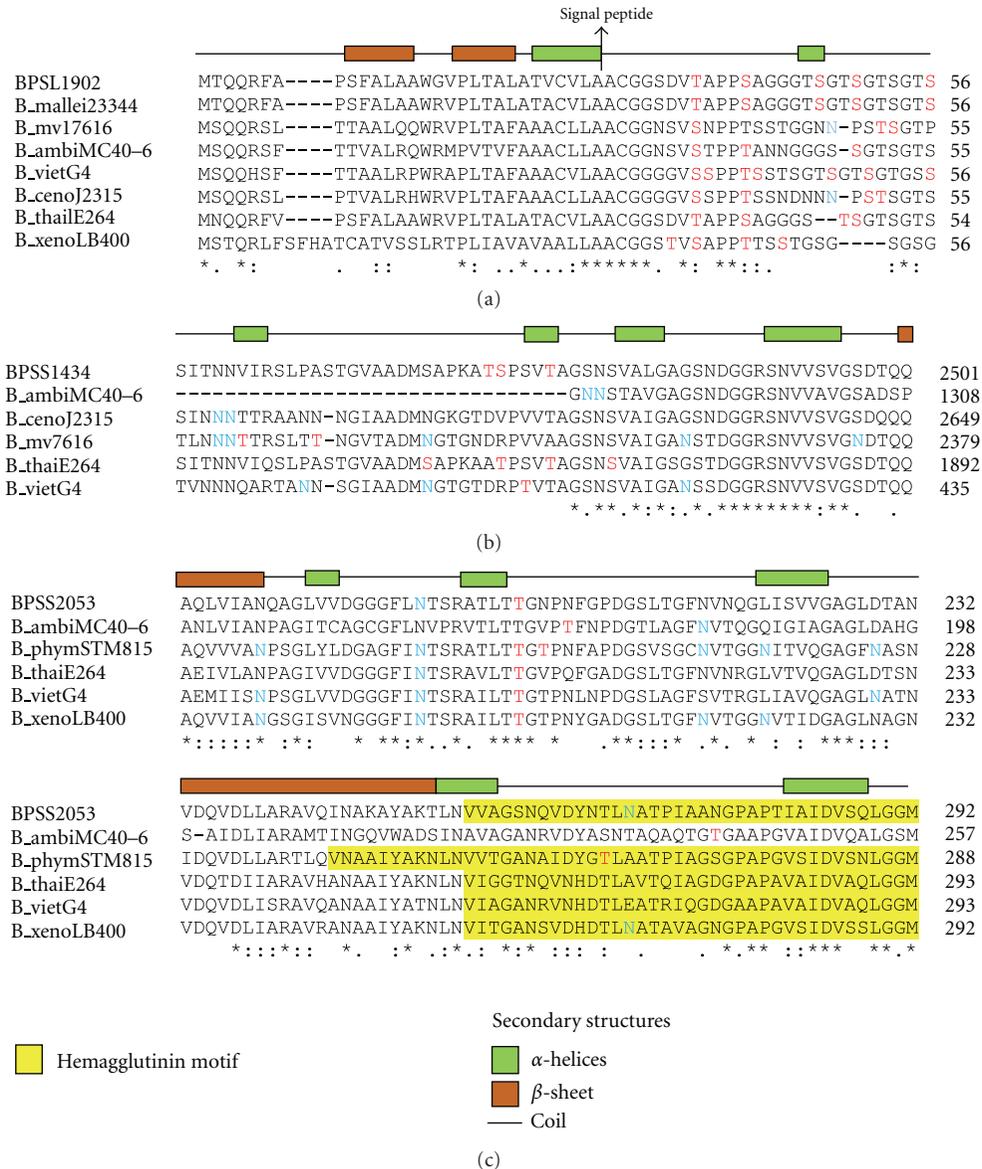


FIGURE 5: Multiple sequence alignment showed glycosylated amino acid in (a) BPSL1902, (b) BPSS1434, and (c) BPSS2053 with its respective orthologs. The amino acids in red represent O-glycosylation; blue color represents N-glycosylation and secondary structures shown are *B. pseudomallei* K96243 secondary structures. Note. BPSL1902, BPSS1434, and BPSS2053 are *B. pseudomallei* K96243 protein, B_mallei23344 is *B. mallei* ATCC 23344, B_mv17616 is *B. multivorans* ATCC 17616, B_ambifariaMC40-6 is *B. ambifaria* MC40-6, B_vietG4 is *B. vietnamiensis* G4, B_cenoJ2315 is *B. cenocepacia* J2315, B_thaiE264 is *B. thailandensis* E264, and B_xenoLB400 is *B. xenovorans* LB400.

then measured by calculating the number of nonsynonymous (amino acid changing) substitutions to synonymous (silent) substitutions rates, termed as K_a and K_s , respectively. We found that the rates of synonymous substitution (K_s) were much higher than nonsynonymous substitutions (K_a) (Table 6) indicating that resulting amino acid sequences carry only a small number of amino acid changes. These proteins are therefore considered as conservative towards functional and structural changes.

It is often that when host immune systems coevolve with the bacterial pathogens, the surface proteins will undergo

positive selection in an attempt to evade the host immune system and will have $K_a > K_s$ [84–86]. This diversifying selection acts to change the protein in ways that increase prevalence of advantageous traits which is the driving force behind adaptive evolution [87].

However, given the high synonymous substitution rate (K_s) of the proteins in Table 6, with their K_a/K_s ratios being lower than 1, the amino acid diversity is therefore quite restricted. Due to this selective pressure (high K_s) and the fact that these glycan-associated proteins are likely involved in interactions with the host due to their being located

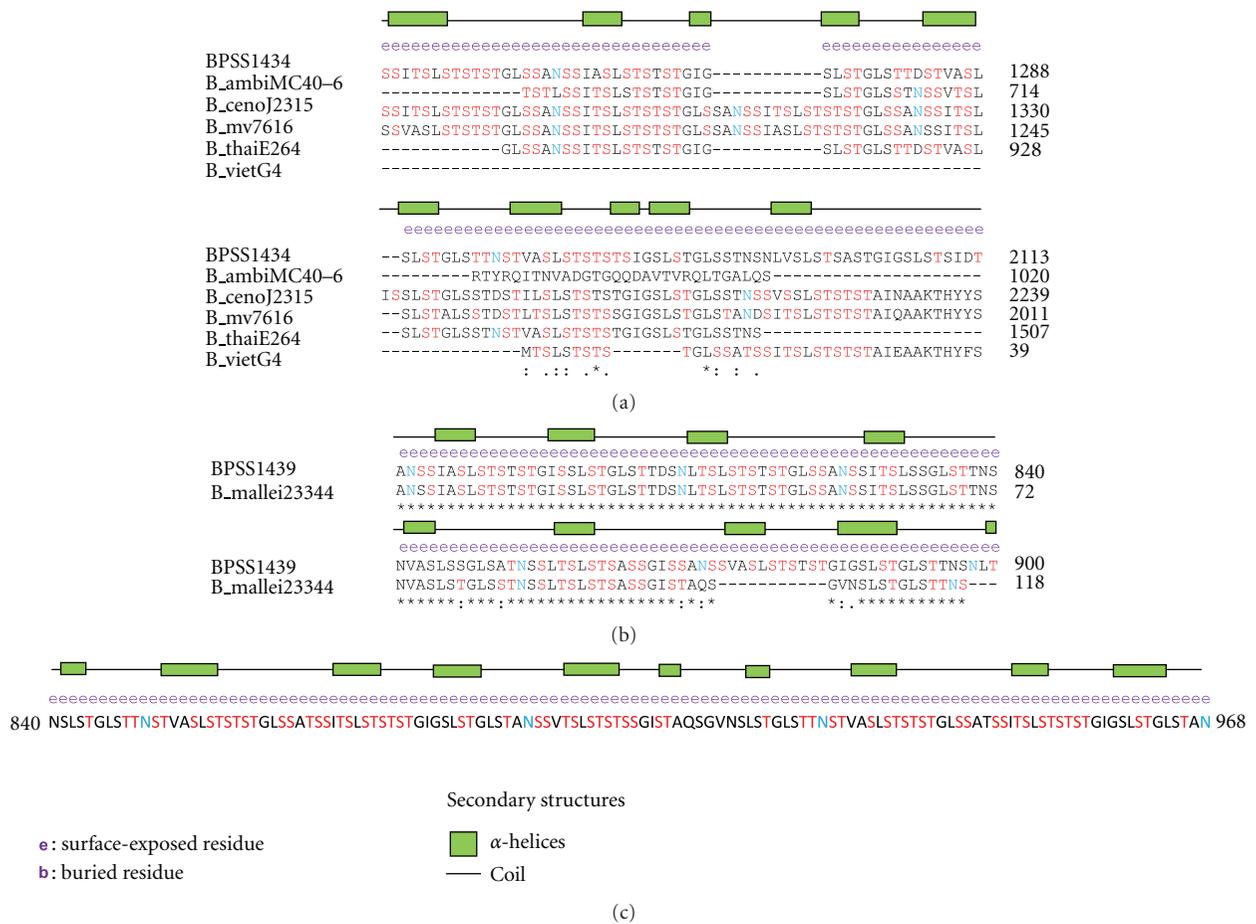


FIGURE 6: Glycosylated serine and threonine tandem repeats in (a) BPSS1434, (b) BPSS1439 and (c) BPSL1705 with its respective orthologs in *Burkholderia* spp. (if any). The amino acids in red represent O-glycosylation; blue color represents N-glycosylation. *Note.* BPSS1434, BPSL1705, and BPSS1439 are *B. pseudomallei* K96243 proteins, B.mv17616 is *B. multivorans* ATCC 17616, B.ambifariaMC40-6 is *B. ambifaria* MC40-6, B_vietG4 is *B. vietnamiensis* G4, B_thaiE264 is *B. thailandensis* E264, and B_xenoLB400 is *B. xenovorans* LB400.

on the cell surface, successful bacterial pathogens would most likely have developed other strategies to circumvent host defense mechanisms. Therefore, this scenario leads to the possibility that one of the mechanisms to increase the adaptation of virulence-associated proteins lies behind the structurally diverse sugar moieties that binds to these proteins (shown with the presence of glycosylation sites). Furthermore, the lack of a clear delineation of a specific glycoprotein or glycan-associated protein to only pathogenic members of the *Burkholderia* spp. implies that while glycan-associated proteins may play roles in virulence, they are not involved directly in the pathogenesis mechanisms of the known *Burkholderia* pathogens.

4. Conclusions

Our use of this dataset has enabled us to contrast the genomes for members of the *Burkholderia* genus resulting in the identification of *Burkholderia* conserved orthologs and several proteins unique to pathogenic *Burkholderia* species. Further analysis of the data was also able to provide

insights into the possible acquisition of species unique genes through gene transfer events. The selective measurement of the surface proteins signified that these proteins are quite conservative towards structural and functional changes unlike several other known surface proteins thereby implying a more prominent role played by the glycan components for diversification of the cell surface presentation to the host immune system.

Authors' Contribution

H. S. Ong compiled the datasets, analyzed the data, and drafted the paper. R. Mohamed participated in the design of the study. M. F. Raih conceived the overall study, designed the analysis, and participated in drafting the paper. All authors participated in revising the paper.

Acknowledgments

This work was supported by the Universiti Kebangsaan Malaysia Arus Perdana Grants UKM-AP-BPB-13-2009 and

TABLE 6: Nonsynonymous substitution (K_a) and synonymous substitution (K_s) rates between protein pairs and its orthologs as described.

Sequence 1	Sequence 2	Nonsynonymous substitution; K_a	Synonymous substitution; K_s	K_a/K_s
BPSS1439	<i>B.mallei</i> ATCC 23344 (Genbank: YP_105520.1)	0.0217786	0.160713	0.135513
	<i>B.mallei</i> ATCC 23344 (Genbank: YP_102742.1)	0.001901	0.014441	0.131671
	<i>B. thailandensis</i> E264 (Genbank: YP_443066.1)	0.022584	0.366201	0.061671
BPSL1902	<i>B. multivorans</i> ATCC 17616 (Genbank: YP_001579923.1)	0.178427	2.7173	0.065663
	<i>B. ambifaria</i> MC40-6 (Genbank: YP_001808141.1)	0.185665	1.8318	0.101357
	<i>B. vietnamiensis</i> G4 (Genbank: YP_001119324.1)	0.161194	1.62651	0.099104
	<i>B. cenocepacia</i> J2315 (Genbank: YP_002230652.1)	0.170878	2.93372	0.058246
BPSS2053	<i>B. xenovorans</i> LB400 (Genbank: YP_558210.1)	0.343855	3.80685	0.090325
	<i>B. thailandensis</i> E264 (Genbank: YP_443237.1)	0.366385	5.64031	0.064958
	<i>B. ambifaria</i> MC40-6 (Genbank: YP_001812417.1)	1.14359	3.06456	0.373167
	<i>B. vietnamiensis</i> G4 (Genbank: YP_001116438.1)	0.383681	4.61058	0.083218
	<i>B. xenovorans</i> LB400 (Genbank: YP_557393.1)	0.570889	5.6423	0.10118
BPSS1434	<i>B. phymatum</i> STM815 (Genbank: YP_001861848.1)	0.4192	2.5801	0.162475
	<i>B. thailandensis</i> E264 (Genbank: YP_439154.1)	0.091527	0.885211	0.103395
	<i>B. multivorans</i> ATCC 17616 (Genbank: YP_001583453.1)	0.360264	2.33222	0.154473
	<i>B. ambifaria</i> MC40-6 (Genbank: YP_001815920.1)	0.555408	5.0961	0.108987
	<i>B. vietnamiensis</i> G4 (Genbank: YP_001118486.1)	0.943879	5.11658	0.184474
	<i>B. xenovorans</i> LB400 (Genbank: YP_560399.1)	0.79562	5.64563	0.140927
	<i>B. phymatum</i> STM815 (Genbank: YP_001858800.1)	0.803449	5.62495	0.142837

UKM-AP-2011-1. H. S. Ong gratefully acknowledges funding from the Ministry of Science, Technology and Innovation, Malaysia, through the National Science Fellowship award.

References

- [1] T. Coenye and P. Vandamme, "Diversity and significance of *Burkholderia* species occupying diverse ecological niches," *Environmental Microbiology*, vol. 5, no. 9, pp. 719–729, 2003.
- [2] S. Compant, B. Duffy, J. Nowak, C. Clément, and E. A. Barka, "Use of plant growth-promoting bacteria for biocontrol of plant diseases: principles, mechanisms of action, and future prospects," *Applied and Environmental Microbiology*, vol. 71, no. 9, pp. 4951–4959, 2005.
- [3] A. Sessitsch, T. Coenye, A. V. Sturz et al., "*Burkholderia phytofirmans* sp. nov., a novel plant-associated bacterium with plant-beneficial properties," *International Journal of Systematic and Evolutionary Microbiology*, vol. 55, no. 3, pp. 1187–1192, 2005.
- [4] E. Ait Barka, J. Nowak, and C. Clément, "Enhancement of chilling resistance of inoculated grapevine plantlets with a plant growth-promoting rhizobacterium, *Burkholderia phytofirmans* strain PsJN," *Applied and Environmental Microbiology*, vol. 72, no. 11, pp. 7246–7252, 2006.
- [5] V. Ngauy, Y. Lemeshev, L. Sadkowski, and G. Crawford, "Cutaneous melioidosis in a man who was taken as a prisoner of war by the Japanese during World War II," *Journal of Clinical Microbiology*, vol. 43, no. 2, pp. 970–972, 2005.
- [6] A. Cruz-Migoni, G. M. Hautbergue, P. J. Artymiuk et al., "A *Burkholderia pseudomallei* toxin inhibits helicase activity of translation factor eIF4A," *Science*, vol. 334, no. 6057, pp. 821–824, 2011.
- [7] C. Steichen, P. Chen, J. F. Kearney, and C. L. Turnbough, "Identification of the immunodominant protein and other proteins of the *Bacillus anthracis* exosporium," *Journal of Bacteriology*, vol. 185, no. 6, pp. 1903–1910, 2003.

- [8] P. Doig, N. Kinsella, P. Guerry, and T. J. Trust, "Characterization of a post-translational modification of *Campylobacter* flagellin: identification of a sero-specific glycosyl moiety," *Molecular Microbiology*, vol. 19, no. 2, pp. 379–387, 1996.
- [9] J. W. McBride, X. J. Yu, and D. H. Walker, "Glycosylation of homologous immunodominant proteins of *Ehrlichia chaffeensis* and *Ehrlichia canis*," *Infection and Immunity*, vol. 68, no. 1, pp. 13–18, 2000.
- [10] I. Lerouge and J. Vanderleyden, "O-antigen structural variation: mechanisms and possible roles in animal/plant-microbe interactions," *FEMS Microbiology Reviews*, vol. 26, no. 1, pp. 17–47, 2002.
- [11] C. Ryder, M. Byrd, and D. J. Wozniak, "Role of polysaccharides in *Pseudomonas aeruginosa* biofilm development," *Current Opinion in Microbiology*, vol. 10, no. 6, pp. 644–648, 2007.
- [12] P. R. Crocker, "Siglecs: sialic-acid-binding immunoglobulin-like lectins in cell-cell interactions and signalling," *Current Opinion in Structural Biology*, vol. 12, no. 5, pp. 609–615, 2002.
- [13] A. Weintraub, "Immunology of bacterial polysaccharide antigens," *Carbohydrate Research*, vol. 338, no. 23, pp. 2539–2547, 2003.
- [14] J. Adam, M. Pokorná, C. Sabin, E. P. Mitchell, A. Imberty, and M. Wimmerová, "Engineering of PA-III lectin from *Pseudomonas aeruginosa*—unravelling the role of the specificity loop for sugar preference," *BMC Structural Biology*, vol. 7, article 36, 2007.
- [15] V. C. Chu and G. R. Whittaker, "Influenza virus entry and infection require host cell N-linked glycoprotein," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 52, pp. 18153–18158, 2004.
- [16] G. F. Audette, L. T. J. Delbaere, and J. Xiang, "Mapping protein: carbohydrate interactions," *Current Protein and Peptide Science*, vol. 4, no. 1, pp. 11–20, 2003.
- [17] C. Shionyu-Mitsuyama, T. Shirai, H. Ishida, and T. Yamane, "An empirical approach for structure-based prediction of carbohydrate-binding sites on proteins," *Protein Engineering*, vol. 16, no. 7, pp. 467–478, 2003.
- [18] M. T. G. Holden, R. W. Titball, S. J. Peacock et al., "Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 39, pp. 14240–14245, 2004.
- [19] W. C. Nierman, D. DeShazer, H. S. Kim et al., "Structural flexibility in the *Burkholderia mallei* genome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 39, pp. 14246–14251, 2004.
- [20] H. S. Kim, M. A. Schell, Y. Yu et al., "Bacterial genome adaptation to niches: divergence of the potential virulence genes in three *Burkholderia* species of different survival strategies," *BMC Genomics*, vol. 6, article 174, 2005.
- [21] J. Goris, P. De Vos, J. Caballero-Mellado et al., "Classification of the biphenyl- and polychlorinated biphenyl-degrading strain LB400T and relatives as *Burkholderia xenovorans* sp. nov.," *International Journal of Systematic and Evolutionary Microbiology*, vol. 54, no. 5, pp. 1677–1681, 2004.
- [22] A. Copeland, S. Lucas, A. Lapidus et al., "Complete sequence of chromosome 1 of *Burkholderia phymatum* STM815," submitted to (APR-2008) to the EMBL/GenBank/DBJ Database, <http://www.ncbi.nlm.nih.gov/nucleotide/NC.010622.1>.
- [23] Y. Ohtsubo, A. Yamashita, K. Kurokawa et al., "Complete genome sequence of *Burkholderia multivorans* ATCC 17616," submitted to (APR-2007) to the EMBL/GenBank/DBJ Databases, <http://www.ncbi.nlm.nih.gov/nucleotide/NC.010804.1>.
- [24] A. Copeland, S. Lucas, A. Lapidus et al., "Complete sequence of chromosome 1 of *Burkholderia ambifaria* MC40-6," submitted to (APR-2008) to the EMBL/GenBank/DBJ Databases, <http://www.ncbi.nlm.nih.gov/nucleotide/NC.010551.1>. In press.
- [25] A. Copeland, S. Lucas, A. Lapidus, Barry et al., "Complete sequence of chromosome 1 of *Burkholderia vietnamiensis* G4," submitted to (MAR-2007) to the EMBL/GenBank/DBJ Databases, <http://www.ncbi.nlm.nih.gov/nucleotide/NC.009256.11>.
- [26] M. T. G. Holden, H. M. B. Seth-Smith, L. C. Crossman et al., "The Genome of *Burkholderia cenocepacia* J2315, an epidemic pathogen of cystic fibrosis patients," *Journal of Bacteriology*, vol. 191, no. 8, p. 2907, 2009.
- [27] D. Kang, E. S. Ji, M. H. Moon, and J. S. Yoo, "Lectin-based enrichment method for glycoproteomics using hollow fiber flow field-flow fractionation: application to *Streptococcus pyogenes*," *Journal of Proteome Research*, vol. 9, no. 6, pp. 2855–2862, 2010.
- [28] M. González-Zamorano, G. M. Hernández, W. Xolalpa et al., "Mycobacterium tuberculosis glycoproteomics based on ConA-lectin affinity capture of mannoseylated proteins," *Journal of Proteome Research*, vol. 8, no. 2, pp. 721–733, 2009.
- [29] A. E. Scott, S. M. Twine, K. M. Fulton et al., "Flagellar glycosylation in *Burkholderia pseudomallei* and *Burkholderia thailandensis*," *Journal of Bacteriology*, vol. 193, no. 14, pp. 3577–3587, 2011.
- [30] B. Boeckmann, A. Bairoch, R. Apweiler et al., "The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003," *Nucleic Acids Research*, vol. 31, no. 1, pp. 365–370, 2003.
- [31] H. M. Berman, J. Westbrook, Z. Feng et al., "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.
- [32] A. Malik and S. Ahmad, "Sequence and structural features of carbohydrate binding in proteins and assessment of predictability using a neural network," *BMC Structural Biology*, vol. 7, article 1, 2007.
- [33] M. Schirm, E. C. Soo, A. J. Aubry, J. Austin, P. Thibault, and S. M. Logan, "Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*," *Molecular Microbiology*, vol. 48, no. 6, pp. 1579–1592, 2003.
- [34] M. F. Mescher, J. L. Strominger, and S. W. Watson, "Protein and carbohydrate composition of the cell envelope of *Halobacterium salinarum*," *Journal of Bacteriology*, vol. 120, no. 2, pp. 945–954, 1974.
- [35] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.
- [36] K. Belov, C. E. Sanderson, J. E. Deakin et al., "Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system," *Genome Research*, vol. 17, no. 7, pp. 982–991, 2007.
- [37] S. Gopal, M. Schroeder, U. Pieper et al., "Homology-based annotation yields 1,042 new candidate genes in the *Drosophila melanogaster* genome," *Nature Genetics*, vol. 27, no. 3, pp. 337–340, 2001.
- [38] S. F. Altschul, T. L. Madden, A. A. Schäffer et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997.
- [39] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.
- [40] A. Bateman, L. Coin, R. Durbin et al., "The Pfam protein families database," *Nucleic Acids Research*, vol. 32, pp. D138–D141, 2004.

- [41] M. R. Remm, C. E. V. Storm, and E. L. L. Sonnhammer, "Automatic clustering of orthologs and in-paralogs from pairwise species comparisons," *Journal of Molecular Biology*, vol. 314, no. 5, pp. 1041–1052, 2001.
- [42] A. Alexeyenko, I. Tamas, G. Liu, and E. L. L. Sonnhammer, "Automatic clustering of orthologs and inparalogs shared by multiple proteomes," *Bioinformatics*, vol. 22, no. 14, pp. e9–e15, 2006.
- [43] P. L. Felgner, M. A. Kayala, A. Vigil et al., "A *Burkholderia pseudomallei* protein microarray reveals serodiagnostic and cross-reactive antigens," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 32, pp. 13499–13504, 2009.
- [44] R. Tiyawisutrisi, M. T. G. Holden, S. Tumapa et al., "*Burkholderia* Hep_Hap autotransporter (BuHA) proteins elicit a strong antibody response during experimental glanders but not human melioidosis," *BMC Microbiology*, vol. 7, article 19, 2007.
- [45] T. J. Carver, K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell, and J. Parkhill, "ACT: the Artemis comparison tool," *Bioinformatics*, vol. 21, no. 16, pp. 3422–3423, 2005.
- [46] B. Petersen, T. N. Petersen, P. Andersen, M. Nielsen, and C. Lundegaard, "A generic method for assignment of reliability scores applied to solvent accessibility predictions," *BMC Structural Biology*, vol. 9, article 51, 2009.
- [47] B. Rost, "Twilight zone of protein sequence alignments," *Protein Engineering*, vol. 12, no. 2, pp. 85–94, 1999.
- [48] T. Schwede, "Protein structure modeling," in *Computational Structural Biology Methods and Applications*, T. Schwede and M. C. Peitsch, Eds., chapter 1, 2008.
- [49] N. Y. Yu, J. R. Wagner, M. R. Laird et al., "PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes," *Bioinformatics*, vol. 26, no. 13, pp. 1608–1615, 2010.
- [50] C. Caragea, J. Sinapov, A. Silvescu, D. Dobbs, and V. Honavar, "Glycosylation site prediction using ensembles of Support Vector Machine classifiers," *BMC Bioinformatics*, vol. 8, article 438, 2007.
- [51] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, and A. Bairoch, "ExPASy: the proteomics server for in-depth protein knowledge and analysis," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3784–3788, 2003.
- [52] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," *Journal of Molecular Biology*, vol. 157, no. 1, pp. 105–132, 1982.
- [53] B. Petersen, T. N. Petersen, P. Andersen, M. Nielsen, and C. Lundegaard, "A generic method for assignment of reliability scores applied to solvent accessibility predictions," *BMC Structural Biology*, vol. 9, article 51, 2009.
- [54] J. Garnier, J. F. Gibrat, and B. Robson, "GOR method for predicting protein secondary structure from amino acid sequence," *Methods in Enzymology*, vol. 266, pp. 540–553, 1996.
- [55] J. D. Bendtsen, H. Nielsen, G. von Heijne, and S. Brunak, "Improved prediction of signal peptides: signalP 3.0," *Journal of Molecular Biology*, vol. 340, no. 4, pp. 783–795, 2004.
- [56] Z. H. Zhang, J. Li, X. Q. Zhao, J. Wang, G. K. S. Wong, and J. Yu, "KaKs_calculator: calculating Ka and Ks through model selection and model averaging," *Genomics, Proteomics and Bioinformatics*, vol. 4, no. 4, pp. 259–263, 2006.
- [57] Z. Yang and R. Nielsen, "Estimating synonymous and non-synonymous substitution rates under realistic evolutionary models," *Molecular Biology and Evolution*, vol. 17, no. 1, pp. 32–43, 2000.
- [58] U. Dobrindt and J. Hacker, "Whole genome plasticity in pathogenic bacteria," *Current Opinion in Microbiology*, vol. 4, no. 5, pp. 550–557, 2001.
- [59] H. Masoud, M. Ho, T. Schollaardt, and M. B. Perry, "Characterization of the capsular polysaccharide of *Burkholderia (Pseudomonas) pseudomallei* 304b," *Journal of Bacteriology*, vol. 179, no. 18, pp. 5663–5669, 1997.
- [60] S. M. Prasad, Y. Yin, E. Rodzinski, E. I. Tuomanen, and H. R. Masure, "Identification of a carbohydrate recognition domain in filamentous hemagglutinin from *Bordetella pertussis*," *Infection and Immunity*, vol. 61, no. 7, pp. 2780–2785, 1993.
- [61] B. M. Q. Sim, N. Chantratita, W. F. Ooi et al., "Genomic acquisition of a capsular polysaccharide virulence cluster by non-pathogenic *Burkholderia* isolates," *Genome Biology*, vol. 11, no. 8, article R89, 2010.
- [62] M. Kuroda, T. Ohta, I. Uchiyama et al., "Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*," *The Lancet*, vol. 357, no. 9264, pp. 1225–1240, 2001.
- [63] H. Ochman, J. G. Lawrence, and E. A. Grolsman, "Lateral gene transfer and the nature of bacterial innovation," *Nature*, vol. 405, no. 6784, pp. 299–304, 2000.
- [64] J. G. Lawrence and H. Hendrickson, "Lateral gene transfer: when will adolescence end?" *Molecular Microbiology*, vol. 50, no. 3, pp. 739–749, 2003.
- [65] B. J. Currie, M. Mayo, N. M. Anstey, P. Donohoe, A. Haase, and D. J. Kemp, "A cluster of melioidosis cases from an endemic region is clonal and is linked to the water supply using molecular typing of *Burkholderia pseudomallei* isolates," *American Journal of Tropical Medicine and Hygiene*, vol. 65, no. 3, pp. 177–179, 2001.
- [66] T. J. J. Inglis, S. C. Garrow, C. Adams, M. Henderson, M. Mayo, and B. J. Currie, "Acute melioidosis outbreak in Western Australia," *Epidemiology and Infection*, vol. 123, no. 3, pp. 437–443, 1999.
- [67] T. J. J. Inglis, S. C. Garrow, M. Henderson et al., "*Burkholderia pseudomallei* traced to water treatment plant in Australia," *Emerging Infectious Diseases*, vol. 6, no. 1, pp. 56–59, 2000.
- [68] W. L. Van Veen, E. G. Mulder, and M. H. Deinema, "The Sphaerotilus-Leptothrix group of bacteria," *Microbiological Reviews*, vol. 42, no. 2, pp. 329–356, 1978.
- [69] D. D. Roberts, L. D. Olson, M. F. Barile, V. Ginsburg, and H. C. Krivan, "Sialic acid-dependent adhesion of *Mycoplasma pneumoniae* to purified glycoproteins," *Journal of Biological Chemistry*, vol. 264, no. 16, pp. 9289–9293, 1989.
- [70] A. L. Mattos-Guaraldi, L. C. Duarte Formiga, and G. A. Pereira, "Cell surface components and adhesion in corynebacterium diphtheriae," *Microbes and Infection*, vol. 2, no. 12, pp. 1507–1512, 2000.
- [71] N. K. Sauter, J. E. Hanson, G. D. Glick et al., "Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography," *Biochemistry*, vol. 31, no. 40, pp. 9609–9621, 1992.
- [72] J. E. Peters and R. A. DiCapua, "Immunochemical characterization of Lymantria dispar NPV hemagglutinin: protein-carbohydrate interaction," *Intervirology*, vol. 9, no. 4, pp. 231–242, 1978.
- [73] Y. Yu, H. S. Kim, H. C. Hui et al., "Genomic patterns of pathogen evolution revealed by comparison of *Burkholderia pseudomallei*, the causative agent of melioidosis, to avirulent *Burkholderia thailandensis*," *BMC Microbiology*, vol. 6, article 46, 2006.

- [74] H. J. Boot and P. H. Pouwels, "Expression, secretion and antigenic variation of bacterial S-layer proteins," *Molecular Microbiology*, vol. 21, no. 6, pp. 1117–1123, 1996.
- [75] M. Sára and U. B. Sleytr, "S-layer proteins," *Journal of Bacteriology*, vol. 182, no. 4, pp. 859–868, 2000.
- [76] J. Lannergard, M. C. U. Gustafsson, J. Waldemarsson, A. Norrby-Teglund, M. Stalhammar-Carlemalm, and G. Lindahl, "The hypervariable region of streptococcus pyogenes M protein escapes antibody attack by antigenic variation and weak immunogenicity," *Cell Host & Microbe*, vol. 10, no. 2, pp. 147–157, 2011.
- [77] L. E. Comstock and D. L. Kasper, "Bacterial glycans: key mediators of diverse host immune responses," *Cell*, vol. 126, no. 5, pp. 847–850, 2006.
- [78] L. Loza, Y. Fu, A. S. Ibrahim, D. C. Sheppard, S. G. Filler, and J. E. Edwards, "Functional analysis of the *Candida albicans* ALSI gene product," *Yeast*, vol. 21, no. 6, pp. 473–482, 2004.
- [79] A. J. Sheets and J. W. S. Geme, "Adhesive activity of the haemophilus cryptic genospecies cha autotransporter is modulated by variation in tandem peptide repeats," *Journal of Bacteriology*, vol. 193, no. 2, pp. 329–339, 2011.
- [80] M. Hussain, A. Haggar, G. Peters et al., "More than one tandem repeat domain of the extracellular adherence protein of *Staphylococcus aureus* is required for aggregation, adherence, and host cell invasion but not for leukocyte activation," *Infection and Immunity*, vol. 76, no. 12, pp. 5615–5623, 2008.
- [81] J. M. Rauceo, R. De Armond, H. Otoo et al., "Threonine-rich repeats increase fibronectin binding in the *Candida albicans* adhesin Als5p," *Eukaryotic Cell*, vol. 5, no. 10, pp. 1664–1673, 2006.
- [82] L. E. Vinall, A. S. Hill, P. Piggy et al., "Variable number tandem repeat polymorphism of the mucin genes located in the complex on 11p15.5," *Human Genetics*, vol. 102, no. 3, pp. 357–366, 1998.
- [83] K. Kyo, T. Muto, H. Nagawa, G. M. Lathrop, and Y. Nakamura, "Associations of distinct variants of the intestinal mucin gene MUC3A with ulcerative colitis and Crohn's disease," *Journal of Human Genetics*, vol. 46, no. 1, pp. 5–20, 2001.
- [84] S. Kakizawa, K. Oshima, H. Y. Jung et al., "Positive selection acting on a surface membrane protein of the plant-pathogenic phytoplasmas," *Journal of Bacteriology*, vol. 188, no. 9, pp. 3424–3428, 2006.
- [85] L. Petersen, J. P. Bollback, M. Dimmic, M. Hubisz, and R. Nielsen, "Genes under positive selection in *Escherichia coli*," *Genome Research*, vol. 17, no. 9, pp. 1336–1343, 2007.
- [86] F. M. Jiggins, G. D. D. Hurst, and Z. Yang, "Host-symbiont conflicts: positive selection on an outer membrane protein of parasitic but not mutualistic Rickettsiaceae," *Molecular Biology and Evolution*, vol. 19, no. 8, pp. 1341–1349, 2002.
- [87] C. Darwin and A. R. Wallace, "On the tendency of species for form varieties; and on the perpetuation of varieties and species by natural means of selection," *Proceedings of the Linnean Society of London*, vol. 3, pp. 45–62, 1858.