

Identification of Sequence Motifs for the Protein Components of Type IX Secretion System (Pengenalpastian Motif Jujukan bagi Komponen Protein Sistem Perembesan Jenis IX)

REEKI EMRIZAL & NOR AZLAN NOR MUHAMMAD*

ABSTRACT

Porphyromonas gingivalis is the bacterium responsible for chronic periodontitis, a severe periodontal disease. Virulence factors produced by this bacterium are secreted by the Type IX Secretion System (T9SS). The specific functions for each protein component of the T9SS have yet to be characterized thus limiting our understanding of the mechanisms associated with the translocation and modification processes of the T9SS. This study aims to identify the sequence motifs for each T9SS component and predict the functions associated with each discovered motif using motif comparisons. We extracted the sequences of 20 T9SS components from the *P. gingivalis* proteome that were experimentally identified to be important for T9SS function and used them for homology searching against fully sequenced bacterial proteomes. We developed a rigorous pipeline for the identification of seed sequences for each protein family of T9SS components. We verified that each selected seed sequence are true members of the protein family hence sharing conserved sequence motifs using profile Hidden Markov Models. The motifs for each T9SS component are identified and compared to motifs in the Pfam database. The discovered motifs for 11 components with known functions matched the motifs associated with the reported functions. We also suggested the putative functions for four components. *PorM* and *PorW* might form the putative energy transduction complex. *PorP* and *PorT* might be the putative *O*-deacylases. The identified motifs for five components matched the motifs associated with functions that related/unrelated to the T9SS.

Keywords: Motifs; periodontal disease; type IX secretion System (T9SS)

ABSTRAK

Porphyromonas gingivalis ialah bakteria yang menyebabkan periodontitis kronik, sejenis penyakit periodontium yang parah. Faktor kevirulenan yang dihasilkan oleh bakteria ini dirembeskan oleh Sistem Perembesan Jenis IX (T9SS). Fungsi khusus bagi setiap komponen protein T9SS masih belum dicirikan menyebabkan pemahaman mengenai proses translokasi dan pengubahsuaian protein oleh T9SS adalah terhad. Kajian ini bertujuan untuk mengenal pasti motif untuk setiap komponen dan meramal fungsi yang boleh dikaitkan dengan motif tersebut menggunakan kaedah perbandingan motif. Jujukan untuk 20 komponen penting T9SS yang dikenal pasti melalui uji kaji diperoleh daripada proteom *P. gingivalis*. Ia digunakan untuk pencarian homologi terhadap proteom bakteria yang mempunyai jujukan genom yang lengkap. Proses analisis yang rapi telah dihasilkan untuk pengenalanpastian jujukan benih bagi setiap famili protein komponen T9SS. Jujukan benih yang dipilih disahkan sebagai ahli sebenar famili protein tersebut justeru berkongsi motif terpelihara menggunakan profil Hidden Markov Models. Motif untuk setiap komponen telah dikenal pasti dan dibandingkan dengan motif daripada pangkalan data Pfam. Motif untuk 11 komponen yang telah diketahui fungsinya dipadankan dengan motif yang dikaitkan dengan fungsi tersebut. Kami turut mencadangkan fungsi putatif untuk empat komponen. *PorM* dan *PorW* berkemungkinan membentuk kompleks transduksi tenaga. *PorP* dan *PorT* berkemungkinan merupakan *O*-deasilase. Motif yang dikenal pasti untuk lima komponen padan dengan motif yang dikaitkan dengan fungsi berkaitan/tidak berkaitan dengan T9SS.

Kata kunci: Motif; penyakit periodontium; sistem perembesan jenis IX (T9SS)

INTRODUCTION

Periodontal diseases are caused by inflammatory conditions that occur in the oral cavity. It leads to the deterioration of the structures (bone, gingiva and periodontal ligament) that support the teeth. Prolonged inflammatory conditions in chronic periodontitis, a form of periodontal disease, would cause the destruction of the supporting structures of teeth and eventually would lead to tooth loss (Escobar et al. 2018). The study to assess the prevalence of periodontitis

among dentate adults (20-79 years) from Northern Norway reported that, out of 1,911 participants, 49.5% of them have periodontitis and 9.1% have severe periodontitis (Holde et al. 2017). Due to its prevalence, many treatments for periodontitis have been developed such as surgery, antimicrobial therapy, vaccination, tissue engineering, laser therapy and host response modulation. However, those approaches in treating periodontitis have their own weaknesses such as invasive and non-specific (Tonetti et

al. 2007). It is found that *Porphyromonas gingivalis*, one of the bacteria that forms the microbial biofilm observed in periodontitis, has acquired a unique secretion system that mediates its pathogenic capabilities as the causative agent in periodontitis (Sato et al. 2010). Therefore, knowledge about the molecular machine that contributes to the development of periodontitis might lead to a new therapeutic method for treating periodontal diseases that is non-invasive and specific.

Bacterial secretion systems are crucial molecular machines that aid numerous bacterial physiological processes important for bacterial survival. Those processes range from defense against host immune system, interaction with the environment, nutrient acquisition, self-attachment to various surfaces and transport of virulence factors to a specific destination (Letoffe et al. 1994). It was discovered that *P. gingivalis* had possibly acquired the Type IX Secretion System (T9SS) and it is thought to be exclusive to bacteria under *Bacteroidetes* phylum (Abby et al. 2016). Despite being the late colonizer during the microbial biofilm formation and the presence of many bacterial species that capable to colonize the oral cavity, *P. gingivalis* is the bacterial species that frequently associated with the severe forms of periodontitis. The T9SS enables it to acquire the key traits as an opportunistic pathogen (Haffajee et al. 1997). *P. gingivalis* depends on the T9SS to access its large arsenal of virulence factors such as gingipains, hemagglutinin and peptidylarginine deiminase as they are translocated and modified by the T9SS in order for these virulence factors to be presence on the bacterial cell surface in their mature functional forms (Gorasia et al. 2015).

Currently there are 20 protein components that have been identified to be essential for T9SS function (Sato et al. 2010). PorK and PorN formed a ring-shaped structure of 50 nm in diameter possibly associated with the T9SS secretion pore on the outer membrane that has not been identified (Gorasia et al. 2016). PorL and PorM formed the inner membrane complex where PorM spanned across the periplasm, interacted with the PorK-N complex and anchored by PorL to the inner membrane (Leone et al. 2018; Vincent et al. 2017). After the cargo protein of T9SS translocated across the inner membrane via the sec translocon, the C-terminal Domain (CTD) cargo protein is recognized by the T9SS to be translocated across the T9SS secretion pore. Then the cargo protein is captured by PorV that function as an outer membrane shuttle protein that delivered it to the attachment complex comprised of PorU and PorZ that are bounded to the outer membrane proteins PorV and PorQ, respectively (Glew et al. 2017). In the attachment complex, PorU cleaved the CTD of cargo protein and conjugated the newly formed C-terminus to an anionic lipopolysaccharide (A-LPS) that will anchor it to the cell surface. The A-LPS is delivered by PorZ as it is capable of binding A-LPS and is involved in the lipid A transport (Lpt) pathway (Glew et al. 2017). There are also a few components (PorX, PorY, PorR, and PG0162) that

are perhaps involved in the regulation of T9SS (Kadowaki et al. 2016; Shoji et al. 2014, 2002).

Despite that, there are details about the translocation and modification processes by T9SS that remain vague and the functions of a few components remained elusive. Therefore, in this study we performed motif identification among sets of sequences (seed sequences) that are proven to be members of the protein families of T9SS components. The discovered motifs are then compared to the Pfam motifs. We have identified the motifs for each component that are associated to functions that consistent with those reported by literatures. We also suggested putative functions for components that might have not yet been reported.

MATERIALS AND METHODS

SELECTION OF SEED SEQUENCES

We constructed a local database with protein sequences from bacteria with completely sequenced genomes. To construct the database, the protein sequences from GenBank were used as they were more likely to represent a larger number of bacterial species. The proteomes of the bacteria with completely sequenced genomes available in the NCBI ftp site (<ftp://ftp.ncbi.nlm.nih.gov/genomes/>) were retrieved in FASTA format in February 2017. A total of 6354 proteomes were retrieved where they represented 4494 different strains of bacteria. The NCBI taxonomy identifiers (taxonomy id) for each protein sequence in the database where each taxonomy id represents a unique bacterial strain were retrieved from the NCBI ftp site (<ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/accession2taxid>) in March 2017 as it is the interest of this research to know which species of bacteria those protein sequences belong to. Both the bacterial proteomes and the taxonomy id were used to construct the local BLAST database of bacterial protein sequences using BLAST+ version (2.6.0). The constructed database contained 22,282,422 bacterial protein sequences.

We retrieved the protein sequences of 20 T9SS components within the proteome of *P. gingivalis* from the NCBI protein database (<https://www.ncbi.nlm.nih.gov/protein/>) in February 2017. *P. gingivalis* was chosen as this type of secretion machinery is well established in this species (Sato et al. 2010). Those sequences were used as the queries for protein BLAST (BLASTP) against the local bacterial protein sequences database to find the homologs for each component (Altschul et al. 1990). The blastp homologs search was done with the maximum target sequences cut-off at a value that will ensure the blastp hits with the e-value up to the default e-value (10) will be reported in the outputs. The outputs were formatted to ensure the corresponding taxonomy id for each homolog will be included. The homologs were filtered for those with the e-value ≤ 0.001 and the query coverage $> 60\%$. Only the homologs that belong to the strains of bacteria

under the *Bacteroidetes* phylum were considered in the selection of seed sequences excluding the homologs from *Bacteroides thetaiotaomicron*, *Bacteroides fragilis* and *Bacteroides vulgatus* that were shown to not have acquired the T9SS. Only the homolog with the lowest e-value was kept for a strain of bacteria with multiple homologs for a particular component (Heath et al. 2016). These steps are necessary to minimize the possibility of false positives inclusion in the seed sequences. To reduce redundancy, sequences with the pairwise identity $\geq 80\%$ and the mutual coverage $\geq 90\%$ were excluded using CD-HIT (V4.6.8), which is a standard based on the construction of profile HMMs by Pfam (Fu et al. 2012; Skewes-Cox et al. 2014). The remaining homologs were selected as the seed sequences for each protein family of T9SS components.

EVALUATION OF THE SELECTED SEED SEQUENCES

To verify that the selected sequences were indeed members of the protein families for each component, we constructed profile HMMs for each component using the seed sequences. The alignments required to build the profile HMMs were constructed in the clustalw format from the seed sequences using T-Coffee (Notredame et al. 2000). The profile HMMs were then constructed from the alignments using hmmbuild, which is one of the command-line tools of HMMER software package (3.1b2) (Eddy 1995). The pipeline to construct the profile HMMs for each component of T9SS is shown in Figure 1.

Profile HMMs are capable of detecting homologs for the multiple sequence alignment it was derived from and it is thought to be more sensitive than sequence homology approaches in the identification of distant homologs (Karplus et al. 1998). These features enable us to verify whether the selected seed sequences were members of the protein families for each T9SS component. Each profile HMM was searched against the local bacterial protein sequences database in FASTA format using hmmsearch (part of HMMER). The homologs obtained were filtered for those with the e-value ≤ 0.001 . The remaining homologs were checked to see whether the profile HMMs could successfully recall all the sequences used to construct the profile HMMs at the threshold e-value of 0.001. To evaluate whether the threshold e-value used is reliable in excluding false positives, we shuffled all sequences in the local database using the shuffleseq, a part of EMBOSS package (6.6.0) (Rice et al. 2000). Shuffleseq rearranged each sequence in a random order while maintaining the sequence length and the residues composition. Then, each profile HMM was searched against the database comprised of shuffled sequences.

MOTIF IDENTIFICATION

MEME, which is one of the command-line tools of MEME Suite Software version (4.12.0) was used to discover motifs in the verified seed sequences for each component (Bailey & Elkan 1994). The motif identification for each component was done with the motif width cut-off at 10 to 100 positions. The maximum number of motifs to be

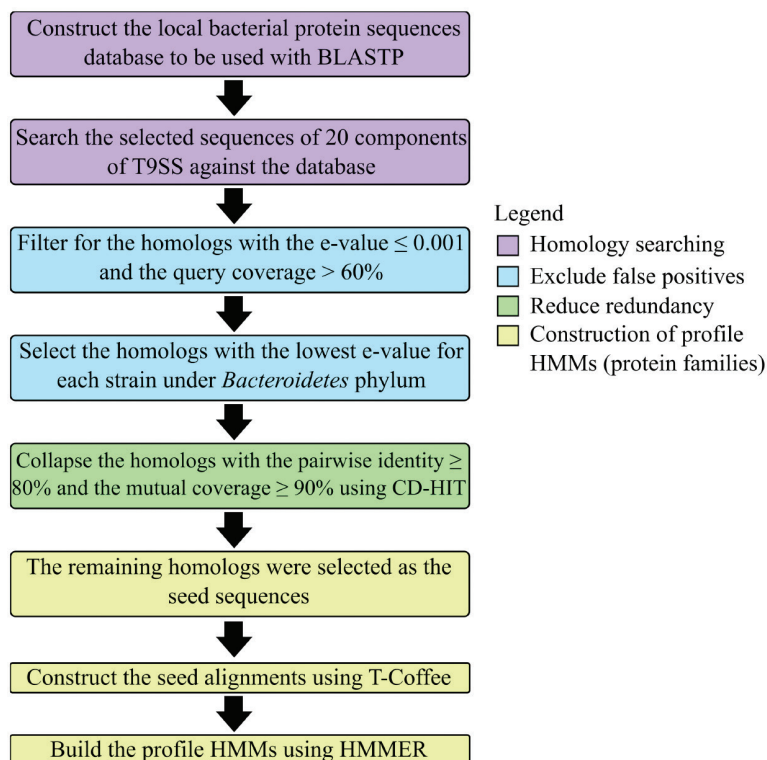


FIGURE 1. The pipeline to construct the profile HMMs for each component of T9SS

found was set to a value that will ensure MEME could find all possible motifs. The motif site distribution of OOPS (Only one site per sequence) was selected because the input sequences were verified to be members of protein families for each component thus each sequence is likely to exhibit the motifs. The discovered motifs were then filtered for overlapping motifs. The motifs that highly probable to represent the signal peptide were excluded from further analysis. The identified motifs with e-value > 0.05 (The threshold for statistical significance used by MEME) were discarded.

MOTIF COMPARISON

The remaining discovered motifs were then compared to the motifs from Pfam (Sonnhammer et al. 1997). The alignments that were used to construct the 16712 protein families in Pfam were retrieved from the FTP site of Pfam (<http://pfam.xfam.org/>) in January 2018. The alignments were then converted into the MEME motif format to construct the Pfam motifs database. The discovered T9SS components' motifs were downloaded in the MEME motif format and were used to search against the Pfam motifs database using Tomtom (part of MEME Suite). Only matches with the e-value ≤ 0.05 were deemed to be significant and reported in the result (Gupta et al. 2007).

RESULT

The constructed profile HMMs successfully recalled all the seed sequences used to construct the profile HMMs at the

e-value threshold of 0.001. At the e-value of 0.001, which is the threshold e-value used to exclude false positives, profile HMMs did not obtain any homolog when they were searched against the database of shuffled protein sequences. This illustrated the reliability of threshold e-value being used to exclude false positives. It also implied the seed sequences used to construct the profile HMMs were members of the protein families for each component of T9SS as they were recalled by the profile HMMs with the e-value ≤ the threshold e-value.

Only the identified motifs with e-value deemed to be significant by MEME were selected for motif comparison analysis. The overlapping motifs and the motifs that represented signal peptide were excluded. An example of output from the motif discovery by MEME is shown in Figure 2.

11 COMPONENTS WITH KNOWN FUNCTIONS MATCHED THE MOTIFS ASSOCIATED WITH THE REPORTED FUNCTIONS

One of the identified PorY motifs matched the motif from PF02518 (Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase) (Figure 3(A)). The PorY motif also matched the motif from PF00512 (His Kinase A (phospho-acceptor) domain) (Figure 3(B)). However, this match has an e-value greater than 0.05, which is the threshold e-value being used. Hence, based on the e-value, the similarity to that motif is insignificant. Despite that, we decide to report it due to its consistency with the reported function from literatures. The PorX motif matched the motif from PF00072 (Response regulator receiver domain) (Figure 3(C)).

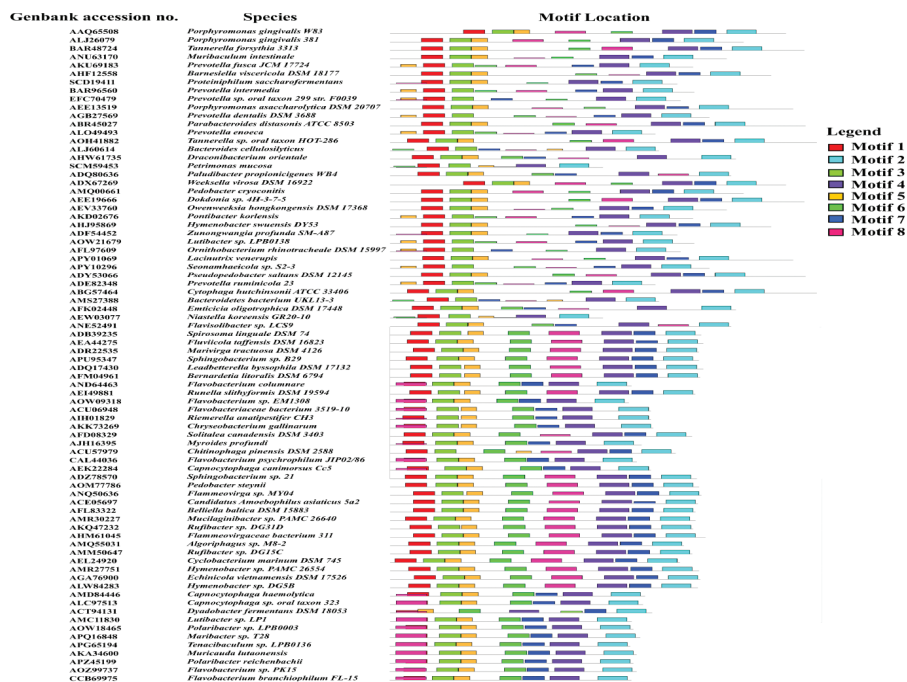


FIGURE 2. The figure shows the motifs discovered by MEME for PorL. The left column shows the Genbank accession numbers for each PorL homolog, the middle column shows the bacterial strains those homologs belong to, and the right column shows the motifs distribution. The overlapping motif (motif 8) is excluded from motif comparison

The PG0162 motif matched the motif from PF08281 (Sigma-70, region 4). The PorL motif matched the motif from PF10046 (Biogenesis of lysosome-related organelles complex-1 (BLOC-1) subunit 2). 12 out of 17 identified PorR motifs matched the motifs from PF01041 (DegT/DnrJ/EryC1/StrS aminotransferase family). The PGN0300 motif matched the motif from PF03938 (Outer membrane protein). The PorN motif matched the motif from PF00041 (Fibronectin type III domain). The identified PG1058 motif matched the motif from PF07676 (WD40-like Beta Propeller Repeat). The PG1058 motif also matched the motif from PF00515 (Tetratricopeptide repeat). 11 out of 40 identified PorU motifs matched the motifs from PF01364 (Peptidase family C25) (Figure 3(D) & 3(E)).

The identified PorV motif matched the motif from PF09411 (Lipid A 3-O-deacylase) (Figure 4(A)). The PorV motif also matched the motif from PF13188 (PAS domain). The identified PorZ motif matched the motif from PF01453 (D-mannose binding lectin).

FOUR COMPONENTS WITH UNKNOWN FUNCTIONS MATCHED THE MOTIFS THAT SUGGESTED THEIR PUTATIVE FUNCTIONS

Similar to PorV (Figure 4(A)), both motifs of PorP (Figure 4(B)) and PorT (Figure 4(C)) matched the motifs from PF09411 (Lipid A 3-O-deacylase).

The PorM motif matched the motif from PF02518 (Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase) (Figure 4(D)). The PorW motif matched the motif from PF00213 (ATP synthase delta subunit) (Figure 4(E)). The PorW motif also matched the motif from PF00515 (Tetratricopeptide repeat) (Figure 4(F)).

FIVE COMPONENTS MATCHED THE MOTIFS ASSOCIATED WITH FUNCTIONS THAT MIGHT BE RELATED OR UNRELATED TO T9SS

There are five components (PorK, PorQ, Sov, PG0189, and PG0534) with motifs that matched the motifs associated with functions that perhaps related or unrelated to T9SS functions as listed in Table 1.

DISCUSSION

There are 20 protein components that made up the T9SS where four of them (PorY, PorX, PorL, and PG0162) might involve in the regulation of T9SS functions (Kadowaki et al. 2016; Vincent et al. 2016). PorX and PorY formed the putative two-component system (TCS) that can regulate the *por* genes (e.g., *porP*, *porK*, *porL*, *porM*, and *porN*) at the transcriptional level. In the TCS, PorY is the sensor histidine kinase which undergoes autophosphorylation at the conserved histidine in response to the environmental signal (Kadowaki et al. 2016).

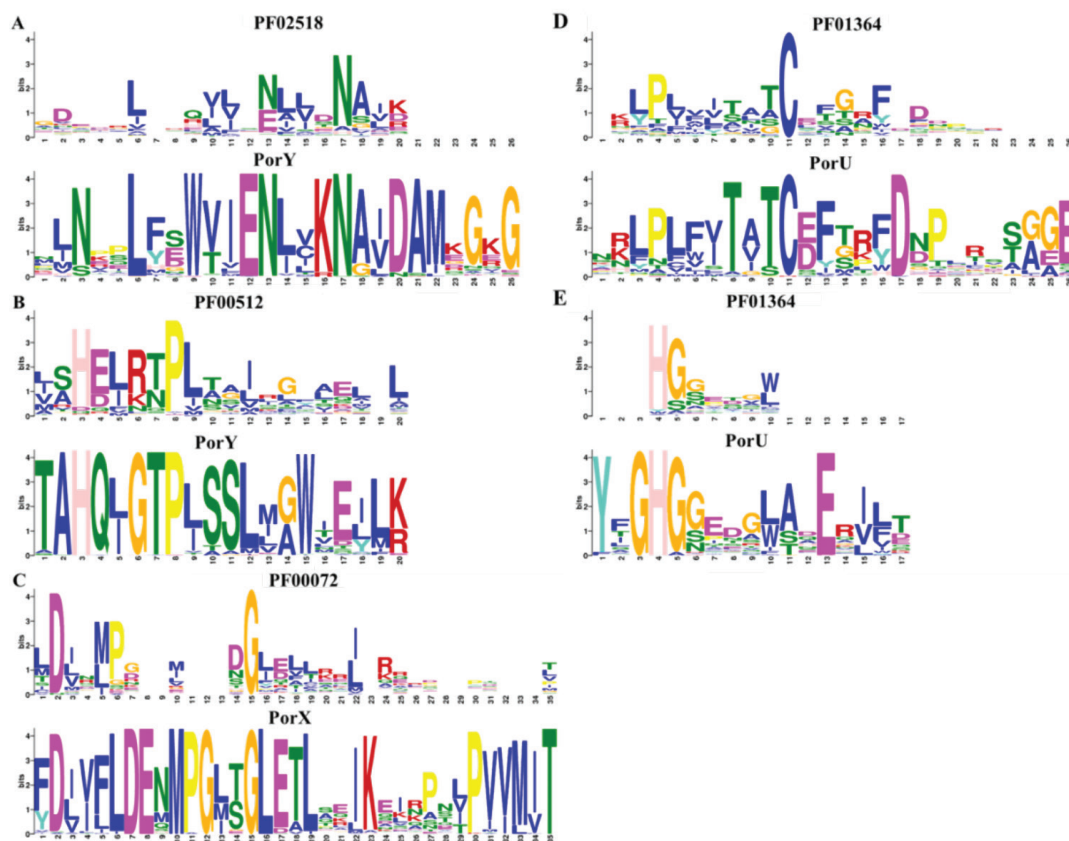


FIGURE 3. The alignments of PorX, PorY and PorU motifs to their respective motif matches

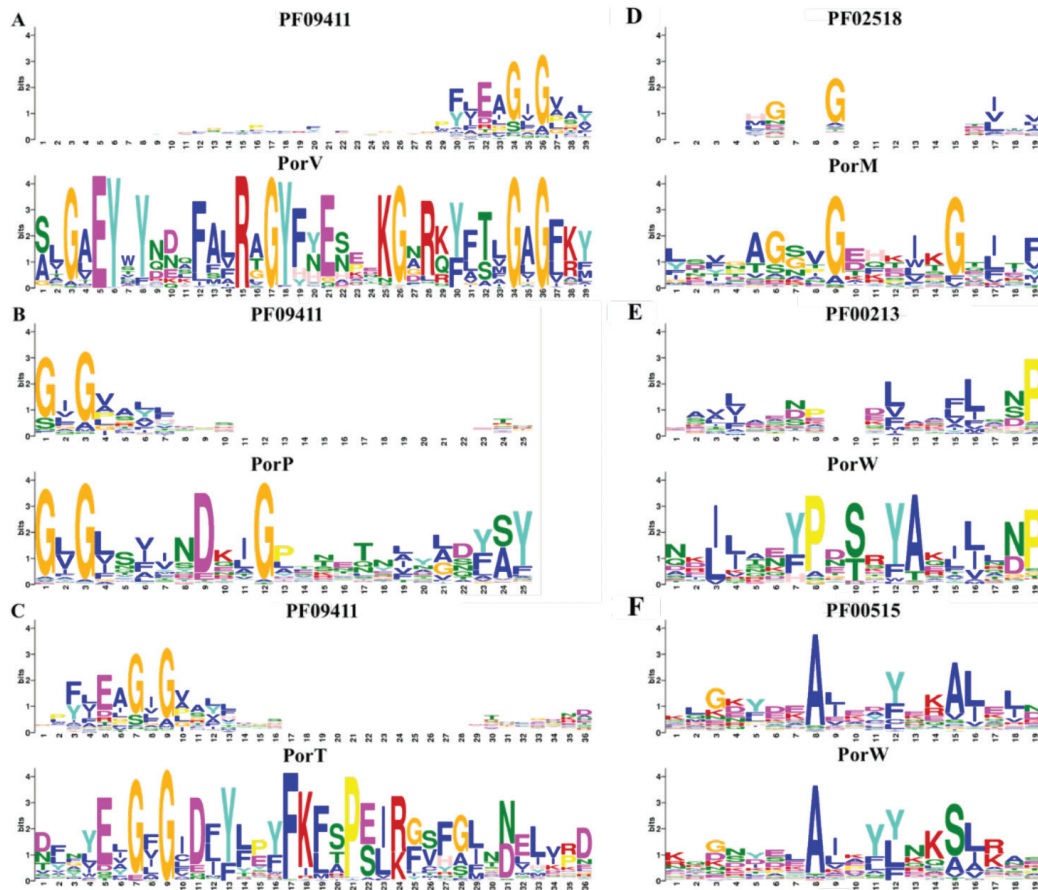


FIGURE 4. The alignments of PorV, PorP, PorT, PorM and PorW motifs to their respective motif matches

TABLE 1. The matches for T9SS component motifs that related/unrelated to T9SS functions

| | Match |
|--------|---|
| PorQ | PF13777 (Domain of unknown function) |
| PorK | PF03781 (Sulfatase-modifying factor enzyme 1) |
| Sov | PF02625 (XdhC and CoxI family) |
| PG0189 | PF02978 (Signal peptide binding domain) |
| PG0534 | PF13715 (CarboxypepD_reg-like domain) |

We found the motifs of PorY that can be associated to its function as the histidine kinase. Figure 3(A) shows the PorY motif that matched the motif of Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase. This motif possibly related with the ATPase activity of histidine kinase where this will allow PorY to bind and hydrolyze ATP. The released phosphoryl group from the ATP hydrolysis will be used to phosphorylate the conserved histidine (Kadowaki et al. 2016). Figure 3(B) shows the PorY motif that matched the motif of His Kinase A (phospho-acceptor) domain. The motif possibly will receive the released phosphoryl group as evidenced by the presence of a conserved histidine at position 3 of the aligned motifs. This conserved histidine will be phosphorylated and it is crucial for the activation of its cognate response regulator partner in the TCS (Kadowaki et al. 2016).

PorX is reported to be the response regulator protein (Kadowaki et al. 2016). Figure 3(C) shows the PorX motif that matched the motif of response regulator receiver domain. This motif will be phosphorylated by PorY through the transfer of phosphoryl group from the histidine of PorY to the aspartate of PorX (Kadowaki et al. 2016). This is evidenced by the presence of a conserved aspartate at position 2 of the aligned motifs.

It is found that the phosphorylated PorX (pPorX) cannot directly bind the promoter region of *por* genes. Hence, to compensate for the lack of DNA-binding domain, PorX depends on PG0162 to bind the promoter region upstream of *por* genes (Kadowaki et al. 2016). We found PG0162 motif that matched the motif from Sigma-70, region 4 that probably formed its DNA-binding domain

that will allow the complex of PG0162 and pPorX to bind the DNA (Dou et al. 2010; Kadowaki et al. 2016).

The TCS also exerts regulatory function through the interaction between pPorX with PorL. In gliding bacteria under *Bacteroidetes* phylum, such as *Flavobacterium johnsoniae*, the T9SS acts as a rotary machine where the binding of pPorX to PorL will induce the reversal of T9SS rotation, hence affecting the movement of gliding bacteria (Vincent et al. 2016). We have identified the PorL motif that matched the motif of Biogenesis of lysosome-related organelles complex-1 (BLOC-1) subunit 2. The PorL motif perhaps responsible for the interaction between pPorX and PorL or PorL and PorM. BLOC-1 is a multi-subunit protein complex comprised of 8 different subunits which one of them is BLOC-1 subunit 2. Through the yeast two-hybrid experiments, this subunit shown to be involved in a network of binary interactions with the other subunits of BLOC-1 (Starcevic & Dell'Angelica 2004). Thus, it is possible that the motif of BLOC-1 subunit 2 that matched the PorL motif is involved in a protein-protein interaction. An important distinction is that *P. gingivalis* is a non-motile bacterium, thus the interaction between PorL and pPorX might have a function different from the gliding associated role it has been known for (Vincent et al. 2016).

PorR is involved in the Wbp pathway that leads to the biosynthesis of A-LPS (Shoji et al. 2018). We have identified the PorR motifs that matched the motifs from the DegT family. This suggested the identified motifs are perhaps associated with the aminotransferase activity that is crucial for the biosynthesis of A-LPS (Shoji et al. 2002). It is not yet known whether PorR has regulatory effect on the biosynthesis of A-LPS and therefore on the function of T9SS.

PGN0300 is the component that can influence the function of PorU (Taguchi et al. 2016). We obtained the PGN0300 motif that matched the motif of the Skp outer membrane protein. The Skp protein acts like a molecular chaperone that interacts with the unfolded proteins as they pass through the sec translocon (Sklar et al. 2007). Thus, the motif might involve in PGN0300 skp-like function and ensure the present of PorU on the cell surface. However, the mechanism behind this function and whether it has regulatory effect remains unclear (Taguchi et al. 2016).

T9SS is suggested to have energetic components that provide energy for the substrate translocation and T9SS components assembly (Vincent et al. 2017). This is similar to the Type IV Secretion System where its energetic components made up of the VirD4-like receptor, VirB11 ATPase and VirB4 ATPase, where they supply energy for the substrate translocation across the inner membrane (Atmakuri et al. 2004).

In the T9SS, two components, PorL, and PorM are proposed to form an energy transducer complex. PorL forms a trimer in the inner membrane while PorM forms a dimer that spans the periplasm and is anchored in the inner membrane by its two transmembrane helices (Vincent et al. 2017). It is suggested that PorL and PorM

interact through their transmembrane helices in the inner membrane. The periplasmic domain of PorM interacts with the PorK-PorN complex that perhaps associated with the secretion pore in the outer membrane (Vincent et al. 2017). It is proposed that the PorK-PorN complex might control the access to the secretion pore and the periplasmic domain of PorM that interacts with the complex perhaps can regulate the opening of PorK-PorN complex. It is also proposed that the PorM is powered by the PorL trimer. Hence, both components formed the energy transduction complex (Leone et al. 2018).

We found a PorM motif that matched the Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase motif as shown in Figure 4(D). This implied that PorM might capable to hydrolyse ATP and enable the conformational change of PorM which can influence the opening of PorK-PorN complex (Leone et al. 2018). However, we could not find a PorL motif that could match the function as the energy transduction complex. It might be that the interaction between PorL and PorM has a regulatory effect rather than an energy transduction function. In the event of TCS receiving an environmental stimulus that does favor the substrate translocation by T9SS, the pPorX that formed will interact with PorL. This interaction might influence the PorM conformation hence the opening of PorK-PorN complex.

PorW is a periplasmic component of T9SS (Sato et al. 2010). We identified a PorW motif that matched the ATP synthase delta subunit motif as shown in Figure 4(E). We also identified a PorW motif that matched the Tetratricopeptide repeat (TPR) motif that reported to be involved in a protein-protein interaction (Das et al. 1998) as shown in Figure 4(F). Not much is known about the specific role of PorW (Sato et al. 2010). However, the localization of PorW in the periplasm (Sato et al. 2010) and the presence of motifs that can be associated to the ATP synthase and the protein-protein interaction functions might suggest PorW and PorM formed the energy transduction complex and energized the substrate translocation by the T9SS. However, experimental validations are needed for the ATPase activities of PorM and PorW and the interaction between both proteins.

As mentioned before, PorK, and PorN formed a complex in which its exact composition is still ambiguous. Gorasia et al. (2016) reported that PorK, and PorN formed a ring-shaped structure of 50 nm in diameter with ~ 32-36 subunits of each component while Leone et al. (2018) reported that the PorK-PorN complex consisted of 2 subunits of each component (PorN₂/ PorK₂). Regardless, in both reports, the PorK-PorN complex is associated with the secretion pore on the outer membrane. We found a PorN motif that matched the Fibronectin type III domain motif. Fibronectin is known to interact with various proteins such as collagen, fibrin, heparin and integrin (Leahy et al. 1996). Hence, the PorN motif might mediate the protein-protein interactions between PorK-PorN, PorM-PorN and PorN-PorN (Vincent et al. 2017). We identified the

PorK motifs that matched the Sulfatase-modifying factor enzyme 1 motifs. However, we are uncertain whether the function associated with the PorK motifs is related to the PorK function as one of the components of T9SS.

After the T9SS substrates pass through the inner membrane via the sec translocon, they will be guided by various protein components to reach the attachment complex on the bacterial cell surface to undergo protein modifications. It is thought that there are components that formed the periplasm complex that will interact with the protein substrates and aid their transit across the periplasm (Heath et al. 2016). The components that make up the periplasm complex are still ambiguous. However, PG1058 is suggested to form a structural scaffold for the assembly of T9SS components that might include the periplasm complex (Heath et al. 2016). We identified the PG1058 motifs that matched the WD40-like Beta Propeller Repeat (b-propeller repeat) and the TPR motifs. Both repeats are reported to be involved in protein-protein interactions, thus the identified motifs might mediate the protein-protein interactions between PG1058 and the other protein components that made up the periplasm complex (Das et al. 1998; Heath et al. 2016; Kopeck & Lupas 2013).

After the substrates transit across the periplasm and reach the periplasmic side of outer membrane, they will be translocated by the secretion pore to the bacterial cell surface. The proteins that made up the secretion pore have not yet been identified. However, the PorK-PorN complex is known to be associated with the secretion pore and controls the entry of substrates into the secretion pore (Gorasia et al. 2016; Leone et al. 2018).

At the cell surface, each substrate is captured by PorV that will transport it to the attachment complex. It is believed that PorV interacts with the CTD of substrate (Glew et al. 2017). We identified the PorV motif that matched the motif of PAS domain. PAS domain has been reported to have protein binding function which implied that the identified PorV motif might involve in mediating the interaction between substrate CTD with PorV (Ponting & Aravind 1997). It is also found that PorV interacts with the CTD of PorU, which is one of the two components of T9SS that acquired the CTD. Thus, PorV is the outer membrane protein that bounded PorU to the cell surface and perhaps orientated it suitable for the cleavage of substrate CTD (Glew et al. 2017).

We have identified the PorU motifs that matched the cysteine peptidase motifs. As shown in Figure 3(D), the PorU motif exhibited the presence of a conserved cysteine at position 11. While in Figure 3(E), the PorU motif displayed the presence of a conserved histidine at position 4. These conserved residues are typical among cysteine peptidases that acquired the catalytic dyads that are important for their peptidase activities (Roberts & Belsham 1995). These motifs further support the role of PorU as the protein component that cleaved off the CTD of T9SS substrate.

The new C-terminus formed after the cleavage of substrate CTD will be conjugated to the A-LPS. The A-LPS is

suggested to be delivered by PorZ. Similar to PorU, PorZ is one of the T9SS components that acquired the CTD and bounded to the cell surface by the other T9SS component (Lasica et al. 2016). However, PorZ interacts with PorQ instead and it is believed that PorQ interacts with the CTD of PorZ. Hence, they formed the PorZ-PorQ subcomplex that binds to the PorU-PorV subcomplex only when PorZ binds to the A-LPS. The interaction between the two subcomplexes constituted the attachment complex. This is where the protein modifications (CTD cleavage and A-LPS attachment) will take place upon binding of the substrate-PorV complex to the attachment complex. After the modifications, the substrate will be anchored by the A-LPS to the bacterial cell surface (Glew et al. 2017).

We discovered the PorZ motif that matched the D-mannose binding lectin motif. This protein is shown to acquire a conserved mannose-binding site (Hester et al. 1995). This suggests that the identified motif might be responsible for PorZ's ability to bind A-LPS, possibly at its mannose moiety (Gorasia et al. 2015). None of the identified PorQ motifs matched the motif related to protein binding. However, some of those motifs matched the domain of unknown function motifs, possibly involve in protein binding.

It is proposed that the substrate translocation pathway as described above is co-ordinated to the Lpt pathway that involved in the transport of A-LPS. PorV is reported to be involved in the Lpt pathway as it performed one of the steps in the pathway which is the O-deacylation of mono-phosphorylated lipid A. The end product of this pathway is the mature form of A-LPS that anchored the substrate on the cell surface (Chen et al. 2011). The dual role played by PorV (binds CTD-containing proteins and as a O-deacylase) proved that it is a multifunctional protein (Glew et al. 2017). As shown in Figure 4(A), we have identified the PorV motif that matched the Lipid A 3-O-deacylase motif. It is possible that the motif is related to the PorV function as a O-deacylase.

Both *porP* and *porT* mutants exhibited the phenotypes that implied the impairment of T9SS functions. However, the specific functions that they perform might have not yet been reported (Nguyen et al. 2009; Vincent et al. 2017). As shown in Figure 4(B) and Figure 4(C), we identified the PorP and PorT motifs that matched the Lipid A 3-O-deacylase motifs. This similarity suggested that both PorP and PorT might be able to perform the deacylation of mono-phosphorylated lipid A, like PorV. The motifs of PorP (positions 1-6), PorT (positions 7-12) and PorV (positions 34-39) matched to the similar motif of Lipid A 3-O-deacylase that further implied that the three proteins might share a similar function. It is reported that the computationally predicted structures of the three proteins are b-barrel structures and those proteins are localized on the outer membrane of the bacterial cell (Glew et al. 2017; Nguyen et al. 2009; Vincent et al. 2017). This further strengthens the possibility of PorP and PorT having acquired a similar function to PorV. Hence, we suggested PorP and

PorT as the putative lipid A 3-O-deacylases. Experimental validations are required to validate the deacylase activities of PorP and PorT proteins.

The identified motifs for five components of the T9SS, PorK, PorQ, Sov (Saiki & Konishi 2010a), PG0189 (Vincent et al. 2017) and PG0534 (Saiki & Konishi 2010b) their secretion and biogenesis process is not yet fully elucidated. The PG0534 gene of *P. gingivalis* W83 encodes a novel protein, PG534, of unknown function. In a PG0534 deletion mutant 83K25, the activities of Arg-gingipains (RgpA and RgpB, matched the motifs associated with the functions that may or may not related to the T9SS. These are listed in Table 1. There is a lack of evidence from literature to support the association between the motifs with those functions. It is also unclear how the predicted functions for each motif fit into the current information regarding the processes of translocation and modification of substrates by the T9SS.

CONCLUSION

The elucidation of the mechanisms behind the translocation and modifications of substrates of the T9SS is still in its early phase. Currently there are 20 components that have been identified to be important for a functional T9SS. We have identified the motifs for 11 components that could be associated to their known reported functions. In doing so, we are able to propose the putative functions for PorM, PorW, PorP and PorT that to our knowledge have not yet been reported. PorM and PorW might form the putative energy transduction complex while PorP and PorT might be the putative O-deacylases. There are identified motifs for some components without clear probable functions.

ACKNOWLEDGEMENTS

The authors acknowledged Universiti Kebangsaan Malaysia (UKM) grant, GGPM-2016-016 for providing funds to support the research.

REFERENCES

- Abby, S.S., Cury, J., Guglielmini, J., Néron, B., Touchon, M. & Rocha, E.P. 2016. Identification of protein secretion systems in bacterial genomes. *Scientific Reports* 6: 23080.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215(3): 403-410.
- Atmakuri, K., Cascales, E. & Christie, P.J. 2004. Energetic components VirD4, VirB11 and VirB4 mediate early DNA transfer reactions required for bacterial type IV secretion. *Molecular Microbiology* 54(5): 1199-1211.
- Bailey, T.L. & Elkan, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in bipolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* 2: 28-36.
- Chen, Y.Y., Peng, B., Yang, Q., Glew, M.D., Veith, P.D., Cross, K.J. & Goldie, K.N. 2011. The outer membrane protein LptO is essential for the O-deacylation of LPS and the co-ordinated secretion and attachment of A-LPS and CTD proteins in *Porphyromonas gingivalis*. *Molecular Microbiology* 79(5): 1380-1401.
- Das, A.K., Cohen, P.T. & Barford, D. 1998. The structure of the tetratricopeptide repeats of protein phosphatase 5: Implications for TPR-mediated protein-protein interactions. *The EMBO Journal* 17(5): 1192-1199.
- Dou, Y., Osbourne, D., McKenzie, R. & Fletcher, H.M. 2010. Involvement of extracytoplasmic function sigma factors in virulence regulation in *Porphyromonas gingivalis* W83. *FEMS Microbiology Letters* 312(1): 24-32.
- Eddy, S.R. 1995. Multiple alignment using hidden Markov models. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology* 3: 114-120.
- Escobar, G.F., Abdalla, D.R., Beghini, M., Gotti, V.B., Rodrigues, V.J., Napimoga, M.H. & Ribeiro, B.M. 2018. Levels of pro and anti-inflammatory cytokines and C-reactive protein in patients with chronic periodontitis submitted to nonsurgical periodontal treatment. *Asian Pacific Journal of Cancer Prevention* 19(7): 1927-1933.
- Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23): 3150-3152.
- Glew, M.D., Veith, P.D., Chen, D., Gorasia, D.G., Peng, B. & Reynolds, E.C. 2017. PorV is an outer membrane shuttle protein for the type IX secretion system. *Scientific Reports* 7(1): 8790.
- Gorasia, D.G., Veith, P.D., Hanssen, E.G., Glew, M.D., Sato, K., Yukitake, H. & Nakayama, K. 2016. Structural insights into the PorK and PorN components of the *Porphyromonas gingivalis* type IX secretion system. *PLoS Pathogens* 12(8): e1005820.
- Gorasia, D.G., Veith, P.D., Chen, D., Seers, C.A., Mitchell, H.A., Chen, Y.Y. & Glew, M.D. 2015. *Porphyromonas gingivalis* type IX secretion substrates are cleaved and modified by a sortase-like mechanism. *PLoS Pathogens* 11(9): e1005152.
- Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. & Noble, W.S. 2007. Quantifying similarity between motifs. *Genome Biology* 8(2): R24.
- Haffajee, A.D., Cugini, M.A., Dibart, S., Smith, C., Kent Jr., R.L. & Socransky, S.S. 1997. The effect of SRP on the clinical and microbiological parameters of periodontal diseases. *Journal of Clinical Periodontology* 24(5): 324-334.
- Heath, J.E., Seers, C.A., Veith, P.D., Butler, C.A., Muhammad, N.A.N., Chen, Y.Y. & Slakeski, N. 2016. PG1058 is a novel multidomain protein component of the bacterial type IX secretion system. *PLoS ONE* 11(10): e0164313.
- Hester, G., Kaku, H., Goldstein, I.J. & Wright, C.S. 1995. Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. *Nature Structural Biology* 2(6): 472-479.
- Holde, G.E., Oscarson, N., Trovik, T.A., Tillberg, A. & Jönsson, B. 2017. Periodontitis prevalence and severity in adults: A cross-sectional study in Norwegian circumpolar communities. *Journal of Periodontology* 88(10): 1012-1022.
- Kadowaki, T., Yukitake, H., Naito, M., Sato, K., Kikuchi, Y., Kondo, Y. & Shoji, M. 2016. A two-component system regulates gene expression of the type IX secretion component proteins via an ECF sigma factor. *Scientific Reports* 6: 23288.
- Karplus, K., Barrett, C. & Hughey, R. 1998. Hidden Markov models for detecting remote protein homologies. *Bioinformatics* 14(10): 846-856.
- Kopec, K.O. & Lupas, A.N. 2013. β -Propeller blades as ancestral peptides in protein evolution. *PLoS ONE* 8(10): e77074.

- Lasica, A.M., Goulas, T., Mizgalska, D., Zhou, X., De Diego, I., Ksiazek, M. & Madej, M. 2016. Structural and functional probing of PorZ, an essential bacterial surface component of the type-IX secretion system of human oral-microbiomic *Porphyromonas gingivalis*. *Scientific Reports* 6: 37708.
- Leahy, D.J., Aukhil, I. & Erickson, H.P. 1996. 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell* 84(1): 155-164.
- Leone, P., Roche, J., Vincent, M.S., Tran, Q.H., Desmyter, A., Cascales, E. & Kellenberger, C. 2018. Type IX secretion system PorM and gliding machinery GldM form arches spanning the periplasmic space. *Nature Communications* 9(1): 429.
- Letoffe, S., Ghigo, J.M. & Wandersman, C. 1994. Secretion of the *Serratia marcescens* HasA protein by an ABC Transporter. *Journal of Bacteriology* 176(17): 5372-5377.
- Nguyen, K.A., Zyllicz, J., Szczesny, P., Sroka, A., Hunter, N. & Potempa, J. 2009. Verification of a topology model of PorT as an integral outer-membrane protein in *Porphyromonas gingivalis*. *Microbiology* 155(2): 328-337.
- Notredame, C., Higgins, D.G. & Heringa, J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* 302(1): 205-217.
- Ponting, C.P. & Aravind, L. 1997. PAS: A multifunctional domain family comes to light. *Current Biology* 7(11): R674-R677.
- Rice, P., Longden, I. & Bleasby, A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* 16(6): 276-277.
- Roberts, P.J. & Belsham, G.J. 1995. Identification of critical amino acids within the foot-and-mouth disease virus leader protein, a cysteine protease. *Virology* 213(1): 140-146.
- Saiki, K. & Konishi, K. 2010a. The role of Sov protein in the secretion of gingipain protease virulence factors of *Porphyromonas gingivalis*. *FEMS Microbiology Letters* 302(2): 166-174.
- Saiki, K. & Konishi, K. 2010b. Identification of a novel *Porphyromonas gingivalis* outer membrane protein, PG534, required for the production of active gingipains. *FEMS Microbiology Letters* 310(2): 168-174.
- Sato, K., Naito, M., Yukitake, H., Hirakawa, H., Shoji, M., McBride, M.J. & Rhodes, R.G. 2010. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proceedings of the National Academy of Sciences of the USA* 107(1): 276-281.
- Shoji, M., Sato, K., Yukitake, H., Kamaguchi, A., Sasaki, Y., Naito, M. & Nakayama, K. 2018. Identification of genes encoding glycosyltransferases involved in lipopolysaccharide synthesis in *Porphyromonas gingivalis*. *Molecular Oral Microbiology* 33(1): 68-80.
- Shoji, M., Sato, K., Yukitake, H., Naito, M. & Nakayama, K. 2014. Involvement of the Wbp pathway in the biosynthesis of *Porphyromonas gingivalis* lipopolysaccharide with anionic polysaccharide. *Scientific Reports* 4: 5056.
- Shoji, M., Ratnayake, D.B., Shi, Y., Kadowaki, T., Yamamoto, K., Yoshimura, F. & Akamine, A. 2002. Construction and characterization of a nonpigmented mutant of *Porphyromonas gingivalis*: Cell surface polysaccharide as an anchorage for gingipains. *Microbiology* 148(4): 1183-1191.
- Skewes-Cox, P., Sharpton, T.J., Pollard, K.S. & DeRisi, J.L. 2014. Profile Hidden Markov Models for the detection of viruses within metagenomic sequence data. *PLoS ONE* 9(8): e105067.
- Sklar, J.G., Wu, T., Kahne, D. & Silhavy, T.J. 2007. Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. *Genes and Development* 21(19): 2473-2484.
- Sonnhammer, E.L., Eddy, S.R. & Durbin, R. 1997. Pfam: A comprehensive database of protein domain families based on seed alignments. *Proteins: Structure, Function and Genetics* 28(3): 405-420.
- Starcevic, M. & Dell'Angelica, E.C. 2004. Identification of Snapin and three novel proteins (BLOS1, BLOS2 and BLOS3/reduced pigmentation) as subunits of biogenesis of Lysosome-related Organelles Complex-1 (BLOC-1). *Journal of Biological Chemistry* 279(27): 28393-28401.
- Taguchi, Y., Sato, K., Yukitake, H., Inoue, T., Nakayama, M., Naito, M. & Kondo, Y. 2016. Involvement of an Skp-Like protein, PGN_0300, in the type IX secretion system of *Porphyromonas gingivalis*. *Infection and Immunity* 84(1): 230-240.
- Tonetti, M.S., D'Aiuto, F., Nibali, L., Donald, A., Storry, C., Parkar, M. & Suvan, J. 2007. Treatment of periodontitis and endothelial function. *The New England Journal of Medicine* 356(9): 911-920.
- Vincent, M.S., Canestrari, M.J., Leone, P., Stathopoulos, J., Ize, B., Zoued, A. & Cambillau, C. 2017. Characterization of the *Porphyromonas gingivalis* type IX secretion trans-envelope PorKLMNP core complex. *Journal of Biological Chemistry* 292(8): 3252-3261.
- Vincent, M.S., Durand, E. & Cascales, E. 2016. The PorX response regulator of the *Porphyromonas gingivalis* PorXY two-component system does not directly regulate the type IX secretion genes but binds the PorL subunit. *Frontiers in Cellular and Infection Microbiology* 6: 96.

Centre for Bioinformatics Research
 Institute of Systems Biology (INBIOSIS)
 Universiti Kebangsaan Malaysia
 43600 UKM Bangi, Selangor Darul Ehsan
 Malaysia

*Corresponding author; email: norazlanm@ukm.edu.my

Received: 30 May 2018

Accepted: 14 September 2018