# Phylogenetic analysis of Type IX Secretion System (T9SS) protein components revealed that PorR undergoes horizontal gene transfer

**Reeki Emrizal** [1], **Nor Azlan Nor Muhammad** [Corresp. 1]

[1] Institute of Systems Biology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

Corresponding Author: Nor Azlan Nor Muhammad
Email address: norazlannm@ukm.edu.my

*Porphyromonas gingivalis* is one of the major bacteria that causes periodontitis. Chronic periodontitis is a severe form of periodontal disease that occurs due to prolong inflammatory conditions. If left untreated, deterioration of the supporting structures such as gingiva, bone, and ligament can ultimately lead to tooth loss. Virulence factors produce by *P. gingivalis* that are responsible for the pathophysiology of periodontitis are secreted by Type IX Secretion System (T9SS). T9SS-acquiring bacteria have been linked to several systemic diseases such as atherosclerosis, aspiration pneumonia, cancer, rheumatoid arthritis, and diabetes mellitus. This study aims to investigate the phylogenetic relationship and taxonomic distribution between putative members of T9SS component protein families. There are 20 protein components of T9SS being investigated in this study. We have constructed multiple sequence alignments for each component using homologs of those components. Then we proceed to phylogenetic analysis by constructing the maximum-likelihood (ML) trees. ML trees for 19 protein components of T9SS exhibit clustering of terminal nodes based on their respective classes under Bacteroidetes phylum. The ML tree of PorR, which is an aminotransferase that involved in Wbp pathway that produces structural sugar of A-LPS, exhibits different clustering pattern of terminal nodes where the nodes do not cluster based on their respective classes. Hence, PorR might evolve independently from the other T9SS protein components which might suggest that PorR is acquired by T9SS-acquiring bacteria through horizontal gene transfer. The part of *P. gingivalis* strain ATCC 33277 genome that contains *porR* gene has been extracted to support the possibility that *porR* gene has been horizontally transferred. Through homology searching using NCBI blastx, we found that seven genes (including *porR*) that involved in the biosynthesis of A-LPS that anchored the virulence factor secreted by T9SS to bacterial cell surface are flanked by insertion sequences (ISs) that encode IS5 family transposase. The IS5 transposons contain a single open reading frame that encodes for the transposase that will cleave the 12 bp inverted repeats that flanked the transposons.

Consequently, this can mobilise the intervening DNA segment that contains *porR* gene and subsequently contributes to the possibility that *porR* gene is subjected to conjugative transfer. The taxonomic distribution of T9SS protein components revealed that they can be found across all classes under Bacteroidetes phylum. Additionally, we have identified species under Chitinophagia, Saprospiria, and unclassified that acquired homologs of T9SS protein components that, to our knowledge, have not been reported. In conclusion, this study can provide a better understanding about the phylogeny and taxonomic distribution of T9SS protein components.

# Phylogenetic analysis of Type IX Secretion System (T9SS) protein components revealed that PorR undergoes horizontal gene transfer

Reeki Emrizal[1], Nor Azlan Nor Muhammad[1]

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

Corresponding Author:
Nor Azlan Nor Muhammad[1]
Centre for Bioinformatics Research, Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia
Email address: norazlannm@ukm.edu.my

## Abstract

*Porphyromonas gingivalis* is one of the major bacteria that causes periodontitis. Chronic periodontitis is a severe form of periodontal disease that occurs due to prolong inflammatory conditions. If left untreated, deterioration of the supporting structures such as gingiva, bone, and ligament can ultimately lead to tooth loss. Virulence factors produce by *P. gingivalis* that are responsible for the pathophysiology of periodontitis are secreted by Type IX Secretion System (T9SS). T9SS-acquiring bacteria have been linked to several systemic diseases such as atherosclerosis, aspiration pneumonia, cancer, rheumatoid arthritis, and diabetes mellitus. This study aims to investigate the phylogenetic relationship and taxonomic distribution between putative members of T9SS component protein families. There are 20 protein components of T9SS being investigated in this study. We have constructed multiple sequence alignments for each component using homologs of those components. Then we proceed to phylogenetic analysis by constructing the maximum-likelihood (ML) trees. ML trees for 19 protein components of T9SS exhibit clustering of terminal nodes based on their respective classes under Bacteroidetes phylum. The ML tree of PorR, which is an aminotransferase that involved in Wbp pathway that produces structural sugar of A-LPS, exhibits different clustering pattern of terminal nodes where the nodes do not cluster based on their respective classes. Hence, PorR might evolve independently from the other T9SS protein components which might suggest that PorR is acquired by T9SS-acquiring bacteria through horizontal gene transfer. The part of *P. gingivalis* strain ATCC 33277 genome that contains *porR* gene has been extracted to support the possibility that *porR* gene has been horizontally transferred. Through homology searching using NCBI blastx, we found that seven genes (including *porR*) that involved in the biosynthesis of A-LPS that anchored the virulence factor secreted by T9SS to bacterial cell surface are flanked by insertion sequences (ISs) that encode IS5 family transposase. The IS5 transposons contain a single open reading frame that encodes for the transposase that will cleave the 12 bp inverted repeats that flanked the transposons. Consequently, this can mobilise the intervening DNA segment that contains *porR* gene and subsequently contributes to the possibility that *porR* gene is subjected to conjugative transfer. The taxonomic distribution of T9SS protein components revealed that they can be found across all classes under Bacteroidetes phylum. Additionally, we have identified species under Chitinophagia, Saprospiria, and unclassified that acquired homologs of T9SS protein components that, to our knowledge, have not been reported. In conclusion, this study can provide a better understanding about the phylogeny and taxonomic distribution of T9SS protein components.


## Introduction

Periodontitis is a form of periodontal disease that is driven by the inflammatory conditions that have deteriorating effects on the structures that support the teeth that include gingiva (gum), alveolar bone, and periodontal ligament. Prolonged inflammatory conditions in chronic

periodontitis can cause the destruction of those supporting structures that ultimately lead to tooth loss and might contribute to the systemic inflammation (Kinane, Stathopoulou & Papapanou, 2017; Escobar et al., 2018). This is evidenced by its implications in systemic diseases such as atherosclerosis (Gotsman et al., 2007), aspiration pneumonia (Benedyk et al., 2016), cancer (Gao et al., 2016), rheumatoid arthritis (Laugisch et al., 2016), and diabetes mellitus (Khader et al., 2006). *Porphyromonas gingivalis* is an oral pathogen that is frequently associated with periodontitis and it is found to acquire Type IX Secretion System (T9SS); a bacterial secretion system that is unique to Bacteroidetes phylum (Sato et al., 2010).

T9SS exhibited diverse roles among species of bacteria under Bacteroidetes phylum. Other than transporting virulence factors such as gingipains and peptidylarginine deiminase in *P. gingivalis* that can cause human oral diseases (Potempa, Pike & Travis, 1995; Maresz et al., 2013), T9SS also transporting virulence factors such as chondroitin sulfate lyases that are associated with fish diseases such as columnaris disease. *Flavobacterium columnare*, a fish pathogen that contributes to the epidemic that occurred among wild and cultured fishes, is found to acquire T9SS. This epidemic poses a problem to the aquaculture industry as columnaris disease caused by the fish pathogen can significantly increase the mortality rate among cultured fishes and thus threatened the industry output (Li et al., 2017). T9SS also involved in the transport of non-virulence factors such as cargo proteins that formed the bacterial gliding motility apparatus in *Flavobacterium johnsoniae* that aids in its motility (Nakane et al., 2013) and enzymes that are importance for lignocellulose digestion in the rumen of ruminants that become the hosts for the *Candidatus Paraporphyromonas polyenzymogenes* (Naas et al., 2018).

Similar to other types of secretion system in bacteria, T9SS is consisted of many different protein components that perform coordinated roles to ensure proper translocation and modification of its cargo proteins. These roles can be categorised into four major functions: translocation, modification, energetic, and regulation (Sato et al., 2010; Lasica et al., 2017). There are six protein components that performed the translocation function. Four of them (PorL, PorM, PorK, and PorN) formed the trans-envelope complex ($PorK_2L_3M_2N_2$) that acts as the main translocation channel that transports cargo proteins from the periplasm to extracellular milieu (Vincent et al., 2017) (Fig. 1). Despite having a translocation channel that spanned both the OM and IM, the presence of N-terminal signal peptide in cargo proteins indicates that T9SS depends on SecYEG to translocate its cargo proteins from the cytoplasm to periplasm which is similar to Type II Secretion System (T2SS) (Overbye, Sandkvist & Bagdasarian, 1993; Costa et al., 2015). During translocation across IM, the signal peptide of cargo proteins is cleaved by type I signal peptidase before they are released into the periplasm (Rahman et al., 2003) (Fig. 1). The cargo proteins of T9SS also acquired C-terminal domain (CTD) that will interact sequentially with the D2, D3, and D4 domains of PorM and then PorN of the trans-envelope complex to be directed to the T9SS secretion pore in OM (Vincent, Chabalier & Cascales, 2018) (Fig. 1). It is also suggested that PG1058 aids the transport of cargo proteins across the periplasm (Heath et al., 2016) (Fig. 1). Recently, it is proposed that SprA (ortholog of Sov in *F. johnsoniae*) is the secretion pore that translocates cargo proteins across OM. SprA is a 36-stranded transmembrane

β-barrel that has a bottom opening that faced the periplasm and a lateral opening that faced the cell surface and it is large enough to transport folded proteins (Lauber et al., 2018) (Fig. 1). SprA is found to interact with PorV, Plug, and peptidyl-prolyl *cis-trans* isomerase (PPI) proteins. When the pore of SprA is vacant, PorV that acts an outer membrane shuttle protein that delivers the cargo proteins to the attachment complex (Glew et al., 2017) is found to interact with SprA at its lateral opening. As a cargo protein moves into the pore, the CTD of cargo protein will interact with PorV and the cargo protein-PorV complex will leave the SprA while Plug immediately interacts with the bottom opening of SprA (Lauber et al., 2018) (Fig. 1). This alternating interactions of PorV and Plug with SprA will ensure that only cargo proteins of T9SS can pass through the SprA pore. The specific function of interaction between SprA and PPI has not been found. However, it is found that deletion of both PPI and Plug does not affect the cargo proteins transport thus they are not crucial for T9SS function (Lauber et al., 2018). There are four protein components (PorU, PorV, PorZ, and PorQ) that performed the modification function. Those proteins formed a complex in OM called the attachment complex (Fig. 1). As the cargo protein-PorV complex binds attachment complex, PorU will cleave the CTD of cargo protein (Glew et al., 2012, 2017). Then, the cargo protein will be glycosylated with anionic lipopolysaccharide (A-LPS) delivered by PorZ at the cleaved site. After both post-translational modifications (CTD cleavage and A-LPS glycosylation), the cargo protein will be anchored to bacterial cell surface by A-LPS (Lasica et al., 2016; Glew et al., 2017) (Fig. 1).

The translocation of cargo proteins by T9SS might be energised by PorL and PorM that formed an energy transducer complex (Vincent et al., 2017). It is suggested that the energy transducer complex converts proton-motive force into mechanical energy that can provide the energy for cargo proteins translocation across T9SS. It has been reported that PorL and PorM bear signatures typical of an energy transducer employing proton-motive force (Vincent et al., 2017). The processes by T9SS are also regulated by its six protein components that is probably to ensure the virulence factors are secreted by bacteria when it is in a favourable environment. PorX and PorY formed a two-component system (TCS) that can regulate the operon of *por* genes (*porP*, *porK*, *porL*, *porM*, and *porN*) (Vincent et al., 2017). However, PorX that acts as a response regulator protein in TCS cannot directly bind the promoter of *por* genes but depends on PG0162 to form a complex that can bind the promoter (Kadowaki et al., 2016) (Fig. 1). PorX is also found to interact with PorL in which this interaction might serve a regulatory purpose (Vincent, Durand & Cascales, 2016). PorR is an aminotransferase that involves in the Wbp pathway that contributes to biosynthesis of the structural sugar of A-LPS (Shoji et al., 2014, 2018) (Fig. 1). PGN0300 can influence the present of PorU on cell surface however the exact mechanisms behind it remained vague (Taguchi et al., 2016).

Despite that, there are experimentally identified T9SS protein components without known functions that are illustrated as white coloured components in Fig. 1 (Nguyen et al., 2009; Saiki & Konishi, 2010; Sato et al., 2010; Gorasia et al., 2016). However, three out of five protein components (PorW, PorP, and PorT) have their functions predicted through motif identification and comparison with protein motifs from Pfam (Emrizal & Muhammad, 2018). Thus, we

performed phylogenetic analysis on those 20 protein components to investigate the phylogenetic relationship and taxonomic distribution among the putative members of T9SS component protein families. We found maximum-likelihood (ML) trees for 19 protein components exhibit clustering of terminal nodes based on their respective classes under Bacteroidetes phylum. Majority of those ML trees exhibit the phylogenetic relationship that is consistent with phylogeny of bacteria under Bacteroidetes phylum based on 16S rRNA sequence (Karlsson et al., 2011). The ML tree of PorR exhibits different clustering pattern of terminal nodes where the nodes do not cluster based on their respective classes. Hence, PorR might evolve independently from the other T9SS protein components which suggest that PorR might be acquired by T9SS-acquiring bacteria through horizontal gene transfer (Plyro et al., 2012). We found that seven genes (including *porR*) that involved in the biosynthesis of A-LPS are flanked by insertion sequences (ISs) that encode IS5 family transposase (Naito et al., 2008). This might suggest the possibility that the intervening DNA segment that contains *porR* gene can be mobilised and possibly subjected to conjugative transfer (Thomas & Nielsen, 2005; Brochet et al., 2009). The taxonomic distribution of T9SS protein components revealed that they can be found across all classes under Bacteroidetes phylum. We also identified additional species under Chitinophagia, Saprospiria, and unclassified that acquired homologs of T9SS protein components (McBride & Zhu, 2013).

## Materials & Methods

**Construction of multiple sequence alignments for protein components of T9SS**

The multiple sequence alignments for the phylogeny reconstruction for each Type IX Secretion System (T9SS) protein component were built using the putative members of T9SS component protein families. The pipeline that was used to select those members has been reported (Emrizal & Muhammad, 2018). The pipeline was used to filter out false positives among the homologs of T9SS protein components that have been identified through homology searching using BLASTP where the protein sequences of T9SS components retrieved from NCBI protein database were searched against the local BLAST database constructed from completely sequenced bacterial proteomes from Genbank. The selection criteria used in the pipeline: e-value $\leq 0.001$, query coverage > 60%, and homolog with the lowest e-value for bacterial strains with multiple hits; can minimise the possibility of false positives inclusion (Emrizal & Muhammad, 2018).

The list of protein homologs, which are the selected putative members for each T9SS component protein family, used to build the multiple sequence alignments for each T9SS protein component was provided as a Supplemental Information (Data S1). The multiple sequence alignments were constructed using T-Coffee version (11.00) (Notredame, Higgins & Heringa, 2000). The multiple sequence alignments in clustal format constructed by T-Coffee were converted into phylip relaxed (interleaved) format using online Format Converter (https://www.hiv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html). The multiple sequence alignments in phylip relaxed (interleaved) format were manually formatted by

removing the 'i' symbol in the first line and deleting the second line of the multiple sequence alignments to make them suitable to be used with PhyML (Data S2).

**Phylogeny reconstruction for protein components of T9SS**

Maximum-likelihood (ML) analysis was performed using those multiple sequence alignments to construct the phylogenetic trees for each T9SS protein component using PhyML version (3.3) under selected amino acid substitution models (Guindon et al., 2010). The node reliability was assessed using the bootstrap algorithm with 100 replicates (Pattengale et al., 2010). Prior to the ML analysis, the amino acid substitution models to be used were determined using ProtTest version (3.4.2) (Abascal, Zardoya & Posada, 2005) under the Bayesian Information Criterion (BIC) using default parameters (Schwarz, 1978). The constructed phylogenetic trees were visualised and annotated using online iTOL (Letunic & Bork, 2016).

**Identification of *porR* and its neighbouring genes arrangement in *P. gingivalis ATCC 33277* genome**

The sequence of *P. gingivalis ATCC 33277* genome and annotation files of the genome were retrieved from Genbank (Naito et al., 2008). The *P. gingivalis ATCC 33277* genome sequence and its annotation files were provided in the Supplemental Information (Data S3). The part of *P. gingivalis ATCC 33277* genome sequence that contains the *porR* and its neighbouring genes was extracted. Then, it was searched against the non-redundant protein sequences (nr) database using online blastx. The search was narrowed down to the proteome of *P. gingivalis ATCC 33277* only. The maximum target sequences was set at highest value available which is 20,000. Other parameters were left at its default values (Altschul et al., 1990). Only the matches with 100% percentage identity and 0 e-value were used to annotate the part of *P. gingivalis ATCC 33277* genome sequence that contains *porR* gene.

## Results

**Phylogenetic trees of protein components of T9SS**

The maximum-likelihood (ML) phylogenetic trees are constructed for each T9SS protein component using the putative members of T9SS component protein families (Emrizal & Muhammad, 2018). The constructed ML trees are rooted using midpoint rooting method (Hess & De Moraes Russo, 2007). The outgroup rooting method is not chosen as no distant homolog, which is homolog from bacterium outside Bacteroidetes phylum, is identified through BLASTP for protein component PorN (Singh et al., 2017; Emrizal & Muhammad, 2018) while unroot method is not chosen because the method is more suitable to establish conservancy and diversity among a set of sequences rather than phylogenetic relationship between them (Du et al., 2018).

We found that the terminal nodes of ML trees for 19 protein components are clustered based on their classes under Bacteroidetes phylum. Three distinct phylogenetic relationships between the three major classes under Bacteroidetes phylum are observed. Major classes are

those with more than five families under them (Bacteroidia, Cytophagia, and Flavobacteriia) while minor classes are those with less than or equal to five families under them (Chitinophagia, Sphingobacteriia, Saprospiria, and unclassified). The phylogenetic relationship between three major classes for 10 protein components is consistently recovered as (Bacteroidia + Flavobacteriia) + Cytophagia. This relationship implied that Bacteroidia is genetically closer to Flavobacteriia than either of them to Cytophagia. The ML trees for 10 protein components that exhibit this relationship are shown in Fig. 2 and Fig. 3. The phylogenetic relationship between three major classes for five protein components is consistently recovered as (Cytophagia + Flavobacteriia) + Bacteroidia. This relationship implied that Cytophagia is genetically closer to Flavobacteriia than either of them to Bacteroidia. The ML trees for five protein components that exhibit this relationship are shown in Fig. 4. The phylogenetic relationship between three major classes for four protein components is consistently recovered as (Bacteroidia + Cytophagia) + Flavobacteriia. This relationship implied that Bacteroidia is genetically closer to Cytophagia than either of them to Flavobacteriia. The ML trees for four protein components that exhibit this relationship are shown in Fig. 5.

However, only the terminal nodes of ML tree of PorR are not clustered based on their classes as shown in Fig. 6. Seven classes under Bacteroidetes phylum (Bacteroidia, Cytophagia, Flavobacteriia, Chitinophagia, Sphingobacteriia, Saprospiria, and unclassified which is for bacterial strains that do not exhibit morphological or sequence similarities with existing classes under Bacteroidetes) are identified to acquire PorR. These classes are clustered into three monophyletic groups. We found that terminal nodes of Flavobacteriia, Saprospiria, and Chitinophagia are clustered into monophyletic group three while terminal nodes of Bacteroidia, Cytophagia, Sphingobacteriia, and unclassified spread out between different monophyletic groups. Thus, unlike other protein components, the ML tree of PorR exhibits a different clustering pattern as its terminal nodes do not cluster based on their classes.

**Arrangement of *porR* and its neighbouring genes in *P. gingivalis* ATCC 33277 genome**

As shown in Fig. 7, *porR* and its neighbouring genes are flanked by IS5 transposons. Both of the IS5 transposons encode for IS5 family transposase that will cleave the flanking 12 bp inverted repeats. This might suggest the possibility that the intervening DNA segment that contains seven genes that involved in A-LPS biosynthesis that are represented by yellow coloured rectangles in Fig. 7 can be mobilised and possibly subjected to conjugative transfer (Thomas & Nielsen, 2005; Brochet et al., 2009). *porR* (PGN_1236) and *ugdA* (PGN_1243) genes have been reported to be involved in Wbp pathway that is important for the biosynthesis of structural sugar (di-acetylated glucuronic acid) of A-LPS (Shoji et al., 2014). *porS* O-antigen flippase and *wzy* O-antigen polymerase genes have been reported to participate in the assembly of A-LPS in bacterial inner membrane (Shoji et al., 2013). *gtfB* and *gtfE* glycosyltransferases genes are important for A-LPS biosynthesis while *rfa* glycosyltransferase gene is important for biosynthesis of lipid A-core portion of A-LPS (Shoji et al., 2018).

**Taxonomic distributions of protein components of T9SS**

As shown in the ML trees of T9SS protein components (Figs. 2-6), only bacteria under Bacteroidia, Flavobacteriia, and Chitinophagia classes acquired the 20 protein components investigated. The bacteria under Cytophagia class acquired only 19 protein components except for PorN. The bacteria under Saprospiria class acquired only 18 protein components except for PorL and PG0189. The bacteria under unclassified acquired only 18 protein components except for PorN and PG0189. The bacteria under Sphingobacteriia class acquired only 17 protein components except for PorQ, PorU, and PorZ.

Our finding is consistent with the taxonomic distribution of T9SS protein components among bacteria under Bacteroidetes phylum where it has been reported that Bacteroidia, Flavobacteriia, Cytophagia, Sphingobacteriia, and Incertae sedis classes acquired those homologs (McBride & Zhu, 2013). It is important to note that the bacterial strains under Incertae sedis (*Rhodothermus marinus DSM 4252* and *Salinibacter ruber DSM 13855*) reported by the literature to acquire homologs of T9SS protein components (McBride & Zhu, 2013) are under unclassified in our research as we follow the classification based on NCBI taxonomy database as of March 2017. However, comparing the reported taxonomic distribution of T9SS protein components to ours, we have identified additional species under Chitinophagia, Saprospiria, and unclassified that acquired homologs of T9SS protein components. Those additional species and the homologs of T9SS protein components they acquired are illustrated in Fig. 8.

## Discussion

The maximum-likelihood (ML) phylogenetic trees of 19 protein components of T9SS exhibit the clustering of terminal nodes based on their classes (Figs. 2-5). To identify the similarities in phylogenetic relationships between those 19 ML trees, we have decided to exclude the minor classes (Chitinophagia, Sphingobacteriia, Saprospiria, and unclassified). The reason behind this is because terminal nodes that represented the minor classes tend to be inconsistent in its positions in those ML trees. Thus, it causes difficulty to draw out similarities between them. This might arise due to insufficient taxa from these classes are provided to construct the phylogenetic trees. Thus, less information is provided which is insufficient to properly resolved the phylogeny of these classes. As more T9SS-acquiring bacterial strains from these classes are sequenced, probably the phylogenies of T9SS protein components will be more resolved (Alvizu et al., 2018).

Three distinct phylogenetic relationships between the three major classes (Bacteroidia, Cytophagia, and Flavobacteriia) are identified: (Bacteroidia + Flavobacteriia) + Cytophagia, (Cytophagia + Flavobacteriia) + Bacteroidia, and (Bacteroidia + Cytophagia) + Flavobacteriia. The ML trees of 10 protein components (PorK, PorM, PorT, PorN, PorZ, PG1058, PorU, Sov, PorX, and PorV) exhibit the phylogenetic relationship (Bacteroidia + Flavobacteriia) + Cytophagia that is consistent with the phylogeny of bacteria under Bacteroidetes phylum based on 16S rRNA sequence that has been reported (Karlsson et al., 2011). The phylogenetic

relationship that is recovered by the reported study is (Bacteroidia + Flavobacteriia) + (Rhodothermaceae + (Chitinophagia + (Sphingobacteriia + Cytophagia))) (Karlsson et al., 2011). This relationship suggests that Bacteroidia is genetically close to Flavobacteria than either of them to Cytophagia which is similar to the relationship exhibited by the ML trees of those 10 protein components. The ML trees of five protein components (PorL, PorQ, PG0162, PG0534, and PGN0300) exhibit the phylogenetic relationship (Cytophagia + Flavobacteriia) + Bacteroidia while the ML trees of four protein components (PorP, PorW, PorY, and PG1058) exhibit the phylogenetic relationship (Bacteroidia + Cytophagia) + Flavobacteriia. In those nine ML trees, there are low bootstrap support values (less than 70) at the nodes leading to the clusters of major classes that are pointed by the black arrows in Fig. 4 and Fig. 5. Hence, there is a low support for the phylogenetic relationship exhibited by those nine ML trees deviates from the phylogeny of bacteria under Bacteroidetes phylum based on 16S rRNA sequence that has been reported (Karlsson et al., 2011).

Only the ML tree of PorR has terminal nodes that are not clustered based on their classes as shown in Fig. 6. It has been shown before that the phylogeny based on a single gene or protein might deviate from the phylogeny based on 16S rRNA sequence (Jeffroy et al., 2006). However, the presence of high bootstrap support values at nodes that are pointed by black arrows in Fig. 6 leading to clusters of terminal nodes that are not grouped based on their classes might suggest there is a strong support for the phylogenetic relationship exhibited by ML tree of PorR deviates from the phylogeny based on 16S rRNA sequence (Karlsson et al., 2011). Thus, there is a possibility that *porR* gene is subjected to horizontal transfer among those bacteria hence causing deviation from the expected phylogeny (Pylro et al., 2012). Hirt, Schlievert & Dunny have demonstrated that virulence factor and antibiotic resistance genes could be horizontally transferred (Hirt, Schlievert & Dunny, 2002). This adds up to the possibility that *porR* gene that encodes one of the virulence factors produced by *P. gingivalis* can be horizontally transferred (Shoji et al., 2014).

We have looked at the sequence of *P. gingivalis strain ATCC 33277* genome. *P. gingivalis strain ATCC 33277* genome is chosen because many gene orthologs that involved in A-LPS biosynthesis have been identified in this genome (Shoji et al., 2018). We found that *porR* and its neighbouring genes are flanked by IS5 transposons. The IS5 transposons contain a single open reading frame that encodes for IS5 family transposase that will cleave the 12 bp inverted repeats that flanked the transposons (Fig. 7). The 12 bp inverted repeats that are represented by purple coloured triangles in Fig. 7 show imperfect homology to each other with the consensus sequence: GAGACCTTTG[CA]A. Both of the IS5 transposons are ~ 1300 bp in length. These features are typical of IS5 family transposons (Mahillon & Chandler, 1998; Naito et al., 2008). The intervening DNA segment and both IS5 transposons that flanked it might form a composite transposon where the cleaving action of IS5 transposases on inverted repeats can mobilise the invervening DNA segment that contains porR gene and possibly subjected it to conjugative transfer (Thomas & Nielsen, 2005; Brochet et al., 2009). The length of the composite transposon is ~ 70 kbp. However, it is also possible for IS5 transposase to cleave the inverted repeat directly

downstream of PGN_1255 (Fig. 7) which will reduce the length of composite transposon to ~ 47 kbp. Brochet et al. have reported a transposon of ~ 47 kbp in length that performed both transposition and conjugative transfer processes (Brochet et al., 2009). Hence, it might be possible for composite transposon of such length to undergo transposition and subsequently being horizontally transferred via bacterial conjugation.

The intervening DNA segment contains seven genes that involved in the biosynthesis of A-LPS (Fig. 7). Both *porR* and *ugdA* genes are involved in Wbp pathway that is important for biosynthesis of di-acetylated glucuronic acid that is the structural sugar of A-LPS (Shoji et al., 2014). *porS*, which is an O-antigen flippase similar to wzx, and *wzy*, which is an O-antigen polymerase, genes are involved in the assembly of A-LPS at the periplasmic side of bacterial inner membrane (Shoji et al., 2013). *gtfB* and *gtfE* glycosyltransferases genes are involved in the biosynthesis of sugar moiety of A-LPS. *rfa* glycosyltransferase gene is involved in the biosynthesis of lipid A-core moiety of A-LPS (Shoji et al., 2018). However, there are many other genes that involved in the biosynthesis of A-LPS and they are spread out throughout the genome (Shoji et al., 2018). This raise a question of why these genes are not clustered in a single operon which is usually the case for genes that involved in a similar pathway. It might be possible that only those seven genes identified in this study are horizontally transferred while the others are not. Thus, explaining why only those seven genes clustered in a composite transposon. Perhaps, those other genes also are flanked by mobile elements that aid in their transposition and conjugative transfer. Further analyses such as comparing the molecular phylogeny of all genes involved in the biosynthesis of A-LPS and comparing the arrangement of those genes in the bacterial genomes that synthesis A-LPS might be able to answer those questions.

T9SS is made up of various protein components that formed the regulation, translocation, energetic, and modification components. Currently, the secretion system is primarily found in bacteria under Bacteroidetes phylum (Abby et al., 2016). We found bacteria from Bacteroidia, Flavobacteriia, Cytophagia, Chitinophagia, Sphingobacteriia, Saprospiria, and unclassified under Bacteroidetes phylum that might acquire T9SS protein components. However, not all bacterial strains that we have identified acquired all 20 protein components reported by literatures (Sato et al., 2010; Lasica et al., 2017).

As shown in the ML trees of T9SS protein components (Figs. 2-6), only bacteria under Bacteroidia, Flavobacteriia, and Chitinophagia classes acquired the 20 protein components investigated. The bacteria under Cytophagia class only acquired 19 protein components excluding PorN. The bacteria under Saprospiria class only acquired 18 protein components excluding PorL and PG0189. The bacteria under unclassified only acquired 18 protein components with the exception of PorN and PG0189. The bacteria under Sphingobacteriia class only acquired 17 protein components with the exception of PorQ, PorU, and PorZ. It is interesting to note that PorU, PorZ, and PorQ formed the modification components of T9SS. Thus, Sphingobacteriia does not acquire the protein components that performed modifications such as cleavage of C-terminal domain (CTD) and A-LPS glycosylation on T9SS cargo proteins. Perhaps, the T9SS acquired by Sphingobacteriia does not cleave the CTD of cargo protein and

glycosylate it with A-LPS but leave the cargo protein bounded to PorV after it is translocated to bacterial cell surface by Sov. Other possible explanation is that Sphingobacteriia does have proteins that performed the functions of missing protein components, but those proteins exhibited limited sequence similarity with any currently known T9SS protein component thus they could not be detected by homology searching method. This explanation could also be the case for the other classes or strains of bacteria that do not acquire the homologs of 20 protein components of T9SS.

We have found additional species under Chitinophagia, Saprospiria, and unclassified that acquired homologs of T9SS protein components that, to our knowledge, might not have been reported (McBride & Zhu, 2013). Those additional species and the homologs of T9SS protein components they acquired are illustrated in Fig. 8. This identification might be due to our analysis covered larger bacterial species because of more bacterial genomes have been completely sequenced in the past few years.

## Conclusions

We found that the maximum-likelihood (ML) phylogenetic trees for 19 protein components of T9SS exhibit clustering of terminal nodes based on their respective classes under Bacteroidetes phylum. Majority of the ML trees of those protein components exhibit the phylogenetic relationship that is consistent with phylogeny of bacteria under Bacteroidetes phylum based on 16S rRNA sequence (Karlsson et al., 2011). The ML tree of PorR, which is an aminotransferase that involved in the Wbp pathway that produces structural sugar of A-LPS, exhibits different clustering pattern of terminal nodes where the nodes do not cluster based on their respective classes. Hence, PorR might evolve independently from the other protein components that might suggest PorR is acquired by bacteria with T9SS through horizontal gene transfer. We found that seven genes (including *porR*) that involved in the biosynthesis of A-LPS that anchored the virulence factor secreted by T9SS to bacterial cell surface are flanked by insertion sequences (ISs) that encode IS5 family transposase. This might suggest that the intervening DNA segment that contains *porR* gene can be mobilized and subsequently contributes to the possibility that *porR* gene is subjected to conjugative transfer.

We found that the 20 protein components of T9SS investigated in this study are not necessarily present in all T9SS-acquiring bacteria under Bacteroidetes phylum. We have identified species under Chitinophagia, Saprospiria, and unclassified that acquired homologs of T9SS protein components that, to our knowledge, might not have been reported (McBride & Zhu, 2013).

## Acknowledgements

The authors acknowledged Universiti Kebangsaan Malaysia (UKM) grant, GGPM-2016-016 for providing funds to support the research. We thank Centre for Bioinformatics Research (CBR) for

providing the facilities to conduct the bioinformatics analyses. We thank the anonymous reviewers for their comments on previous drafts of the manuscript.

## References

**Abascal F, Zardoya R, Posada D**. **2005**. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**(**9**):2104–2105 DOI 10.1093/bioinformatics/bti263.

**Abby SS, Cury J, Guglielmini J, Néron B, Touchon M, Rocha EPC**. **2016**. Identification of protein secretion systems in bacterial genomes. *Scientific reports* **6**:23080 DOI 10.1038/srep23080.

**Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ**. **1990**. Basic local alignment search tool. *Journal of molecular biology* **215**(**3**):403–410 DOI 10.1016/S0022-2836(05)80360-2.

**Alvizu A, Eilertsen MH, Xavier JR, Rapp HT**. **2018**. Increased taxon sampling provides new insights into the phylogeny and evolution of the subclass Calcaronea (Porifera, Calcarea). *Organisms Diversity and Evolution* **18**(**3**):279–290 DOI 10.1007/s13127-018-0368-4.

**Benedyk M, Mydel PM, Delaleu N, Płaza K, Gawron K, Milewska A, Maresz K, Koziel J, Pyrc K, Potempa J**. **2016**. Gingipains: critical factors in the development of aspiration pneumonia caused by Porphyromonas gingivalis. *Journal of innate immunity* **8**(**2**):185–198 DOI 10.1159/000441724.

**Brochet M, Da Cunha V, Couvé E, Rusniok C, Trieu-Cuot P, Glaser P**. **2009**. Atypical association of DDE transposition with conjugation specifies a new family of mobile elements. *Molecular Microbiology* **71(4)**:948–959 DOI 10.1111/j.1365-2958.2008.06579.x.

**Costa TRD, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A, Trokter M, Waksman G**. **2015**. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology* **13**(**6**):343–359 DOI 10.1038/nrmicro3456.

**Du K, Zhong Z, Fang C, Dai W, Shen Y, Gan X, He S**. **2018**. Ancient duplications and functional divergence in the interferon regulatory factors of vertebrates provide insights into the evolution of vertebrate immune systems. *Developmental & Comparative Immunology* **81**:324–333 DOI 10.1016/j.dci.2017.12.016.

**Emrizal R, Muhammad NAN**. **2018**. Identification of sequence motifs for the protein components of type IX secretion system. *Sains Malaysiana* **47**(**12**):2941–2950 DOI 10.17576/jsm-2018-4712-02.

**Escobar GF, Abdalla DR, Beghini M, Gotti VB, Junior VR, Napimoga MH, Ribeiro BM, Rodrigues DBR, Nogueira RD, de Lima Pereira SA**. **2018**. Levels of Pro and Anti-inflammatory Citokynes and C-Reactive Protein in Patients with Chronic Periodontitis Submitted to Nonsurgical Periodontal Treatment. *Asian Pacific journal of cancer prevention : APJCP* **19**(**7**):1927–1933 DOI 10.22034/APJCP.2018.19.7.1927.

**Gao S, Li S, Ma Z, Liang S, Shan T, Zhang M, Zhu X, Zhang P, Liu G, Zhou F, Yuan X, Jia R, Potempa J, Scott DA, Lamont RJ, Wang H, Feng X**. **2016**. Presence of Porphyromonas gingivalis in esophagus and its association with the clinicopathological characteristics and survival in patients with esophageal cancer. *Infectious agents and cancer* **11**(**1**):3 DOI 10.1186/s13027-016-0049-x.

**Glew MD, Veith PD, Chen D, Gorasia DG, Peng B, Reynolds EC**. **2017**. PorV is an outer membrane shuttle protein for the type IX secretion system. *Scientific reports* **7**(**1**):8790 DOI 10.1038/s41598-017-09412-w.

**Glew MD, Veith PD, Peng B, Chen YY, Gorasia DG, Yang Q, Slakeski N, Chen D, Moore C, Crawford S, Reynolds EC**. **2012**. PG0026 is the C-terminal signal peptidase of a novel secretion system of Porphyromonas gingivalis. *Journal of Biological Chemistry* **287**(**29**):24605–24617 DOI 10.1074/jbc.M112.369223.

**Gorasia DG, Veith PD, Hanssen EG, Glew MD, Sato K, Yukitake H, Nakayama K, Reynolds EC**. **2016**. Structural insights into the PorK and PorN components of the Porphyromonas gingivalis type IX secretion system. *PLoS pathogens* **12**(**8**):e1005820 DOI 10.1371/journal.ppat.1005820.

**Gotsman I, Lotan C, Soskolne WA, Rassovsky S, Pugatsch T, Lapidus L, Novikov Y, Masrawa S, Stabholz A**. **2007**. Periodontal destruction is associated with coronary artery disease and periodontal infection with acute coronary syndrome. *Journal of periodontology* **78**(**5**):849–858 DOI 10.1902/jop.2007.060301.

**Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O**. **2010**. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**(**3**):307–321 DOI 10.1093/sysbio/syq010.

**Heath JE, Seers CA, Veith PD, Butler CA, Muhammad NAN, Chen YY, Slakeski N, Peng B, Zhang L, Dashper SG, Cross KJ, Cleal SM, Moore C, Reynolds EC**. **2016**. PG1058 is a novel multidomain protein component of the bacterial type IX secretion system. *PloS one* **11**(**10**):e0164313 DOI 10.1371/journal.pone.0164313.

**Hess PN, De Moraes Russo CA**. **2007**. An empirical test of the midpoint rooting method. *Biological Journal of the Linnean Society* **92**(**4**):669–674 DOI 10.1111/j.1095-8312.2007.00864.x.

**Hirt H, Schlievert PM, Dunny GM**. **2002**. In vivo induction of virulence and antibiotic resistance transfer in Enterococcus faecalis mediated by the sex pheromone-sensing system of pCF10. *Infection and immunity* **70**(**2**):716–723 DOI 10.1128/IAI.70.2.716.

**Jeffroy O, Brinkmann H, Delsuc F, Philippe H**. **2006**. Phylogenomics: the beginning of incongruence? *Trends in Genetics* **22**(**4**):225–231 DOI 10.1016/j.tig.2006.02.003.

**Kadowaki T, Yukitake H, Naito M, Sato K, Kikuchi Y, Kondo Y, Shoji M, Nakayama K**. **2016**. A two-component system regulates gene expression of the type IX secretion component proteins via an ECF sigma factor. *Scientific reports* **6**:23288 DOI 10.1038/srep23288.

**Karlsson FH, Ussery DW, Nielsen J, Nookaew I**. **2011**. A closer look at Bacteroides: phylogenetic relationship and genomic implications of a life in the human gut. *Microbial ecology* **61**(**3**):473–485 DOI 10.1007/s00248-010-9796-1.

**Khader YS, Dauod AS, El-Qaderi SS, Alkafajei A, Batayha WQ**. **2006**. Periodontal status of diabetics compared with nondiabetics: a meta-analysis. *Journal of diabetes and its complications* **20**(**1**):59–68 DOI 10.1016/j.jdiacomp.2005.05.006.

**Kinane DF, Stathopoulou PG, Papapanou PN**. **2017**. Periodontal diseases. *Nature Reviews Disease Primers* **3**:17038 DOI doi:10.1038/nrdp.2017.38.

**Lasica AM, Goulas T, Mizgalska D, Zhou X, De Diego I, Ksiazek M, Madej M, Guo Y, Guevara T, Nowak M, Potempa B, Goel A, Sztukowska M, Prabhakar AT, Bzowska M, Widziolek M, Thøgersen IB, Enghild JJ, Simonian M, Kulczyk AW, Nguyen KA, Potempa J, Gomis-Rüth FX**. **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 DOI 10.1038/srep37708.

**Lasica AM, Ksiazek M, Madej M, Potempa J**. **2017**. The type IX secretion system (T9SS): highlights and recent insights into its structure and function. *Frontiers in cellular and infection microbiology* **7**:215 DOI 10.3389/fcimb.2017.00215.

**Lauber F, Deme JC, Lea SM, Berks BC**. **2018**. Type 9 secretion system structures reveal a new protein transport mechanism. *Nature* **564**(**7734**):77 DOI 10.1038/s41586-018-0693-y.

**Laugisch O, Wong A, Sroka A, Kantyka T, Koziel J, Neuhaus K, Sculean A, Venables PJ, Potempa J, Möller B, Eick S**. **2016**. Citrullination in the periodontium—a possible link between periodontitis and rheumatoid arthritis. *Clinical oral investigations* **20**(**4**):675–683 DOI 10.1007/s00784-015-1556-7.

**Letunic I, Bork P**. **2016**. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research* **44**(**W1**):W242–W245 DOI 10.1093/nar/gkw290.

**Li N, Zhu Y, LaFrentz BR, Evenhuis JP, Hunnicutt DW, Conrad RA, Barbier P, Gullstrand CW, Roets JE, Powers JL, Kulkarni SS, Erbes DH, Garcia JC, Nie P, McBride MJ**. **2017**. The type IX secretion system is required for virulence of the fish pathogen Flavobacterium columnare. *Appl. Environ. Microbiol.* **83**(**23**):e01769-17 DOI 10.1128/aem.01769-17.

**Mahillon J, Chandler M**. **1998**. Insertion sequences. *Microbiol. Mol. Biol. Rev.* **62**(**3**):725–774

**Maresz KJ, Hellvard A, Sroka A, Adamowicz K, Bielecka E, Koziel J, Gawron K, Mizgalska D, Marcinska KA, Benedyk M, Pyrc K, Quirke AM, Jonsson R, Alzabin S, Venables PJ, Nguyen KA, Mydel P, Potempa J**. **2013**. Porphyromonas gingivalis facilitates the development and progression of destructive arthritis through its unique bacterial peptidylarginine deiminase (PAD). *PLoS pathogens* **9**(**9**):e1003627 DOI 10.1371/journal.ppat.1003627.

**McBride MJ, Zhu Y**. **2013**. Gliding motility and Por secretion system genes are widespread among members of the phylum Bacteroidetes. *Journal of bacteriology* **195**(**2**):270–278 DOI 10.1128/JB.01962-12.

**Naas AE, Solden LM, Norbeck AD, Brewer H, Hagen LH, Heggenes IM, McHardy AC, Mackie RI, Paša-Tolic L, Arntzen MØ, Eijsink VGH, Koropatkin NM, Hess M, Wrighton KC, Pope PB**. **2018**. "Candidatus Paraporphyromonas polyenzymogenes" encodes multi-modular cellulases linked to the type IX secretion system. *Microbiome* **6**(**1**):44 DOI 10.1186/s40168-018-0421-8.

**Naito M, Hirakawa H, Yamashita A, Ohara N, Shoji M, Yukitake H, Nakayama K, Toh H, Yoshimura F, Kuhara S, Hattori M**. **2008**. Determination of the genome sequence of Porphyromonas gingivalis strain ATCC 33277 and genomic comparison with strain W83 revealed extensive genome rearrangements in P. gingivalis. *DNA research* **15(4)**:215–225 DOI 10.1093/dnares/dsn013.

**Nakane D, Sato K, Wada H, McBride MJ, Nakayama K**. **2013**. Helical flow of surface protein required for bacterial gliding motility. *Proceedings of the National Academy of Sciences* **110**(**27**):11145–11150 DOI 10.1073/pnas.1219753110.

**Nguyen KA, Zylicz 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 (Reading, England)* **155**(**Pt 2**):328 DOI 10.1099/mic.0.024323-0.

**Notredame C, Higgins DG, Heringa J**. **2000**. T-coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology* **302**(**1**):205–217 DOI 10.1006/jmbi.2000.4042.

**Overbye LJ, Sandkvist M, Bagdasarian M**. **1993**. Genes required for extracellular secretion of enterotoxin are clustered in Vibrio cholerae. *Gene* **132**(**1**):101–106 DOI 10.1016/0378-1119(93)90520-D.

**Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A**. **2010**. How many bootstrap replicates are necessary? *Journal of computational biology* **17**(**3**):337–354

**Potempa J, Pike R, Travis J**. **1995**. The multiple forms of trypsin-like activity present in various strains of Porphyromonas gingivalis are due to the presence of either Arg-gingipain or Lys-gingipain. *Infection and immunity* **63**(**4**):1176–1182

**Pylro VS, Vespoli LDS, Duarte GF, Yotoko KSC**. **2012**. Detection of horizontal gene transfers from phylogenetic comparisons. *International journal of evolutionary biology* **2012** DOI 10.1155/2012/813015.

**Rahman MS, Simser JA, Macaluso KR, Azad AF**. **2003**. Molecular and functional analysis of the lepB gene, encoding a type I signal peptidase from Rickettsia rickettsii and Rickettsia typhi. *Journal of Bacteriology* **185**(**15**):4578–4584 DOI 10.1128/JB.185.15.4578-4584.2003.

**Saiki K, Konishi K**. **2010**. Identification of a novel Porphyromonas gingivalis outer membrane protein, PG534, required for the production of active gingipains. *FEMS microbiology letters* **310**(**2**):168–174 DOI 10.1111/j.1574-6968.2010.02059.x.

**Sato K, Naito M, Yukitake H, Hirakawa H, Shoji M, McBride MJ, Rhodes RG, Nakayama K**. **2010**. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proceedings of the National Academy of Sciences* **107**(**1**):276–281 DOI 10.1073/pnas.0912010107.

**Schwarz G**. **1978**. Estimating the dimension of a model. *The annals of statistics* **6**(**2**):461–464

**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 DOI 10.1111/omi.12200.

**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 DOI 10.1038/srep05056.

**Shoji M, Yukitake H, Sato K, Shibata Y, Naito M, Aduse-Opoku J, Abiko Y, Curtis MA, Nakayama K**. **2013**. Identification of an O-antigen chain length regulator, WzzP, in Porphyromonas gingivalis. *Microbiologyopen* **2**(**3**):383–401 DOI 10.1002/mbo3.84.

**Singh SP, Groeneveld JC, Al-Marzouqi A, Willows-Munro S**. **2017**. A molecular phylogeny of the spiny lobster Panulirus homarus highlights a separately evolving lineage from the Southwest Indian Ocean. *PeerJ* **5**:e3356 DOI 10.7717/peerj.3356.

**Taguchi Y, Sato K, Yukitake H, Inoue T, Nakayama M, Naito M, Kondo Y, Kano K, Hoshino T, Nakayama K, Takashiba S, Ohara N**. **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 DOI 10.1128/IAI.01308-15.

**Thomas CM, Nielsen KM**. **2005**. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature reviews microbiology* **3(9)**:711 DOI 10.1038/nrmicro1234.

**Vincent MS, Canestrari MJ, Leone P, Stathopulos J, Ize B, Zoued A, Cambillau C, Kellenberger C, Roussel A, Cascales E**. **2017**. Characterization of the Porphyromonas gingivalis type IX secretion trans-envelope PorKLMNP core complex. *Journal of Biological Chemistry* **292**(**8**):3252–3261 DOI 10.1074/jbc.M116.765081.

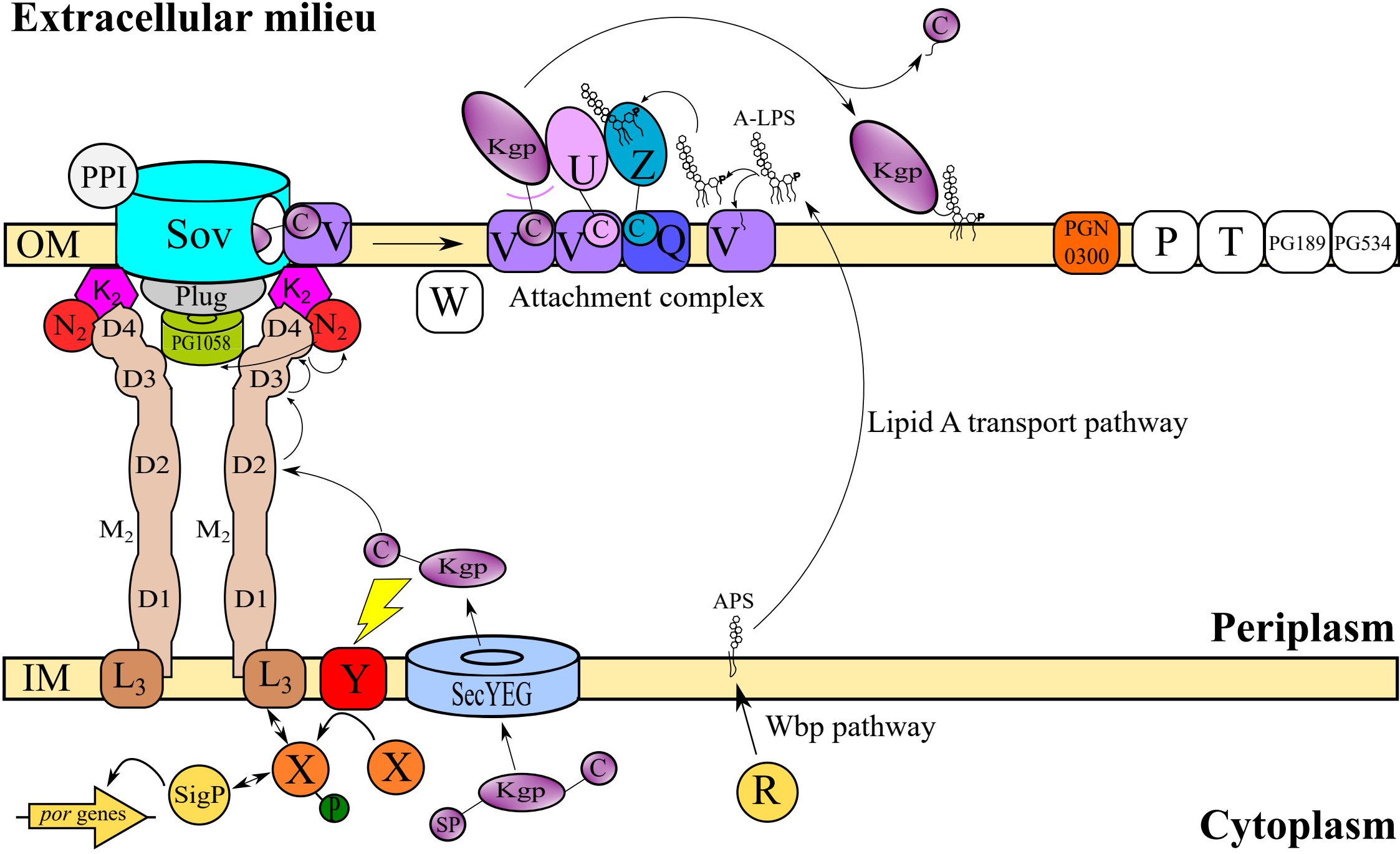
**Vincent MS, Chabalier M, Cascales E**. **2018**. A conserved motif of Porphyromonas Type IX secretion effectors C-terminal secretion signal specifies interactions with the PorKLMN core complex. *bioRxiv*:483123 DOI 10.1101/483123.

**Vincent MS, 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 DOI 10.3389/fcimb.2016.00096.

## Figure 1(on next page)

The T9SS protein components on the bacterial cell double membranes.

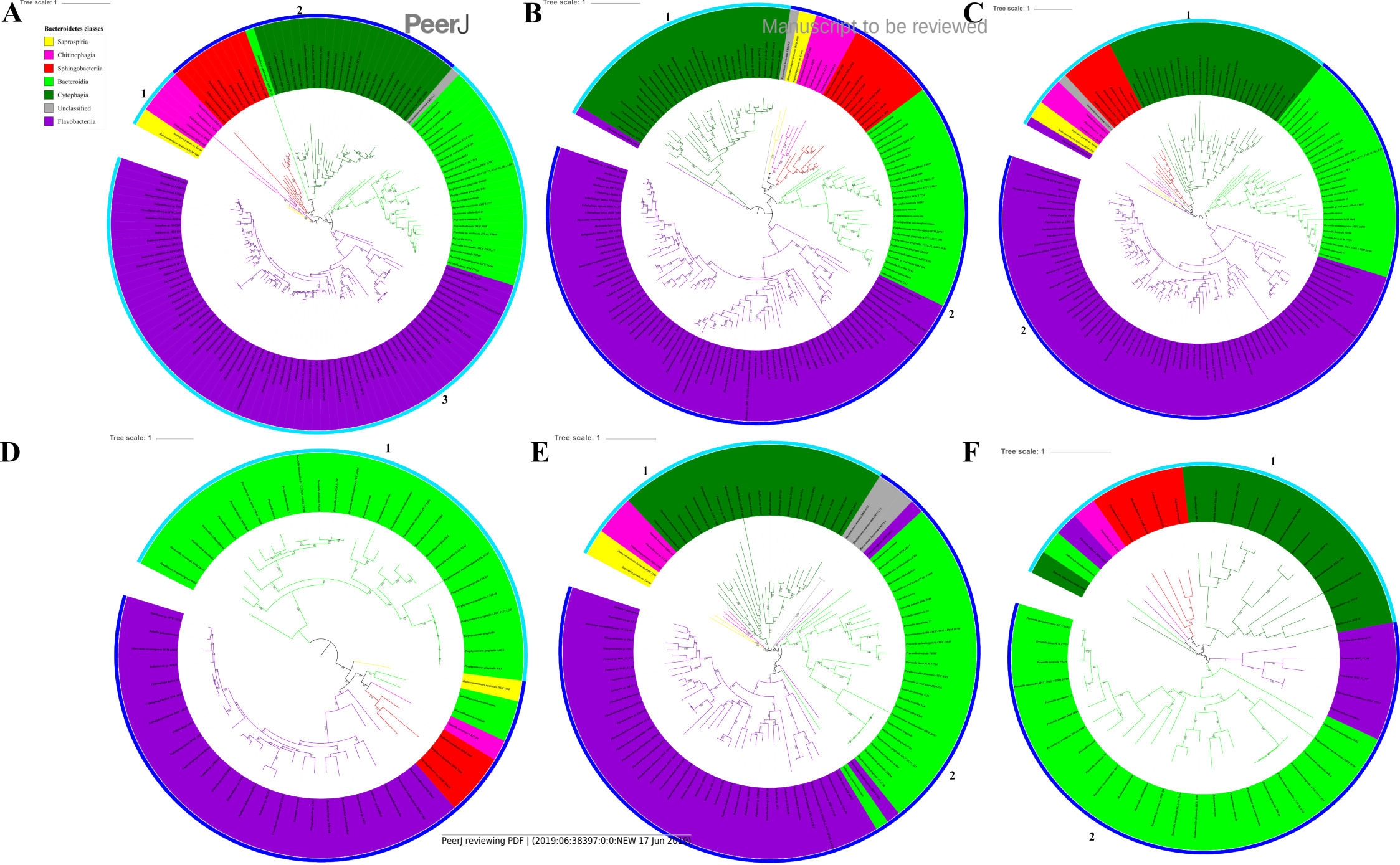
The protein components with known functions are coloured while those with unknown functions are coloured in white. The pathway for cargo protein lysine gingipain (Kgp) translocation and modifications is illustrated. How those processes performed by T9SS are regulated by its protein components are also exhibited.


Extracellular milieu
C
Kgp
U
Z
A-LPS
Kgp
PPI
Sov
C
V
OM
V
C
V
C
C
Q
V
PGN
0300
P
T
PG189
PG534
K2
K2
Plug
W
Attachment complex
N2
D4
PG1058
D4
N2
D3
D3
D2
D2
Lipid A transport pathway
M2
M2
C
Kgp
D1
D1
APS
Periplasm
IM
L3
L3
Y
SecYEG
Wbp pathway
X
X
R
SigP
por genes
P
Kgp
C
SP
Cytoplasm

## Figure 2(on next page)

The maximum-likelihood (ML) phylogenetic trees of six protein components of T9SS that exhibit phylogenetic relationship: (Bacteroidia + Flavobacteriia) + Cytophagia.

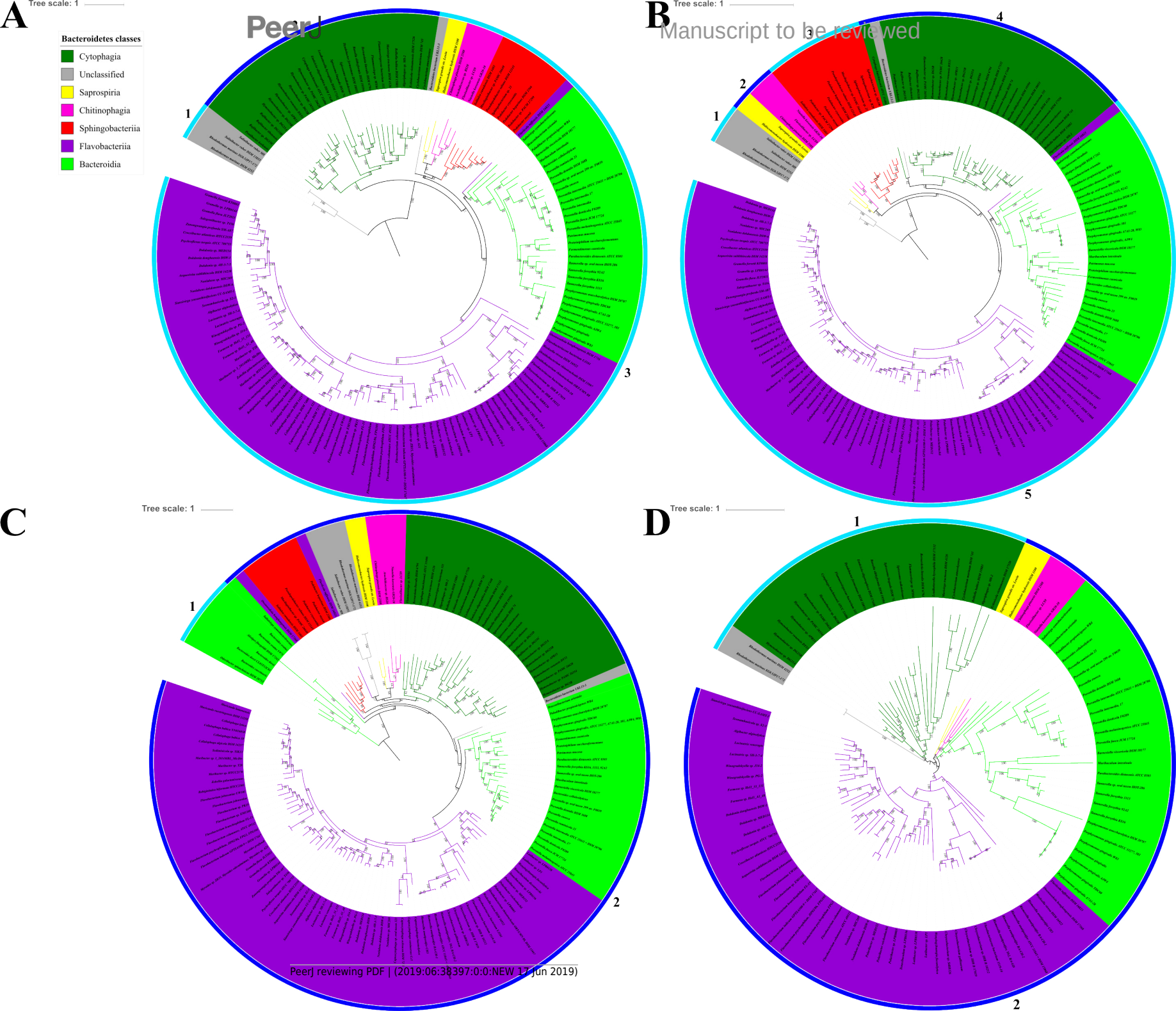
(A) ML tree of PorK. (B) ML tree of PorM. (C) ML tree of PorT. (D) ML tree of PorN. (E) ML tree of PorZ. (F) ML tree of PG0189. The terminal nodes are labelled with the bacterial strains that the homologs belong to and the colours denote the classes for each bacterial strain. A terminal node can represent multiple bacterial strains in the case where their homologs are completely identical and this terminal node can be identified by the present of commas separating the bacterial strain names. Bootstrap support values are indicated on each node. The outer strips and its corresponding numbers denote the monophyletic groups identified.



## Figure 3(on next page)

The maximum-likelihood (ML) phylogenetic trees of four protein components of T9SS that exhibit phylogenetic relationship: (Bacteroidia + Flavobacteriia) + Cytophagia.

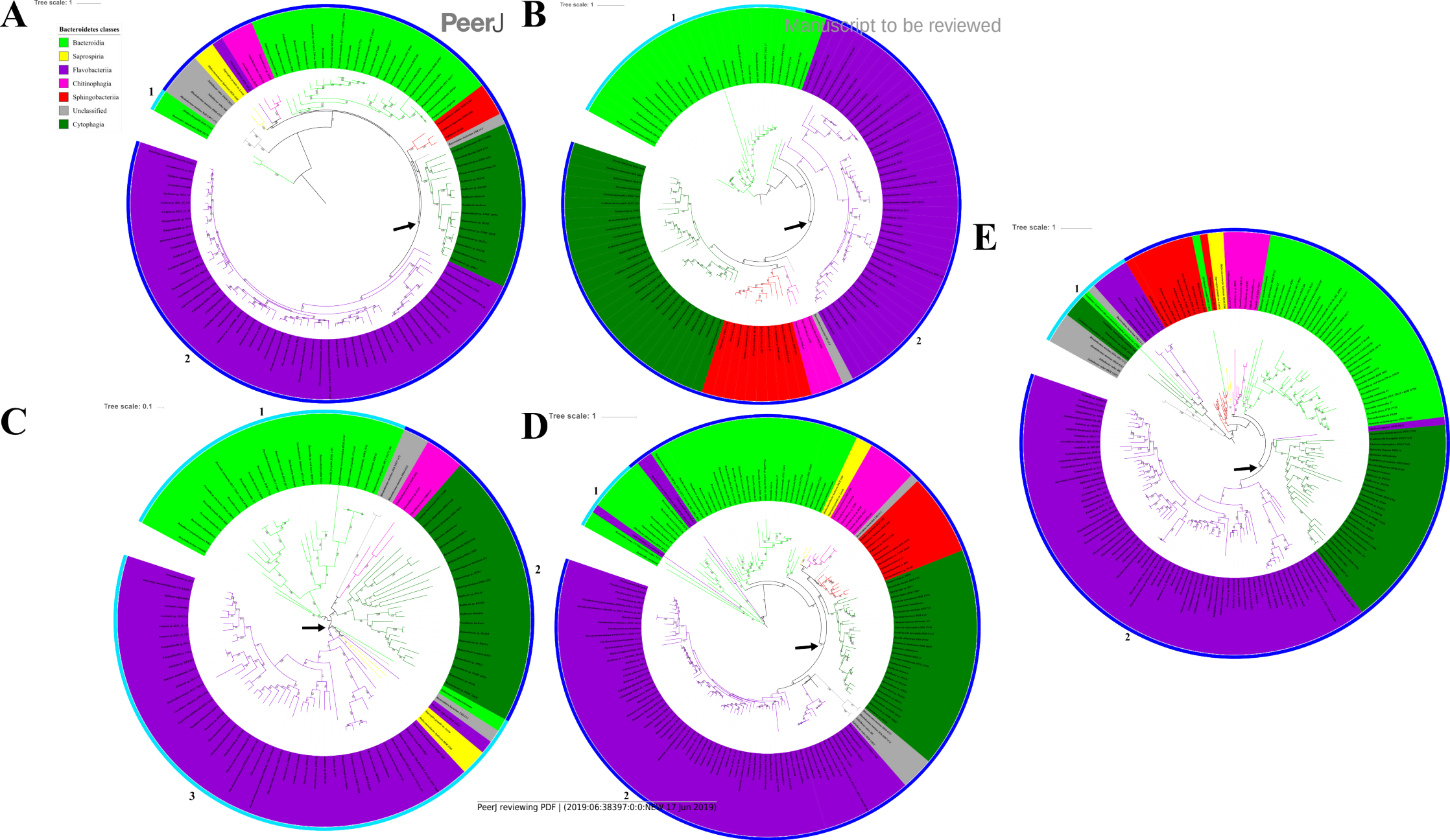
(A) ML tree of Sov. (B) ML tree of PorX. (C) ML tree of PorV. (D) ML tree of PorU.



## Figure 4(on next page)

The maximum-likelihood (ML) phylogenetic trees of five protein components of T9SS that exhibit phylogenetic relationship: (Cytophagia + Flavobacteriia) + Bacteroidia.

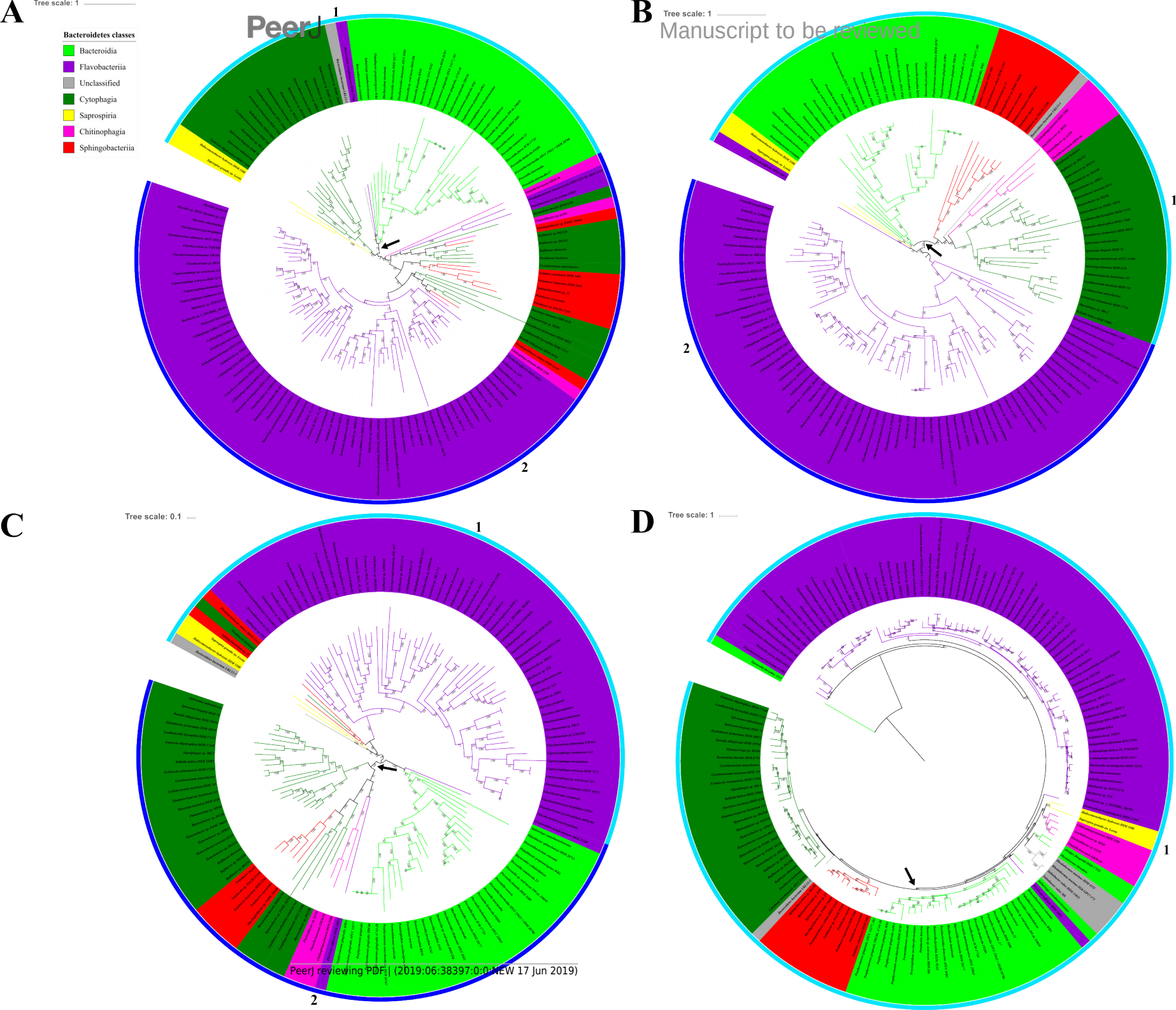
(A) ML tree of PG0534. (B) ML tree of PorL. (C) ML tree of PorQ. (D) ML tree of PG0162. (E) ML tree of PGN0300. The nodes leading to clusters of major classes with low bootstrap support values (less than 70) are pointed by the black arrows.



## Figure 5(on next page)

The maximum-likelihood (ML) phylogenetic trees of four protein components of T9SS that exhibit phylogenetic relationship: (Bacteroidia + Cytophagia) + Flavobacteriia.

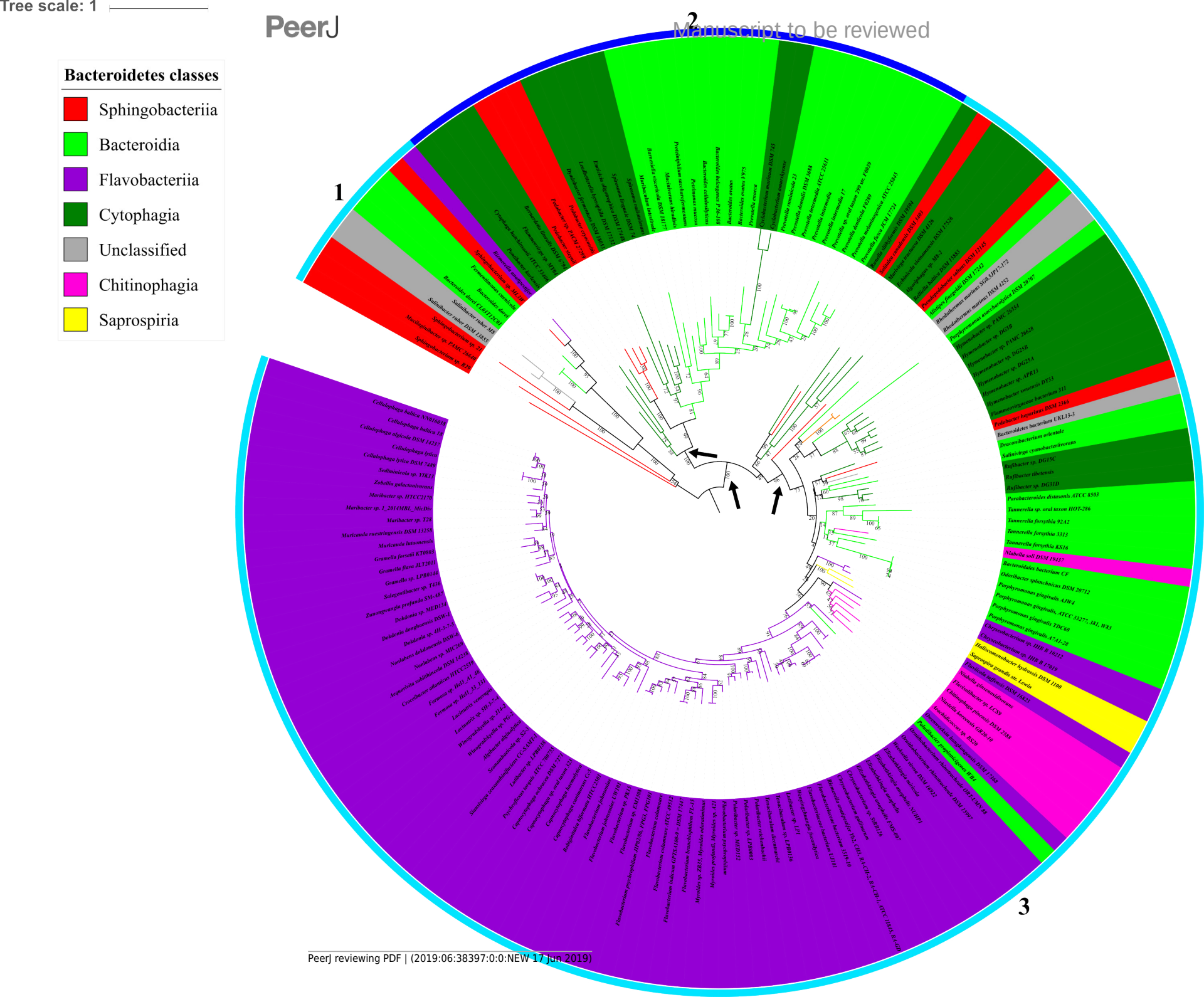
(A) ML tree of PorP. (B) ML tree of PorW. (C) ML tree of PG1058. (D) ML tree of PorY. The nodes leading to clusters of major classes with low bootstrap support values (less than 70) are pointed by the black arrows.



## Figure 6(on next page)

The maximum-likelihood (ML) phylogenetic tree of PorR. The terminal nodes of ML tree of PorR do not cluster based on their classes under Bacteroidetes phylum unlike the other 19 protein components of T9SS.

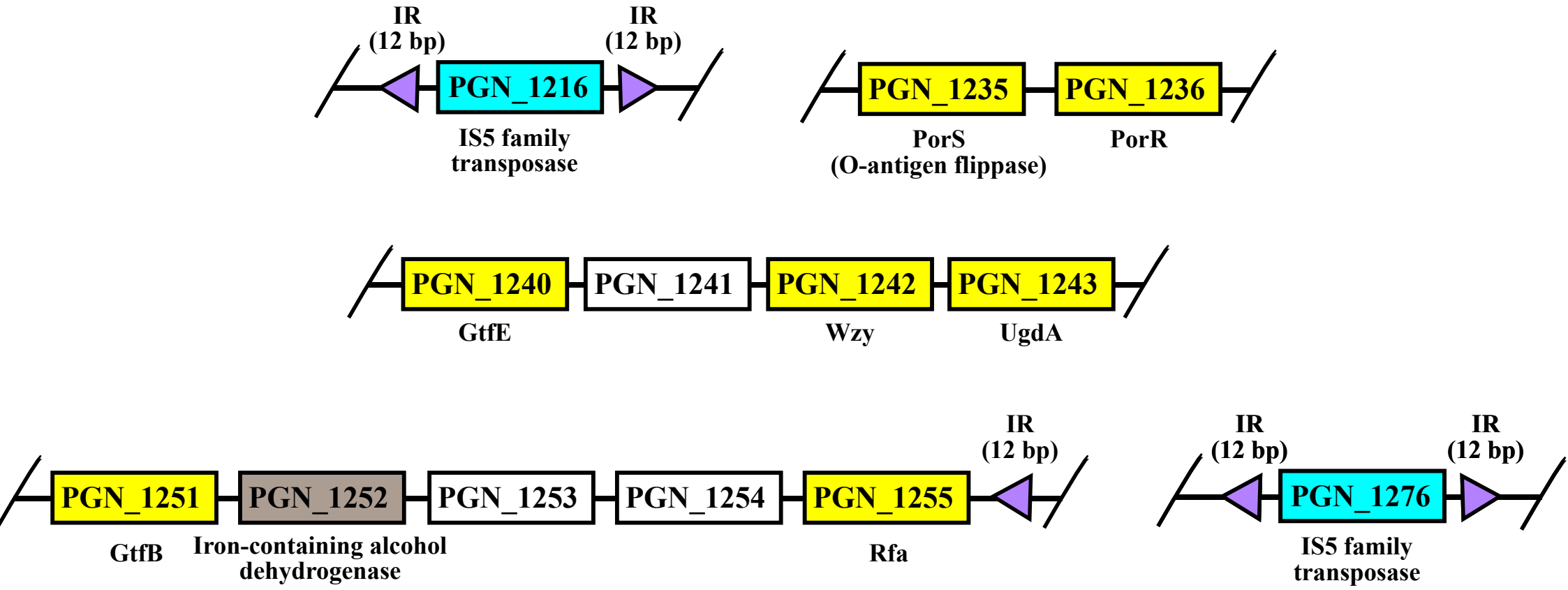
The nodes leading to clusters of major classes with high bootstrap support values (higher or equal 70) are pointed by the black arrows.


Tree scale: 1
Bacteroidetes classes
Sphingobacteriia
Bacteroidia
Flavobacteriia
Cytophagia
Unclassified
Chitinophagia
Saprospiriia
1
2
3

## Figure 7(on next page)

The arrangement of porR and its neighbouring genes in P. gingivalis ATCC 33277 genome.

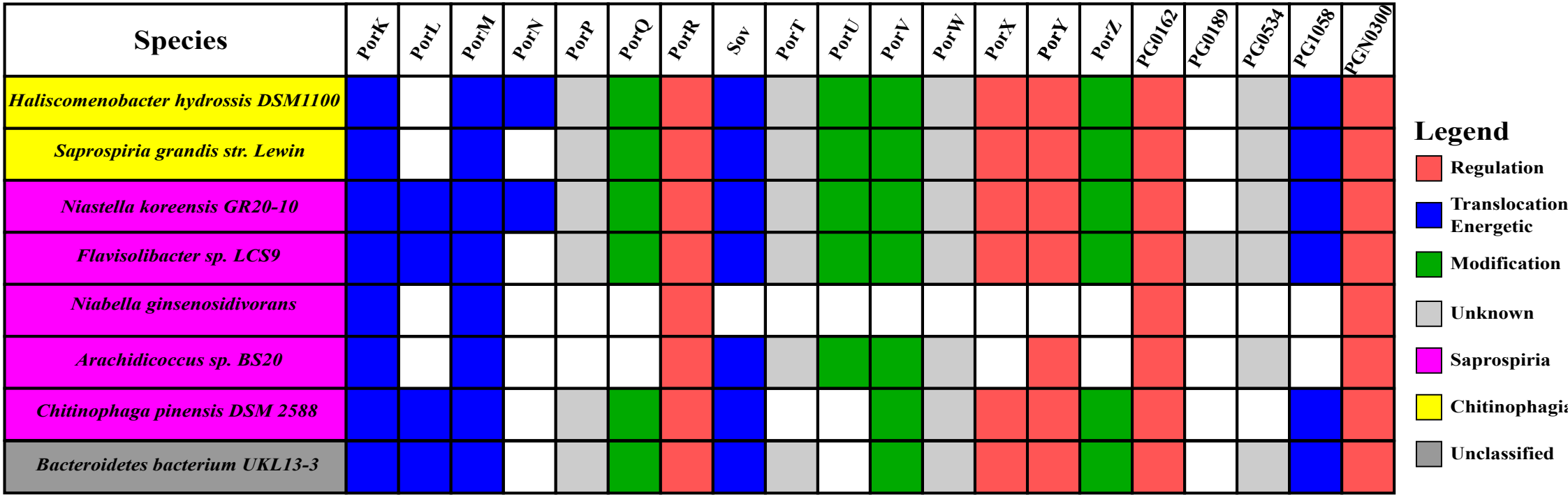
porR (PGN_1236) and its neighbouring genes are flanked by IS5 transposons that formed a composite transposon of 70 kbp in length. The genes that involved in biosynthesis of A-LPS are represented by yellow coloured rectangles while the gene that does not involve is represented by grey coloured rectangle. The genes for hypothetical proteins are represented by white coloured rectangles. The genes for IS5 family transposases are represented by cyan coloured rectangles. The purple triangles represented 12 bp inverted repeats that flanked the genes for IS5 family transposases. Putatives proteins encoded by the genes are shown under rectangles that represented the genes. The slashes indicated gaps in the genome.


IR
(12 bp)
IR
(12 bp)
PGN_1216
IS5 family
transposase
PGN_1235
PGN_1236
PorS
(O-antigen flippase)
PorR
PGN_1240
PGN_1241
PGN_1242
PGN_1243
GtfE
Wzy
UgdA
IR
(12 bp)
PGN_1251
PGN_1252
PGN_1253
PGN_1254
PGN_1255
GtfB
Iron-containing alcohol
dehydrogenase
Rfa
IR
(12 bp)
IR
(12 bp)
PGN_1276
IS5 family
transposase

## Figure 8(on next page)

The additional species from Chitinophagia, Saprospiria, and unclassified under Bacteroidetes phylum that acquired homologs of T9SS protein components.

The colours denote the classes those bacterial species belong to. Coloured squares indicate the homologs of T9SS protein components acquired by the bacterial species where the different colours denote different functions those components performed. White squares indicate the homologs of T9SS protein components absent in the bacterial species.


Species
PorK
PorL
PorM
PorN
PorP
PorQ
PorR
Sov
PorT
PorU
PorV
PorW
PorX
PorY
PorZ
PG0162
PG0189
PG0534
PG1058
PGN0300
Haliscomenobacter hydrossis DSM1100
Saprospiria grandis str. Lewin
Niastella koreensis GR20-10
Flavisolibacter sp. LCS9
Niabella ginsenosidivorans
Arachidicoccus sp. BS20
Chitinophaga pinensis DSM 2588
Bacteroidetes bacterium UKL13-3
Legend
Regulation
Translocation
Energetic
Modification
Unknown
Saprospiria
Chitinophagia
Unclassified