

A peer-reviewed version of this preprint was published in PeerJ on 26 June 2020.

[View the peer-reviewed version](https://peerj.com/articles/9019) (peerj.com/articles/9019), which is the preferred citable publication unless you specifically need to cite this preprint.

Emrizal R, Nor Muhammad NA. 2020. Phylogenetic comparison between Type IX Secretion System (T9SS) protein components suggests evidence of horizontal gene transfer. PeerJ 8:e9019
<https://doi.org/10.7717/peerj.9019>

1 **Phylogenetic analysis of Type IX Secretion System**
2 **(T9SS) protein components revealed that PorR**
3 **undergoes horizontal gene transfer**

4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

Reeki Emrizal¹, Nor Azlan Nor Muhammad¹

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 Muhammad¹

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

Email address: norazlanm@ukm.edu.my

37 Abstract

38

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

71

72 Introduction

73

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

77 periodontitis can cause the destruction of those supporting structures that ultimately lead to tooth
78 loss and might contribute to the systemic inflammation (Kinane, Stathopoulou & Papapanou,
79 2017; Escobar et al., 2018). This is evidenced by its implications in systemic diseases such as
80 atherosclerosis (Gotsman et al., 2007), aspiration pneumonia (Benedyk et al., 2016), cancer (Gao
81 et al., 2016), rheumatoid arthritis (Laugisch et al., 2016), and diabetes mellitus (Khader et al.,
82 2006). *Porphyromonas gingivalis* is an oral pathogen that is frequently associated with
83 periodontitis and it is found to acquire Type IX Secretion System (T9SS); a bacterial secretion
84 system that is unique to Bacteroidetes phylum (Sato et al., 2010).

85 T9SS exhibits diverse roles among species of bacteria under Bacteroidetes phylum. Other
86 than transporting virulence factors such as gingipains and peptidylarginine deiminase in *P.*
87 *gingivalis* that can cause human oral diseases (Potempa, Pike & Travis, 1995; Maresz et al.,
88 2013), T9SS also transporting virulence factors such as chondroitin sulfate lyases that are
89 associated with fish diseases such as columnaris disease. *Flavobacterium columnare*, a fish
90 pathogen that contributes to the epidemic that occurred among wild and cultured fishes, is found
91 to acquire T9SS. This epidemic poses a problem to the aquaculture industry as columnaris
92 disease caused by the fish pathogen can significantly increase the mortality rate among cultured
93 fishes and thus threatened the industry output (Li et al., 2017). T9SS also involved in the
94 transport of non-virulence factors such as cargo proteins that formed the bacterial gliding
95 motility apparatus in *Flavobacterium johnsoniae* that aids in its motility (Nakane et al., 2013)
96 and enzymes that are importance for lignocellulose digestion in the rumen of ruminants that
97 become the hosts for the *Candidatus Paraporphyromonas polyenzymogenes* (Naas et al., 2018).

98 Similar to other types of secretion system in bacteria, T9SS is consisted of many different
99 protein components that perform coordinated roles to ensure proper translocation and
100 modification of its cargo proteins. These roles can be categorised into four major functions:
101 translocation, modification, energetic, and regulation (Sato et al., 2010; Lasica et al., 2017).
102 There are six protein components that performed the translocation function. Four of them (PorL,
103 PorM, PorK, and PorN) formed the trans-envelope complex ($\text{PorK}_2\text{L}_3\text{M}_2\text{N}_2$) that acts as the main
104 translocation channel that transports cargo proteins from the periplasm to extracellular milieu
105 (Vincent et al., 2017) (Fig. 1). Despite having a translocation channel that spanned both OM and
106 IM, the presence of N-terminal signal peptide in cargo proteins indicates that T9SS depends on
107 SecYEG to translocate its cargo proteins from the cytoplasm to periplasm which is similar to
108 Type II Secretion System (T2SS) (Overbye, Sandkvist & Bagdasarian, 1993; Costa et al., 2015).
109 During translocation across IM, the signal peptide of cargo proteins is cleaved by type I signal
110 peptidase before they are released into the periplasm (Rahman et al., 2003) (Fig. 1). The cargo
111 proteins of T9SS also acquired C-terminal domain (CTD) that will interact sequentially with the
112 D2, D3, and D4 domains of PorM and then PorN of the trans-envelope complex to be directed to
113 the T9SS secretion pore in OM (Vincent, Chabaliere & Cascales, 2018) (Fig. 1). It is also
114 suggested that PG1058 aids the transport of cargo proteins across the periplasm (Heath et al.,
115 2016) (Fig. 1). Recently, it is proposed that SprA (ortholog of Sov in *F. johnsoniae*) is the
116 secretion pore that translocates cargo proteins across OM. SprA is a 36-stranded transmembrane

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

136 The translocation of cargo proteins by T9SS might be energised by PorL and PorM that
137 formed an energy transducer complex (Vincent et al., 2017). It is suggested that the energy
138 transducer complex converts proton-motive force into mechanical energy that can provide the
139 energy for cargo proteins translocation across T9SS. It has been reported that PorL and PorM
140 bear signatures typical of an energy transducer employing proton-motive force (Vincent et al.,
141 2017). The processes by T9SS are also regulated by its six protein components that is probably
142 to ensure the virulence factors are secreted by bacteria when it is in a favourable environment.
143 PorX and PorY formed a two-component system (TCS) that can regulate the operon of *por* genes
144 (*porP*, *porK*, *porL*, *porM*, and *porN*) (Vincent et al., 2017). However, PorX that acts as a
145 response regulator protein in TCS cannot directly bind the promoter of *por* genes but depends on
146 PG0162 to form a complex that can bind the promoter (Kadowaki et al., 2016) (Fig. 1). PorX is
147 also found to interact with PorL in which this interaction might serve a regulatory purpose
148 (Vincent, Durand & Cascales, 2016). PorR is an aminotransferase that involves in the Wbp
149 pathway that contributes to biosynthesis of the structural sugar of A-LPS (Shoji et al., 2014,
150 2018) (Fig. 1). PGN0300 can influence the present of PorU on cell surface however the exact
151 mechanisms behind it remained vague (Taguchi et al., 2016).

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

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

174

175 **Materials & Methods**

176

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

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

188 The list of protein homologs, which are the selected putative members for each T9SS
189 component protein family, used to build the multiple sequence alignments for each T9SS protein
190 component was provided as a Supplemental Information (Data S1). The multiple sequence
191 alignments were constructed using T-Coffee version (11.00) (Notredame, Higgins & Heringa,
192 2000). The multiple sequence alignments in clustal format constructed by T-Coffee were
193 converted into phylip relaxed (interleaved) format using online Format Converter
194 (https://www.hiv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html). The
195 multiple sequence alignments in phylip relaxed (interleaved) format were manually formatted by

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

198

199 **Phylogeny reconstruction for protein components of T9SS**

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

208

209 **Identification of *porR* and its neighbouring genes arrangement in *P. gingivalis* ATCC 33277 210 genome**

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

221

222 **Results**

223

224 **Phylogenetic trees of protein components of T9SS**

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

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

236 those with more than five families under them (Bacteroidia, Cytophagia, and Flavobacteriia)
237 while minor classes are those with less than or equal to five families under them (Chitinophagia,
238 Sphingobacteriia, Saprospira, and unclassified). The phylogenetic relationship between three
239 major classes for 10 protein components is consistently recovered as (Bacteroidia +
240 Flavobacteriia) + Cytophagia. This relationship implied that Bacteroidia is genetically closer to
241 Flavobacteriia than either of them to Cytophagia. The ML trees for 10 protein components that
242 exhibit this relationship are shown in Fig. 2 and Fig. 3. The phylogenetic relationship between
243 three major classes for five protein components is consistently recovered as (Cytophagia +
244 Flavobacteriia) + Bacteroidia. This relationship implied that Cytophagia is genetically closer to
245 Flavobacteriia than either of them to Bacteroidia. The ML trees for five protein components that
246 exhibit this relationship are shown in Fig. 4. The phylogenetic relationship between three major
247 classes for four protein components is consistently recovered as (Bacteroidia + Cytophagia) +
248 Flavobacteriia. This relationship implied that Bacteroidia is genetically closer to Cytophagia than
249 either of them to Flavobacteriia. The ML trees for four protein components that exhibit this
250 relationship are shown in Fig. 5.

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

261

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

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

275

276 Taxonomic distributions of protein components of T9SS

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

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

295

296 Discussion

297

298 The maximum-likelihood (ML) phylogenetic trees of 19 protein components of T9SS exhibit the
299 clustering of terminal nodes based on their classes (Figs. 2-5). To identify the similarities in
300 phylogenetic relationships between those 19 ML trees, we have decided to exclude the minor
301 classes (Chitinophagia, Sphingobacteriia, Saprospira, and unclassified). The reason behind this
302 is because terminal nodes that represented the minor classes tend to be inconsistent in its
303 positions in those ML trees. Thus, it causes difficulty to draw out similarities between them. This
304 might arise due to insufficient taxa from these classes are provided to construct the phylogenetic
305 trees. Thus, less information is provided which is insufficient to properly resolved the phylogeny
306 of these classes. As more T9SS-acquiring bacterial strains from these classes are sequenced,
307 probably the phylogenies of T9SS protein components will be more resolved (Alvizu et al.,
308 2018).

309 Three distinct phylogenetic relationships between the three major classes (Bacteroidia,
310 Cytophagia, and Flavobacteriia) are identified: (Bacteroidia + Flavobacteriia) + Cytophagia,
311 (Cytophagia + Flavobacteriia) + Bacteroidia, and (Bacteroidia + Cytophagia) + Flavobacteriia.
312 The ML trees of 10 protein components (PorK, PorM, PorT, PorN, PorZ, PG1058, PorU, Sov,
313 PorX, and PorV) exhibit the phylogenetic relationship (Bacteroidia + Flavobacteriia) +
314 Cytophagia (Figs. 2 and 3) that is consistent with the phylogeny of bacteria under Bacteroidetes
315 phylum based on 16S rRNA sequence that has been reported (Karlsson et al., 2011). The

316 phylogenetic relationship that is recovered by the reported study is (Bacteroidia + Flavobacteriia)
317 + (Rhodothermaceae + (Chitinophagia + (Sphingobacteriia + Cytophagia))) (Karlsson et al.,
318 2011). This relationship suggests that Bacteroidia is genetically close to Flavobacteria than either
319 of them to Cytophagia which is similar to the relationship exhibited by the ML trees of those 10
320 protein components. The ML trees of five protein components (PorL, PorQ, PG0162, PG0534,
321 and PGN0300) exhibit the phylogenetic relationship (Cytophagia + Flavobacteriia) + Bacteroidia
322 (Fig. 4) while the ML trees of four protein components (PorP, PorW, PorY, and PG1058) exhibit
323 the phylogenetic relationship (Bacteroidia + Cytophagia) + Flavobacteriia (Fig. 5). In those nine
324 ML trees, there are low bootstrap support values (less than 70) at the nodes leading to the
325 clusters of major classes that are pointed by the black arrows in (Figs. 4 and 5). Hence, there is a
326 low support for the phylogenetic relationship exhibited by those nine ML trees deviates from the
327 phylogeny of bacteria under Bacteroidetes phylum based on 16S rRNA sequence that has been
328 reported (Karlsson et al., 2011).

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

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

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

361 The intervening DNA segment contains seven genes that involved in the biosynthesis of
362 A-LPS (Fig. 7). Both *porR* and *ugdA* genes are involved in Wbp pathway that is important for
363 biosynthesis of di-acetylated glucuronic acid that is the structural sugar of A-LPS (Shoji et al.,
364 2014). *porS*, which is an O-antigen flippase similar to *wzx*, and *wzy*, which is an O-antigen
365 polymerase, genes are involved in the assembly of A-LPS at the periplasmic side of bacterial
366 inner membrane (Shoji et al., 2013). *gtfB* and *gtfE* glycosyltransferases genes are involved in the
367 biosynthesis of sugar moiety of A-LPS. *rfa* glycosyltransferase gene is involved in the
368 biosynthesis of lipid A-core moiety of A-LPS (Shoji et al., 2018). However, there are many other
369 genes that involved in the biosynthesis of A-LPS and they are spread out throughout the genome
370 (Shoji et al., 2018). This raise a question of why these genes are not clustered in a single operon
371 which is usually the case for genes that involved in a similar pathway. It might be possible that
372 only those seven genes identified in this study are horizontally transferred while the others are
373 not. Thus, explaining why only those seven genes clustered in a composite transposon. Perhaps,
374 those other genes also are flanked by mobile elements that aid in their transposition and
375 conjugative transfer. Further analyses such as comparing the molecular phylogeny of all genes
376 involved in the biosynthesis of A-LPS and comparing the arrangement of those genes in the
377 bacterial genomes that synthesis A-LPS might be able to answer those questions.

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

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

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

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

409

410 Conclusions

411

412 We found that the maximum-likelihood (ML) phylogenetic trees for 19 protein components of
413 T9SS exhibit clustering of terminal nodes based on their respective classes under Bacteroidetes
414 phylum. Majority of the ML trees of those protein components exhibit the phylogenetic
415 relationship that is consistent with phylogeny of bacteria under Bacteroidetes phylum based on
416 16S rRNA sequence (Karlsson et al., 2011). The ML tree of PorR, which is an aminotransferase
417 that involved in the Wbp pathway that produces structural sugar of A-LPS, exhibits different
418 clustering pattern of terminal nodes where the nodes do not cluster based on their respective
419 classes. Hence, PorR might evolve independently from the other protein components that might
420 suggest PorR is acquired by bacteria with T9SS through horizontal gene transfer. We found that
421 seven genes (including *porR*) that involved in the biosynthesis of A-LPS that anchored the
422 virulence factor secreted by T9SS to bacterial cell surface are flanked by insertion sequences
423 (ISs) that encode IS5 family transposase. This might suggest that the intervening DNA segment
424 that contains *porR* gene can be mobilized and subsequently contributes to the possibility that
425 *porR* gene is subjected to conjugative transfer.

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

431

432 Acknowledgements

433

434 We thank Centre for Bioinformatics Research (CBR) for providing the facilities to conduct the
435 bioinformatics analyses. We thank the anonymous reviewers for their comments on previous
436 drafts of the manuscript.

437

438 References

439

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

442 **Abby SS, Cury J, Guglielmini J, Néron B, Touchon M, Rocha EPC. 2016.** Identification of
443 protein secretion systems in bacterial genomes. *Scientific reports* **6**:23080 DOI
444 10.1038/srep23080.

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

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

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

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

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

460 **Du K, Zhong Z, Fang C, Dai W, Shen Y, Gan X, He S. 2018.** Ancient duplications and
461 functional divergence in the interferon regulatory factors of vertebrates provide insights into
462 the evolution of vertebrate immune systems. *Developmental & Comparative Immunology*
463 **81**:324–333 DOI 10.1016/j.dci.2017.12.016.

464 **Emrizal R, Muhammad NAN. 2018.** Identification of sequence motifs for the protein
465 components of type IX secretion system. *Sains Malaysiana* **47(12)**:2941–2950 DOI
466 10.17576/jsm-2018-4712-02.

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

472 **Gao S, Li S, Ma Z, Liang S, Shan T, Zhang M, Zhu X, Zhang P, Liu G, Zhou F, Yuan X,
473 Jia R, Potempa J, Scott DA, Lamont RJ, Wang H, Feng X. 2016.** Presence of
474 *Porphyromonas gingivalis* in esophagus and its association with the clinicopathological
475 characteristics and survival in patients with esophageal cancer. *Infectious agents and cancer*
476 **11(1)**:3 DOI 10.1186/s13027-016-0049-x.

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

- 479 10.1038/s41598-017-09412-w.
- 480 **Glew MD, Veith PD, Peng B, Chen YY, Gorasia DG, Yang Q, Slakeski N, Chen D, Moore**
481 **C, Crawford S, Reynolds EC. 2012.** PG0026 is the C-terminal signal peptidase of a novel
482 secretion system of *Porphyromonas gingivalis*. *Journal of Biological Chemistry*
483 **287(29):24605–24617** DOI 10.1074/jbc.M112.369223.
- 484 **Gorasia DG, Veith PD, Hanssen EG, Glew MD, Sato K, Yukitake H, Nakayama K,**
485 **Reynolds EC. 2016.** Structural insights into the PorK and PorN components of the
486 *Porphyromonas gingivalis* type IX secretion system. *PLoS pathogens* **12(8):e1005820** DOI
487 10.1371/journal.ppat.1005820.
- 488 **Gotsman I, Lotan C, Soskolne WA, Rassovsky S, Pugatsch T, Lapidus L, Novikov Y,**
489 **Masrawa S, Stabholz A. 2007.** Periodontal destruction is associated with coronary artery
490 disease and periodontal infection with acute coronary syndrome. *Journal of periodontology*
491 **78(5):849–858** DOI 10.1902/jop.2007.060301.
- 492 **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010.** New
493 algorithms and methods to estimate maximum-likelihood phylogenies: assessing the
494 performance of PhyML 3.0. *Systematic Biology* **59(3):307–321** DOI
495 10.1093/sysbio/syq010.
- 496 **Heath JE, Seers CA, Veith PD, Butler CA, Muhammad NAN, Chen YY, Slakeski N, Peng**
497 **B, Zhang L, Dashper SG, Cross KJ, Cleal SM, Moore C, Reynolds EC. 2016.** PG1058
498 is a novel multidomain protein component of the bacterial type IX secretion system. *PloS*
499 *one* **11(10):e0164313** DOI 10.1371/journal.pone.0164313.
- 500 **Hess PN, De Moraes Russo CA. 2007.** An empirical test of the midpoint rooting method.
501 *Biological Journal of the Linnean Society* **92(4):669–674** DOI 10.1111/j.1095-
502 8312.2007.00864.x.
- 503 **Hirt H, Schlievert PM, Dunny GM. 2002.** In vivo induction of virulence and antibiotic
504 resistance transfer in *Enterococcus faecalis* mediated by the sex pheromone-sensing system
505 of pCF10. *Infection and immunity* **70(2):716–723** DOI 10.1128/IAI.70.2.716.
- 506 **Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006.** Phylogenomics: the beginning of
507 incongruence? *Trends in Genetics* **22(4):225–231** DOI 10.1016/j.tig.2006.02.003.
- 508 **Kadowaki T, Yukitake H, Naito M, Sato K, Kikuchi Y, Kondo Y, Shoji M, Nakayama K.**
509 **2016.** A two-component system regulates gene expression of the type IX secretion
510 component proteins via an ECF sigma factor. *Scientific reports* **6:23288** DOI
511 10.1038/srep23288.
- 512 **Karlsson FH, Ussery DW, Nielsen J, Nookaew I. 2011.** A closer look at Bacteroides:
513 phylogenetic relationship and genomic implications of a life in the human gut. *Microbial*
514 *ecology* **61(3):473–485** DOI 10.1007/s00248-010-9796-1.
- 515 **Khader YS, Dauod AS, El-Qaderi SS, Alkafajei A, Batayha WQ. 2006.** Periodontal status of
516 diabetics compared with nondiabetics: a meta-analysis. *Journal of diabetes and its*
517 *complications* **20(1):59–68** DOI 10.1016/j.jdiacomp.2005.05.006.
- 518 **Kinane DF, Stathopoulou PG, Papananou PN. 2017.** Periodontal diseases. *Nature Reviews*
519 *Disease Primers* **3:17038** DOI doi:10.1038/nrdp.2017.38.
- 520 **Lasica AM, Goulas T, Mizgalska D, Zhou X, De Diego I, Ksiazek M, Madej M, Guo Y,**
521 **Guevara T, Nowak M, Potempa B, Goel A, Sztukowska M, Prabhakar AT, Bzowska**
522 **M, Widziol M, Thøgersen IB, Enghild JJ, Simonian M, Kulczyk AW, Nguyen KA,**
523 **Potempa J, Gomis-Rüth FX. 2016.** Structural and functional probing of PorZ, an essential
524 bacterial surface component of the type-IX secretion system of human oral-microbiomic

- 525 Porphyromonas gingivalis. *Scientific reports* **6**:37708 DOI 10.1038/srep37708.
- 526 **Lasica AM, Ksiazek M, Madej M, Potempa J. 2017.** The type IX secretion system (T9SS):
527 highlights and recent insights into its structure and function. *Frontiers in cellular and*
528 *infection microbiology* **7**:215 DOI 10.3389/fcimb.2017.00215.
- 529 **Lauber F, Deme JC, Lea SM, Berks BC. 2018.** Type 9 secretion system structures reveal a
530 new protein transport mechanism. *Nature* **564(7734)**:77 DOI 10.1038/s41586-018-0693-y.
- 531 **Laugisch O, Wong A, Sroka A, Kantyka T, Koziel J, Neuhaus K, Sculean A, Venables PJ,**
532 **Potempa J, Möller B, Eick S. 2016.** Citrullination in the periodontium—a possible link
533 between periodontitis and rheumatoid arthritis. *Clinical oral investigations* **20(4)**:675–683
534 DOI 10.1007/s00784-015-1556-7.
- 535 **Letunic I, Bork P. 2016.** Interactive tree of life (iTOL) v3: an online tool for the display and
536 annotation of phylogenetic and other trees. *Nucleic acids research* **44(W1)**:W242–W245
537 DOI 10.1093/nar/gkw290.
- 538 **Li N, Zhu Y, LaFrentz BR, Evenhuis JP, Hunnicutt DW, Conrad RA, Barbier P,**
539 **Gullstrand CW, Roets JE, Powers JL, Kulkarni SS, Erbes DH, Garcia JC, Nie P,**
540 **McBride MJ. 2017.** The type IX secretion system is required for virulence of the fish
541 pathogen *Flavobacterium columnare*. *Appl. Environ. Microbiol.* **83(23)**:e01769-17 DOI
542 10.1128/aem.01769-17.
- 543 **Mahillon J, Chandler M. 1998.** Insertion sequences. *Microbiol. Mol. Biol. Rev.* **62(3)**:725–774
- 544 **Maresz KJ, Hellvard A, Sroka A, Adamowicz K, Bielecka E, Koziel J, Gawron K,**
545 **Mizgalska D, Marcinska KA, Benedyk M, Pyrc K, Quirke AM, Jonsson R, Alzabin S,**
546 **Venables PJ, Nguyen KA, Mydel P, Potempa J. 2013.** *Porphyromonas gingivalis*
547 facilitates the development and progression of destructive arthritis through its unique
548 bacterial peptidylarginine deiminase (PAD). *PLoS pathogens* **9(9)**:e1003627 DOI
549 10.1371/journal.ppat.1003627.
- 550 **McBride MJ, Zhu Y. 2013.** Gliding motility and Por secretion system genes are widespread
551 among members of the phylum Bacteroidetes. *Journal of bacteriology* **195(2)**:270–278 DOI
552 10.1128/JB.01962-12.
- 553 **Naas AE, Solden LM, Norbeck AD, Brewer H, Hagen LH, Heggenes IM, McHardy AC,**
554 **Mackie RI, Paša-Tolic L, Arntzen MØ, Eijsink VGH, Koropatkin NM, Hess M,**
555 **Wrighton KC, Pope PB. 2018.** “Candidatus *Paraporphyromonas polyenzymogenes*”
556 encodes multi-modular cellulases linked to the type IX secretion system. *Microbiome*
557 **6(1)**:44 DOI 10.1186/s40168-018-0421-8.
- 558 **Naito M, Hirakawa H, Yamashita A, Ohara N, Shoji M, Yukitake H, Nakayama K, Toh H,**
559 **Yoshimura F, Kuhara S, Hattori M. 2008.** Determination of the genome sequence of
560 *Porphyromonas gingivalis* strain ATCC 33277 and genomic comparison with strain W83
561 revealed extensive genome rearrangements in *P. gingivalis*. *DNA research* **15(4)**:215–225
562 DOI 10.1093/dnares/dsn013.
- 563 **Nakane D, Sato K, Wada H, McBride MJ, Nakayama K. 2013.** Helical flow of surface
564 protein required for bacterial gliding motility. *Proceedings of the National Academy of*
565 *Sciences* **110(27)**:11145–11150 DOI 10.1073/pnas.1219753110.
- 566 **Nguyen KA, Zylicz J, Szczesny P, Sroka A, Hunter N, Potempa J. 2009.** Verification of a
567 topology model of PorT as an integral outer-membrane protein in *Porphyromonas*
568 *gingivalis*. *Microbiology (Reading, England)* **155(Pt 2)**:328 DOI 10.1099/mic.0.024323-0.
- 569 **Notredame C, Higgins DG, Heringa J. 2000.** T-coffee: A novel method for fast and accurate
570 multiple sequence alignment. *Journal of molecular biology* **302(1)**:205–217 DOI

- 571 10.1006/jmbi.2000.4042.
- 572 **Overbye LJ, Sandkvist M, Bagdasarian M. 1993.** Genes required for extracellular secretion of
573 enterotoxin are clustered in *Vibrio cholerae*. *Gene* **132(1)**:101–106 DOI 10.1016/0378-
574 1119(93)90520-D.
- 575 **Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. 2010.** How
576 many bootstrap replicates are necessary? *Journal of computational biology* **17(3)**:337–354
- 577 **Potempa J, Pike R, Travis J. 1995.** The multiple forms of trypsin-like activity present in
578 various strains of *Porphyromonas gingivalis* are due to the presence of either Arg-gingipain
579 or Lys-gingipain. *Infection and immunity* **63(4)**:1176–1182
- 580 **Pylro VS, Vespoli LDS, Duarte GF, Yotoko KSC. 2012.** Detection of horizontal gene transfers
581 from phylogenetic comparisons. *International journal of evolutionary biology* **2012** DOI
582 10.1155/2012/813015.
- 583 **Rahman MS, Simser JA, Macaluso KR, Azad AF. 2003.** Molecular and functional analysis of
584 the lepB gene, encoding a type I signal peptidase from *Rickettsia rickettsii* and *Rickettsia*
585 *typhi*. *Journal of Bacteriology* **185(15)**:4578–4584 DOI 10.1128/JB.185.15.4578-
586 4584.2003.
- 587 **Saiki K, Konishi K. 2010.** Identification of a novel *Porphyromonas gingivalis* outer membrane
588 protein, PG534, required for the production of active gingipains. *FEMS microbiology letters*
589 **310(2)**:168–174 DOI 10.1111/j.1574-6968.2010.02059.x.
- 590 **Sato K, Naito M, Yukitake H, Hirakawa H, Shoji M, McBride MJ, Rhodes RG, Nakayama**
591 **K. 2010.** A protein secretion system linked to bacteroidete gliding motility and
592 pathogenesis. *Proceedings of the National Academy of Sciences* **107(1)**:276–281 DOI
593 10.1073/pnas.0912010107.
- 594 **Schwarz G. 1978.** Estimating the dimension of a model. *The annals of statistics* **6(2)**:461–464
- 595 **Shoji M, Sato K, Yukitake H, Kamaguchi A, Sasaki Y, Naito M, Nakayama K. 2018.**
596 Identification of genes encoding glycosyltransferases involved in lipopolysaccharide
597 synthesis in *Porphyromonas gingivalis*. *Molecular oral microbiology* **33(1)**:68–80 DOI
598 10.1111/omi.12200.
- 599 **Shoji M, Sato K, Yukitake H, Naito M, Nakayama K. 2014.** Involvement of the Wbp pathway
600 in the biosynthesis of *Porphyromonas gingivalis* lipopolysaccharide with anionic
601 polysaccharide. *Scientific reports* **4**:5056 DOI 10.1038/srep05056.
- 602 **Shoji M, Yukitake H, Sato K, Shibata Y, Naito M, Aduse-Opoku J, Abiko Y, Curtis MA,**
603 **Nakayama K. 2013.** Identification of an O-antigen chain length regulator, WzzP, in
604 *Porphyromonas gingivalis*. *Microbiologyopen* **2(3)**:383–401 DOI 10.1002/mbo3.84.
- 605 **Singh SP, Groeneveld JC, Al-Marzouqi A, Willows-Munro S. 2017.** A molecular phylogeny
606 of the spiny lobster *Panulirus homarus* highlights a separately evolving lineage from the
607 Southwest Indian Ocean. *PeerJ* **5**:e3356 DOI 10.7717/peerj.3356.
- 608 **Taguchi Y, Sato K, Yukitake H, Inoue T, Nakayama M, Naito M, Kondo Y, Kano K,**
609 **Hoshino T, Nakayama K, Takashiba S, Ohara N. 2016.** Involvement of an Skp-like
610 protein, PGN_0300, in the Type IX secretion system of *Porphyromonas gingivalis*. *Infection*
611 *and immunity* **84(1)**:230–240 DOI 10.1128/IAI.01308-15.
- 612 **Thomas CM, Nielsen KM. 2005.** Mechanisms of, and barriers to, horizontal gene transfer
613 between bacteria. *Nature reviews microbiology* **3(9)**:711 DOI 10.1038/nrmicro1234.
- 614 **Vincent MS, Chabaliier M, Cascales E. 2018.** A conserved motif of *Porphyromonas* Type IX
615 secretion effectors C-terminal secretion signal specifies interactions with the PorKLMN
616 core complex. *bioRxiv*:483123 DOI 10.1101/483123.

- 617 **Vincent MS, Canestrari MJ, Leone P, Stathopoulos J, Ize B, Zoued A, Cambillau C,**
618 **Kellenberger C, Roussel A, Cascales E. 2017.** Characterization of the Porphyromonas
619 gingivalis type IX secretion trans-envelope PorKLMNP core complex. *Journal of*
620 *Biological Chemistry* **292(8)**:3252–3261 DOI 10.1074/jbc.M116.765081.
- 621 **Vincent MS, Durand E, Cascales E. 2016.** The PorX response regulator of the Porphyromonas
622 gingivalis PorXY two-component system does not directly regulate the type IX secretion
623 genes but binds the PorL subunit. *Frontiers in cellular and infection microbiology* **6**:96 DOI
624 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

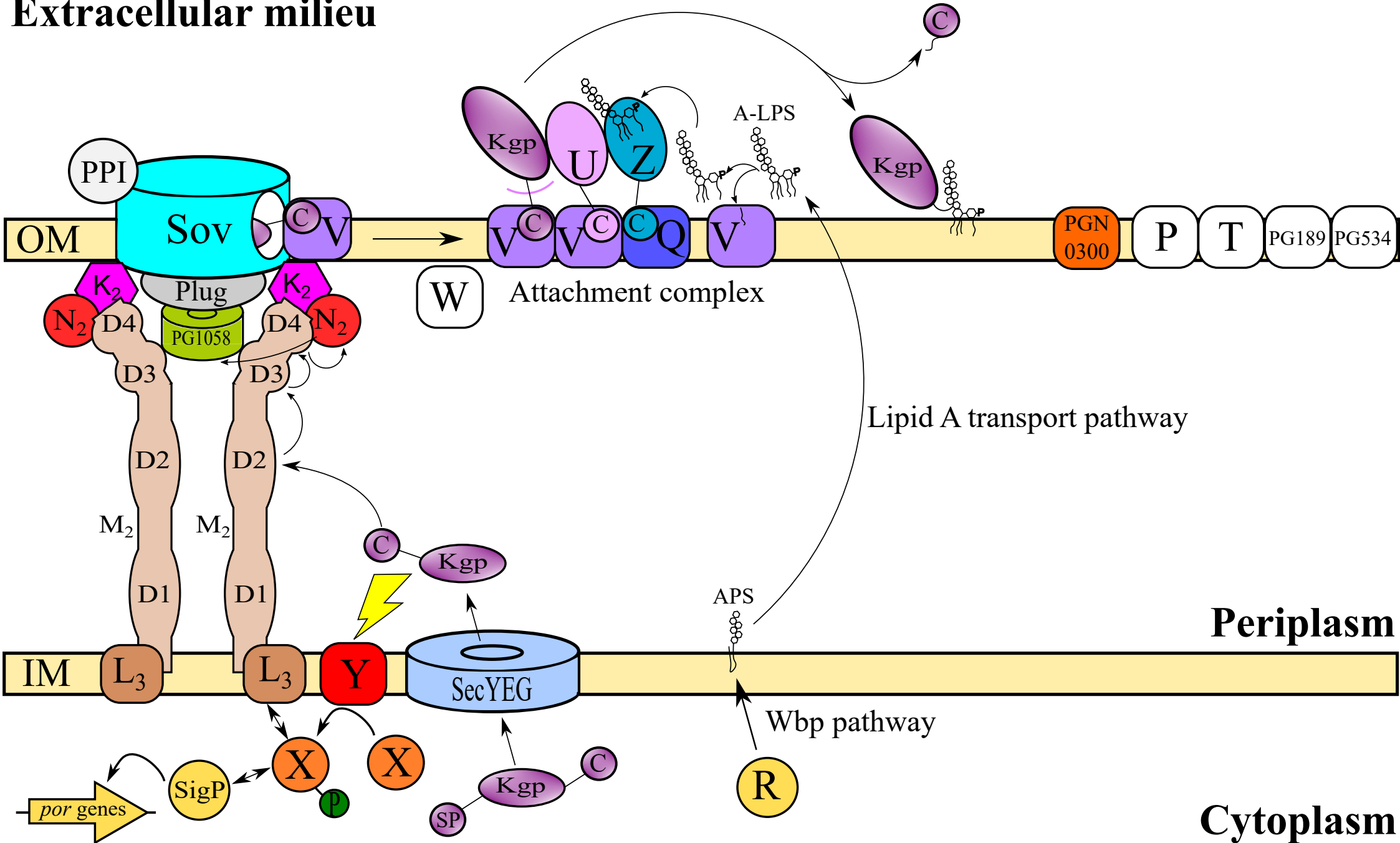


Figure 2 (on next page)

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.

- Bacteroidetes classes**
- Saprospiria
 - Chitinophagia
 - Sphingobacteria
 - Bacteroidia
 - Cytophagia
 - Unclassified
 - Flavobacteria

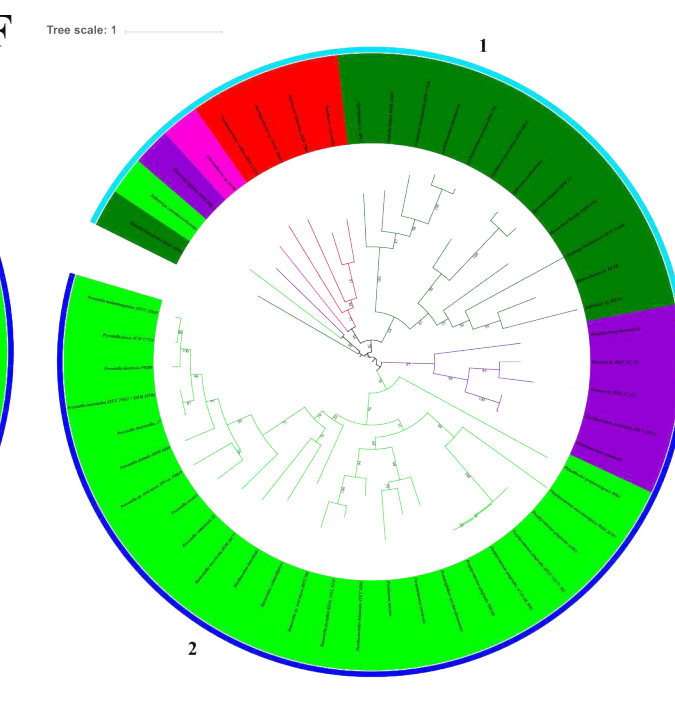
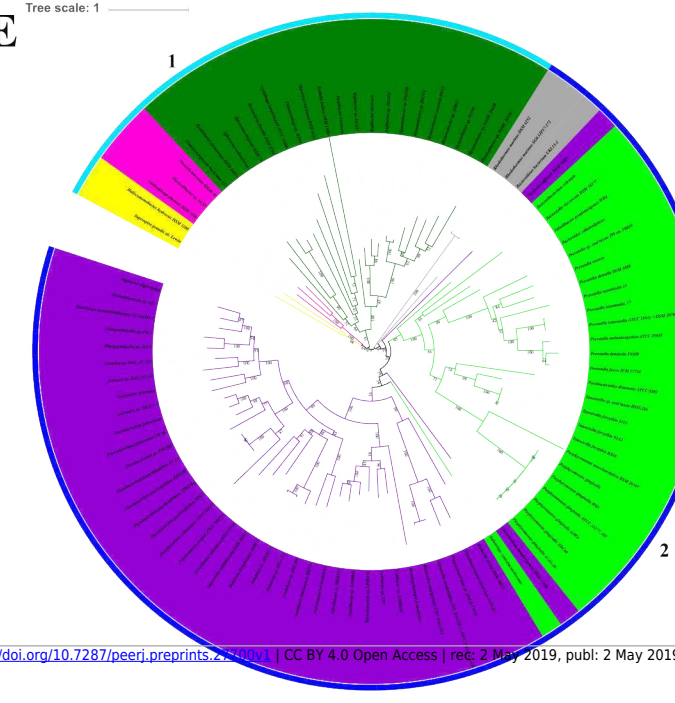
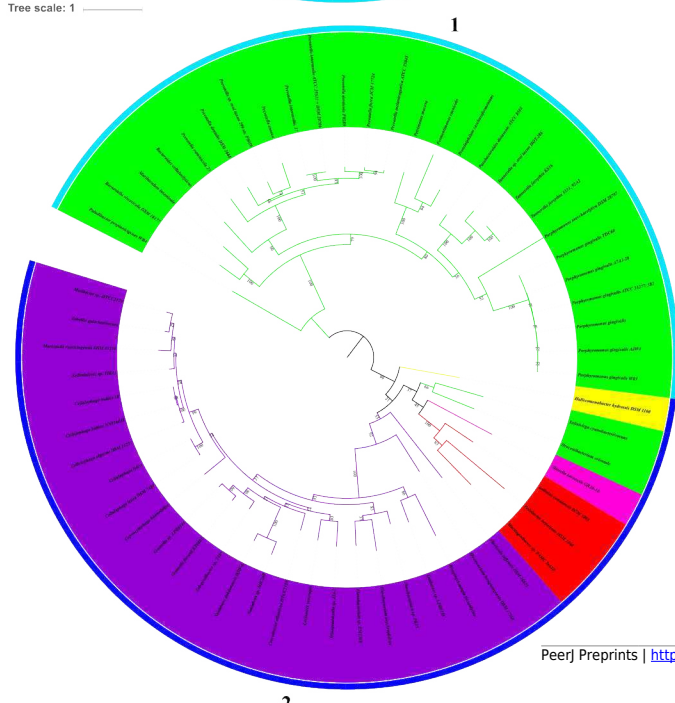
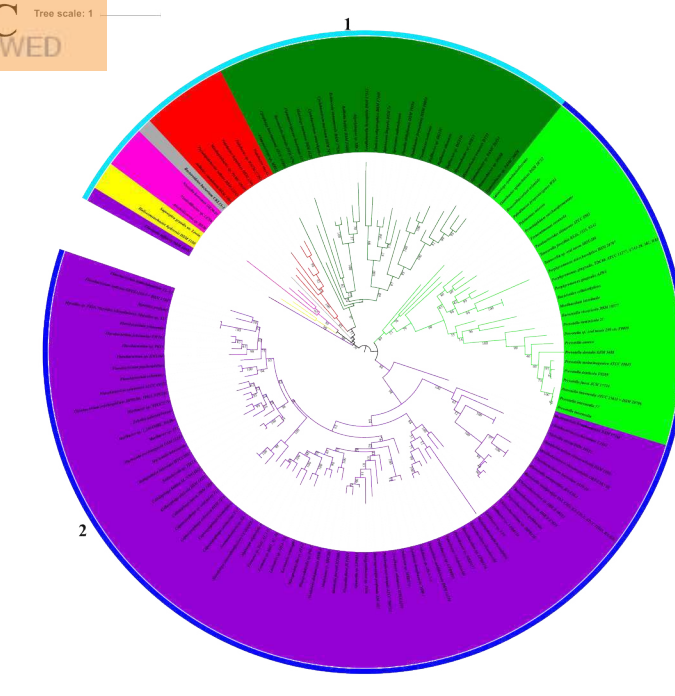
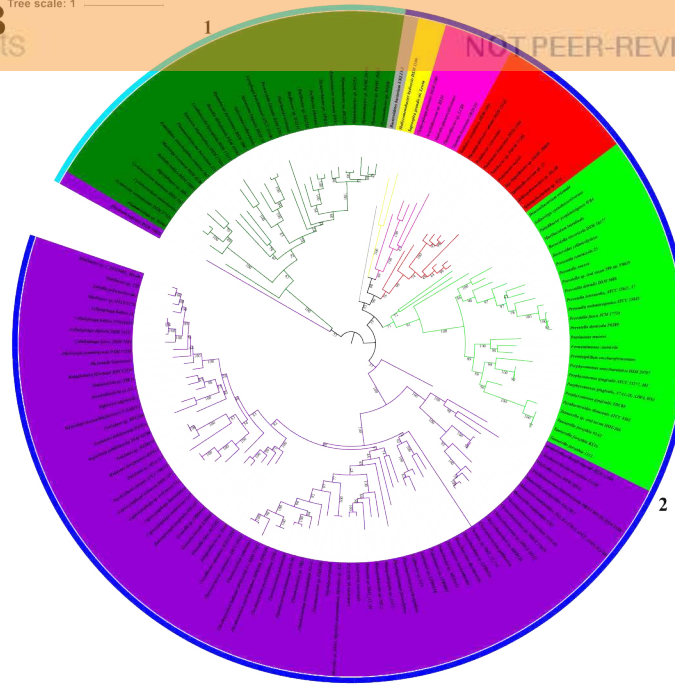
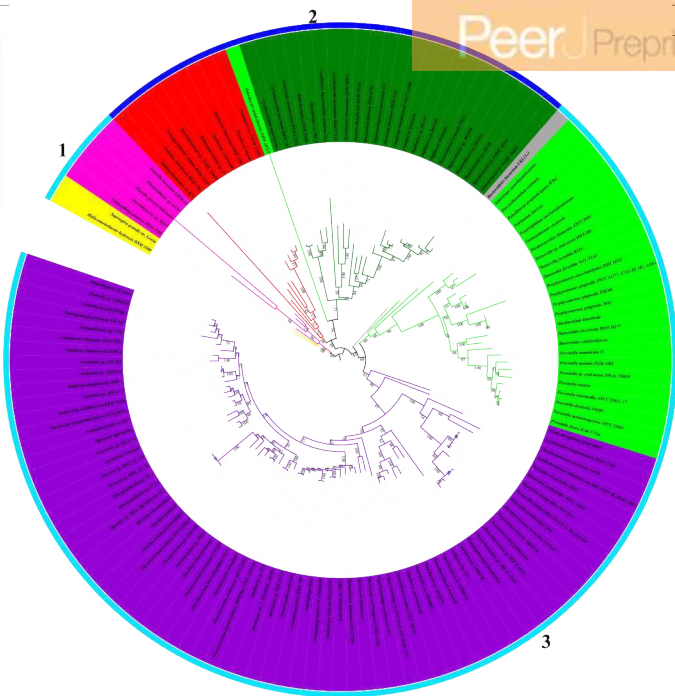


Figure 3 (on next page)

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.

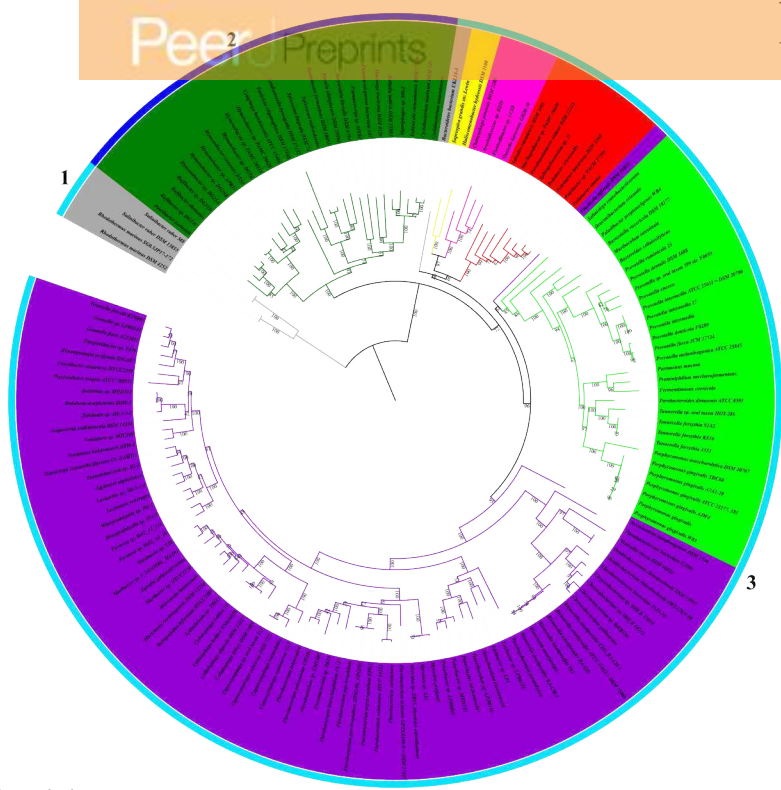
A

Tree scale: 1

PeerPreprints

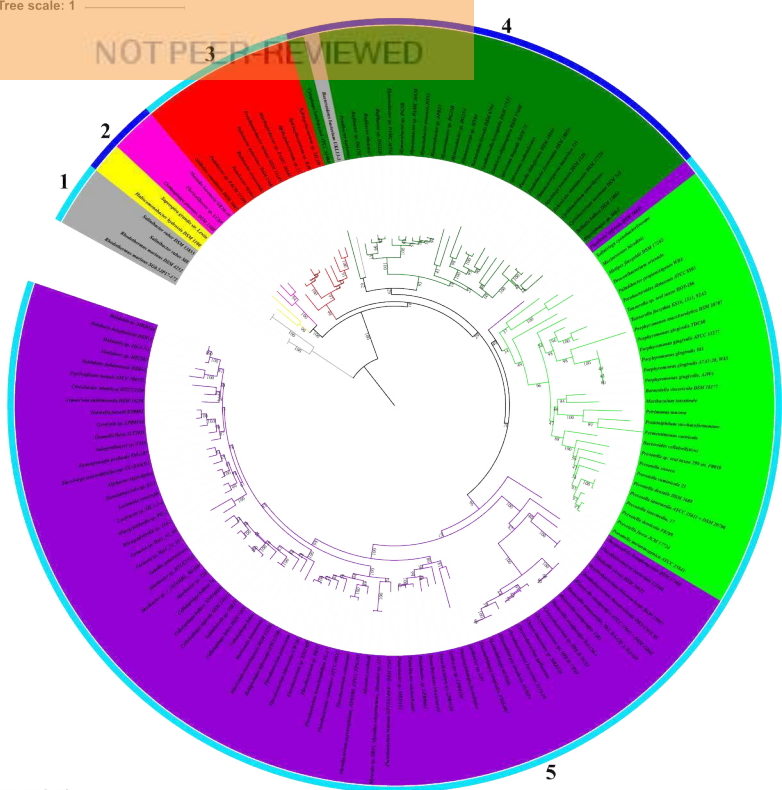
Bacteroidetes classes

- Cytophagia
- Unclassified
- Saprospira
- Chitinophagia
- Sphingobacteriia
- Flavobacteriia
- Bacteroidia

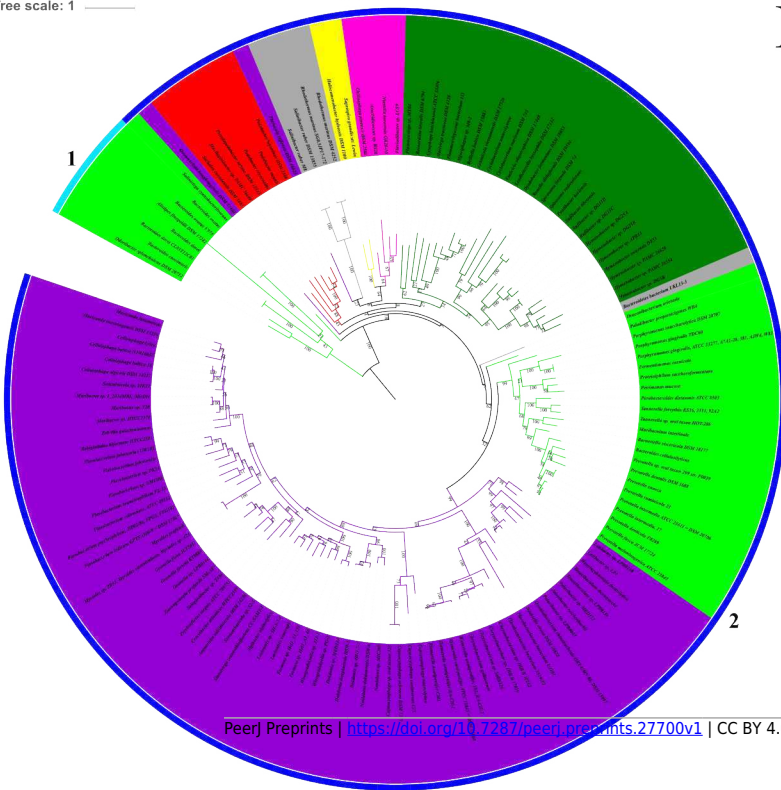
**B**

Tree scale: 1

NOT PEER REVIEWED

**C**

Tree scale: 1

**D**

Tree scale: 1

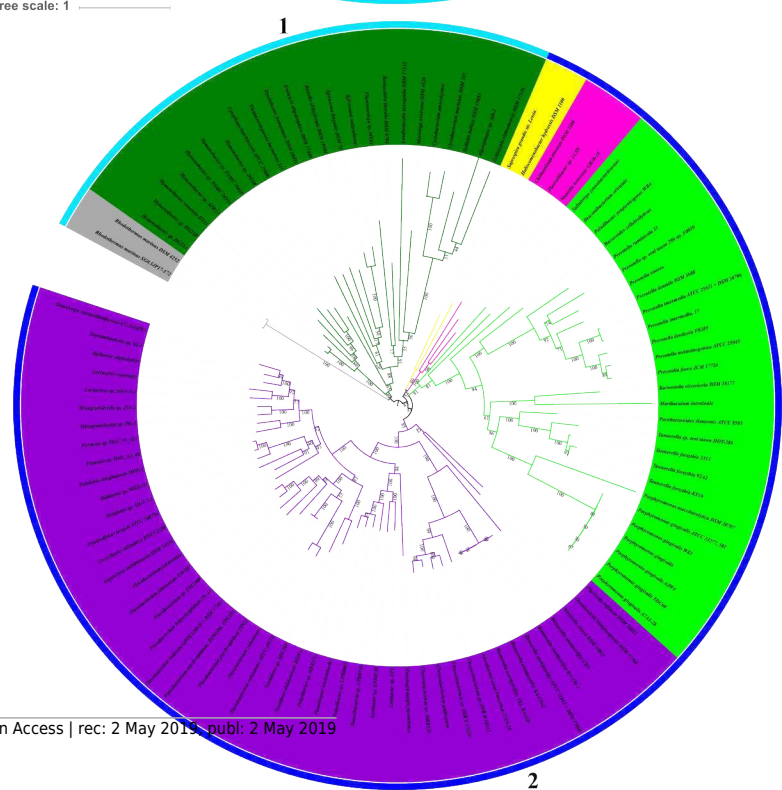


Figure 4 (on next page)

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

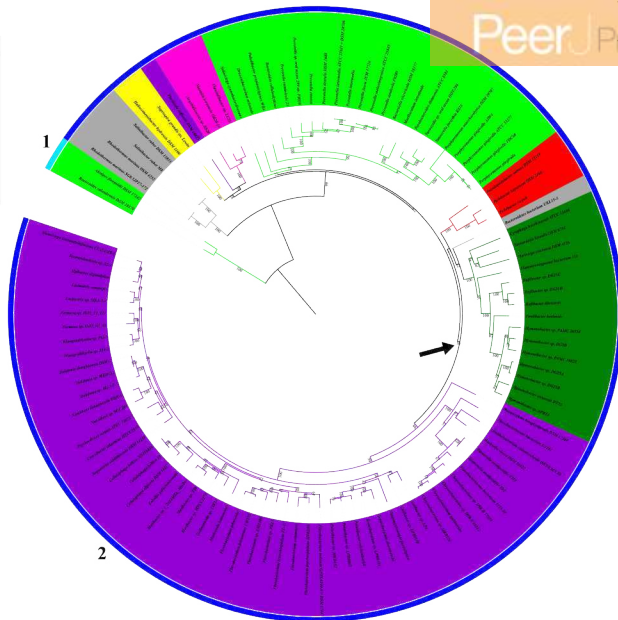
(A) ML tree of PG0534. (B) ML tree of PorL. (C) ML tree of PorQ. (D) ML tree of PG0162. (E) ML tree of PGN0300. The nodes leading to clusters of major classes with low bootstrap support values (less than 70) are pointed by the black arrows.

A

Tree scale: 1

Bacteroidetes classes

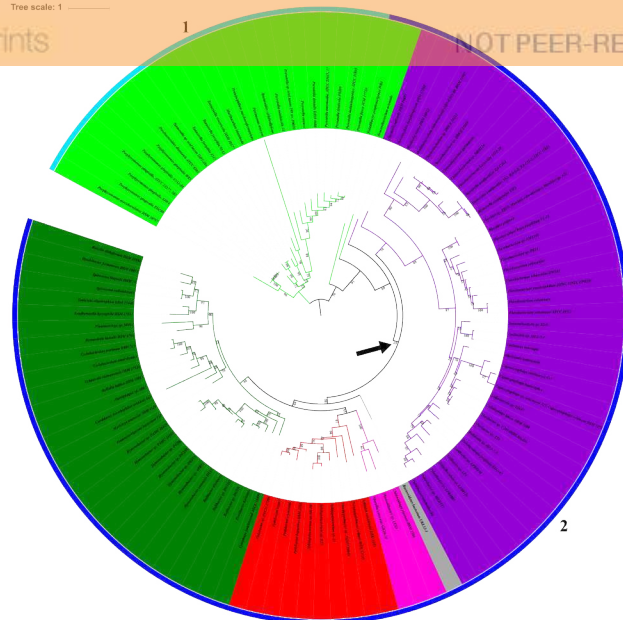
- Bacteroidia
- Saprospiria
- Flavobacteria
- Chitinophagia
- Sphingobacteria
- Unclassified
- Cytophagia



Peer Preprints

B

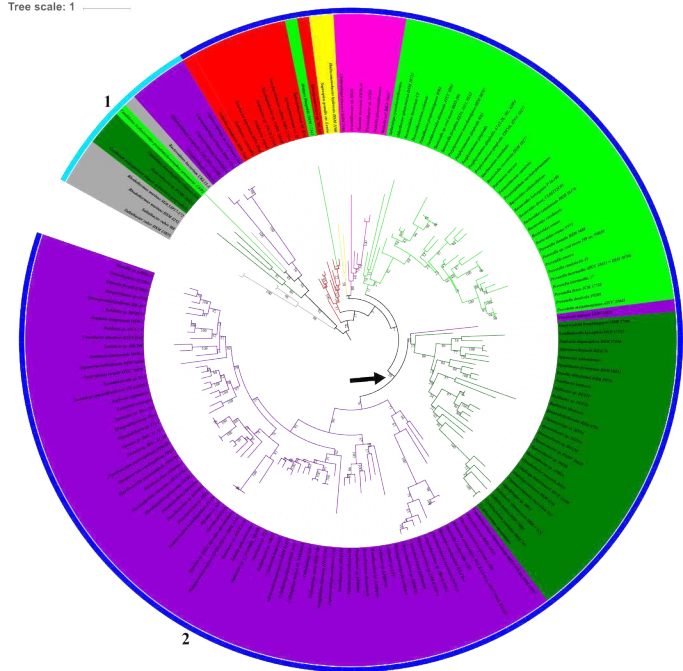
Tree scale: 1



NOT PEER-REVIEWED

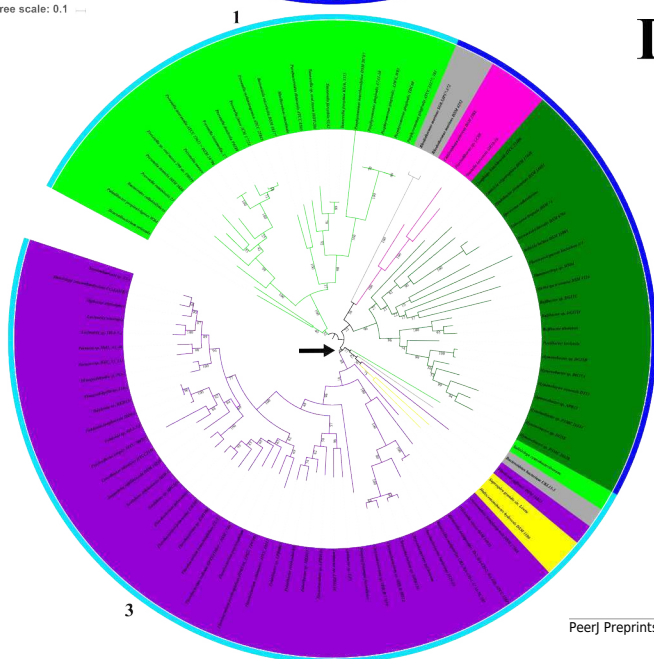
E

Tree scale: 1



C

Tree scale: 0.1



D

Tree scale: 1

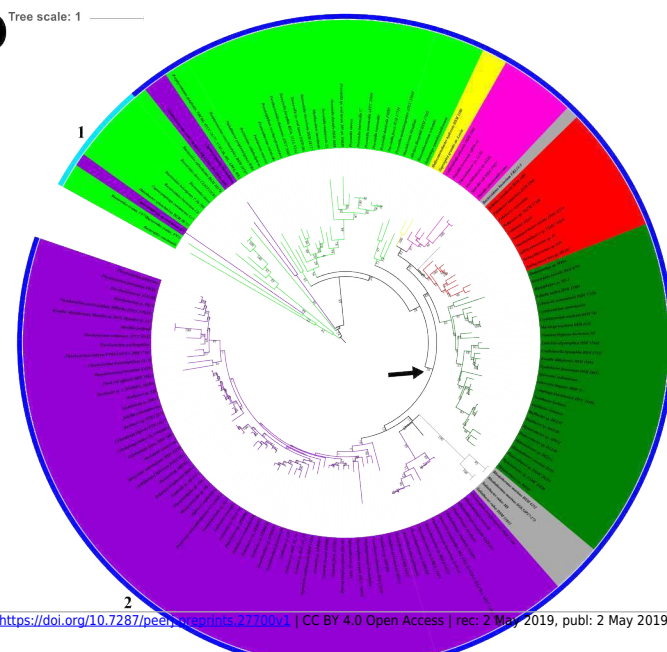


Figure 5 (on next page)

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

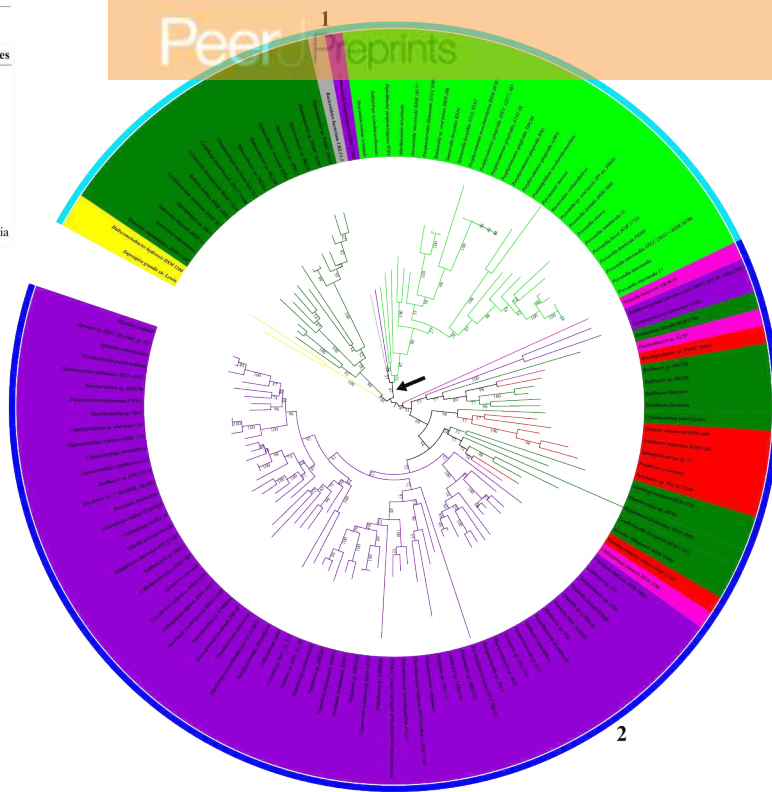
(A) ML tree of PorP. (B) ML tree of PorW. (C) ML tree of PG1058. (D) ML tree of PorY. The nodes leading to clusters of major classes with low bootstrap support values (less than 70) are pointed by the black arrows.

A

Tree scale: 1

Bacteroidetes classes

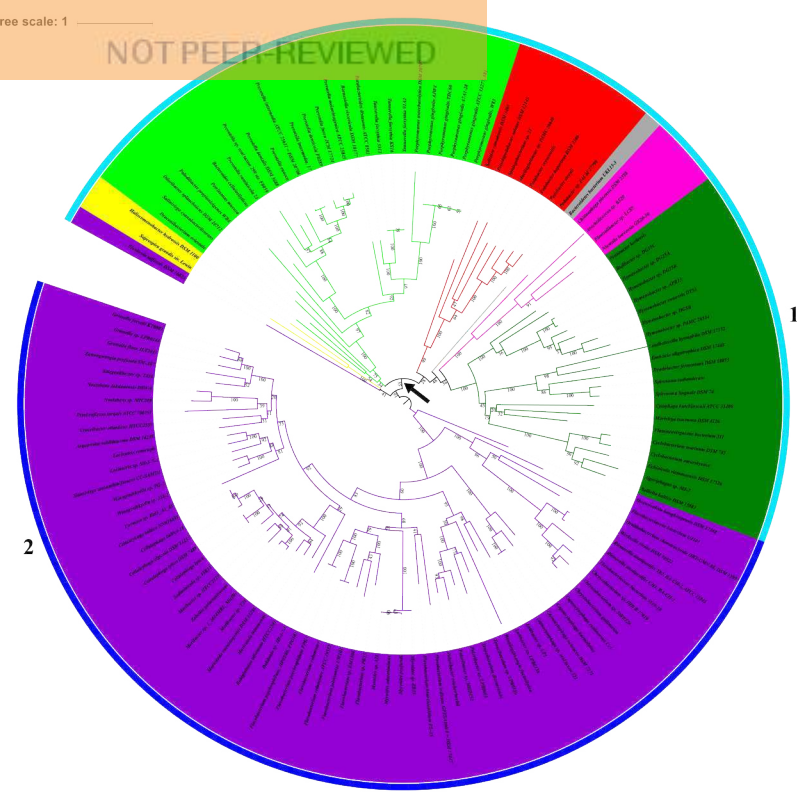
- Bacteroidia
- Flavobacteria
- Unclassified
- Cytophagia
- Saprospiria
- Chitinophagia
- Sphingobacteria



B

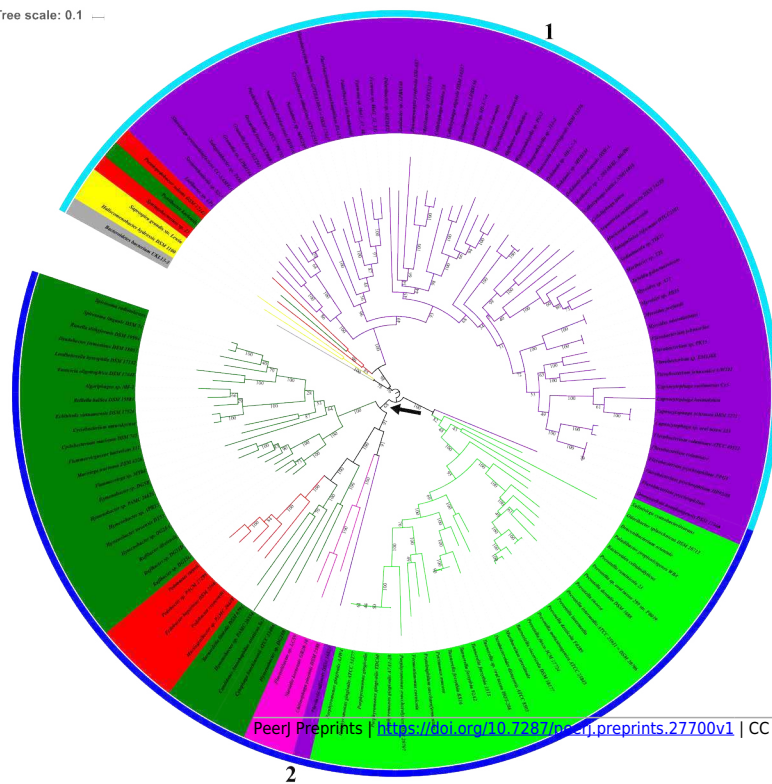
Tree scale: 1

NOT PEER-REVIEWED



C

Tree scale: 0.1



D

Tree scale: 1

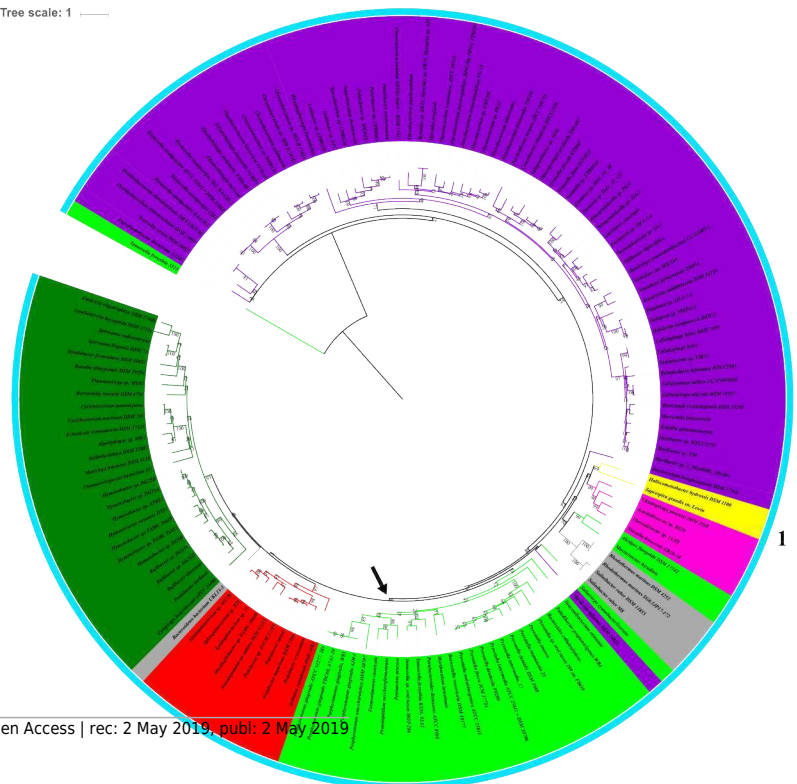


Figure 6 (on next page)

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.

Bacteroidetes classes

- Sphingobacteriia
- Bacteroidia
- Flavobacteriia
- Cytophagia
- Unclassified
- Chitinophagia
- Saprospiria

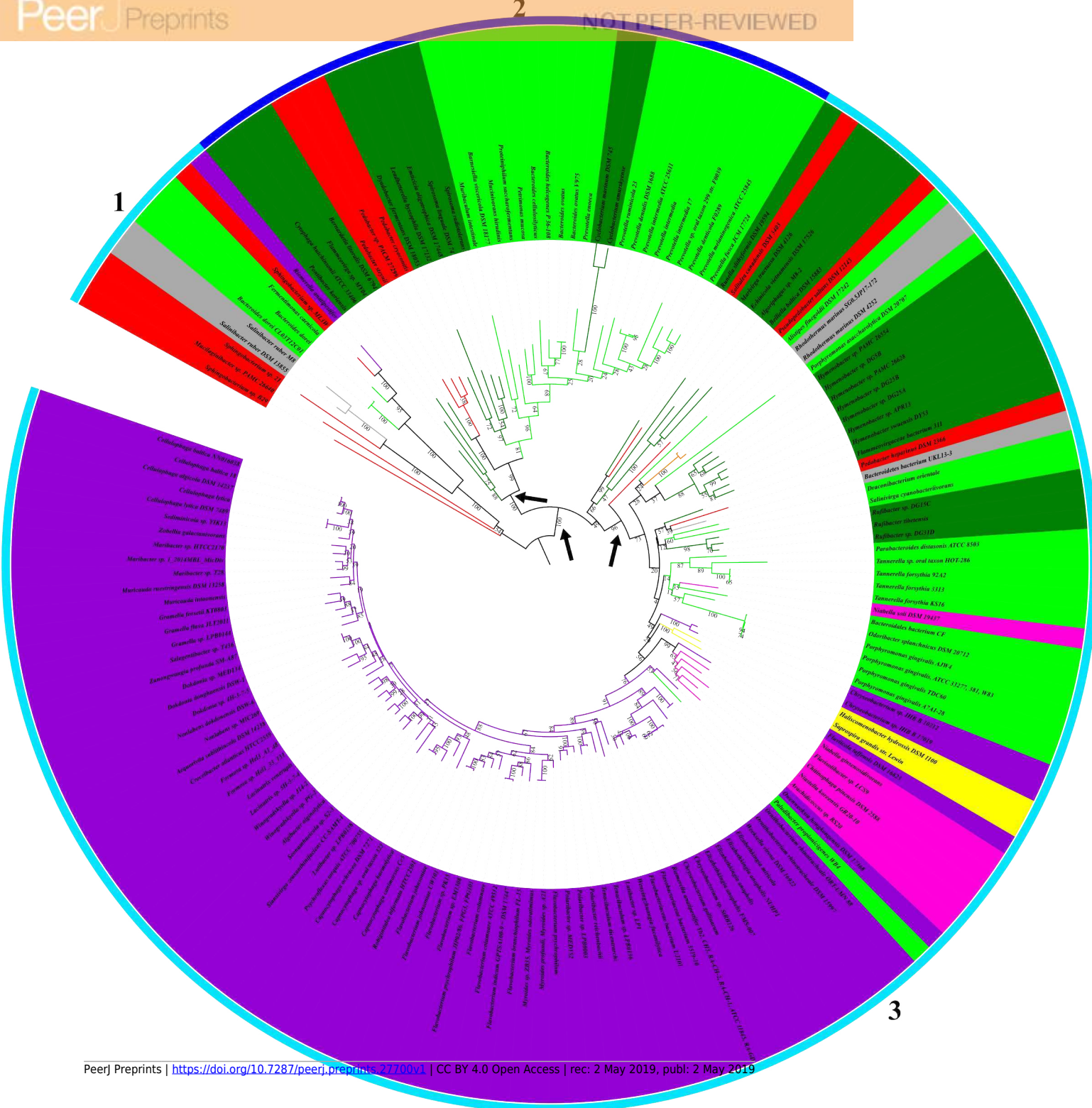


Figure 7 (on next page)

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.

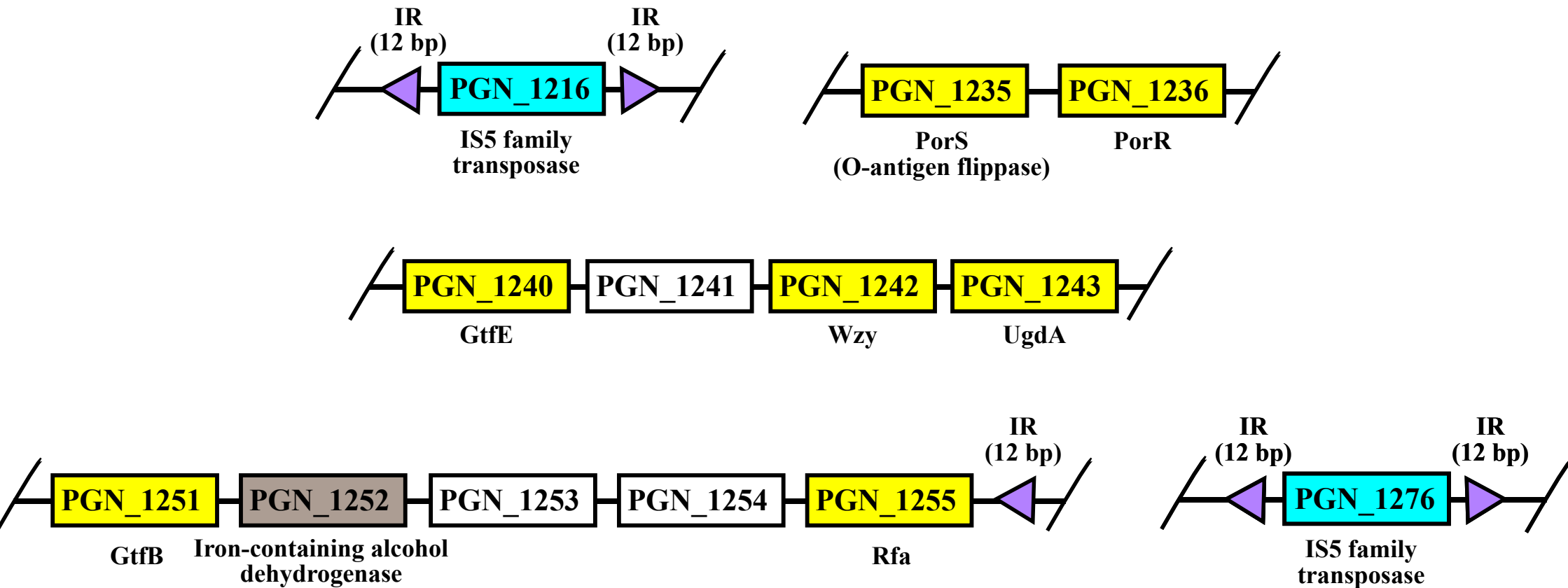


Figure 8 (on next page)

The additional species from Chitinophagia, Saprospira, 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	Soy	PorT	PorU	PorV	PorW	PorX	PorY	PorZ	PG0162	PG0189	PG0534	PG1058	PGN0300
<i>Haliscomenobacter hydrossis</i> DSM1100	Regulation	Unknown	Translocation Energetic	Translocation Energetic	Unknown	Modification	Regulation	Translocation Energetic	Unknown	Modification	Modification	Unknown	Regulation	Regulation	Modification	Regulation	Unknown	Unknown	Translocation Energetic	Regulation
<i>Saprospira grandis</i> str. Lewin	Regulation	Unknown	Translocation Energetic	Unknown	Unknown	Modification	Regulation	Translocation Energetic	Unknown	Modification	Modification	Unknown	Regulation	Regulation	Modification	Regulation	Unknown	Unknown	Translocation Energetic	Regulation
<i>Niastella koreensis</i> GR20-10	Regulation	Translocation Energetic	Translocation Energetic	Translocation Energetic	Unknown	Modification	Regulation	Translocation Energetic	Unknown	Modification	Modification	Unknown	Regulation	Regulation	Modification	Regulation	Unknown	Unknown	Translocation Energetic	Regulation
<i>Flavisolibacter</i> sp. LCS9	Regulation	Translocation Energetic	Translocation Energetic	Unknown	Unknown	Modification	Regulation	Translocation Energetic	Unknown	Modification	Modification	Unknown	Regulation	Regulation	Modification	Regulation	Unknown	Unknown	Translocation Energetic	Regulation
<i>Niabella ginsenosidivorans</i>	Regulation	Unknown	Translocation Energetic	Unknown	Unknown	Unknown	Regulation	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Regulation	Unknown	Unknown	Unknown	Regulation
<i>Arachidicoccus</i> sp. BS20	Regulation	Unknown	Translocation Energetic	Unknown	Unknown	Unknown	Regulation	Translocation Energetic	Unknown	Modification	Modification	Unknown	Unknown	Regulation	Unknown	Regulation	Unknown	Unknown	Unknown	Regulation
<i>Chitinophaga pinensis</i> DSM 2588	Regulation	Translocation Energetic	Translocation Energetic	Unknown	Unknown	Modification	Regulation	Translocation Energetic	Unknown	Unknown	Modification	Unknown	Regulation	Regulation	Modification	Regulation	Unknown	Unknown	Translocation Energetic	Regulation
<i>Bacteroidetes bacterium</i> UKL13-3	Regulation	Translocation Energetic	Translocation Energetic	Unknown	Unknown	Modification	Regulation	Translocation Energetic	Unknown	Unknown	Modification	Unknown	Regulation	Regulation	Modification	Regulation	Unknown	Unknown	Translocation Energetic	Regulation

Legend

- Regulation
- Translocation Energetic
- Modification
- Unknown
- Saprospira
- Chitinophagia
- Unclassified