Author/s:
Veith, PD;Gorasia, DG;Reynolds, EC

Title:
Towards defining the outer membrane proteome of Porphyromonas gingivalis

Date:
2021-02-01

Citation:
Veith, P. D., Gorasia, D. G. & Reynolds, E. C. (2021). Towards defining the outer membrane proteome of Porphyromonas gingivalis. Molecular Oral Microbiology, 36 (1), pp.25-36. https://doi.org/10.1111/omi.12320.

PROFESSOR ERIC C REYNOLDS (Orcid ID : 0000-0002-6618-4856)

Article type : Original Article

Submission to Mol Oral Microbiol

Towards defining the outer membrane proteome of *Porphyromonas gingivalis*

Paul D. Veith[a]*, Dhana G. Gorasia[a], Eric C. Reynolds[a]*

*corresponding authors
e.reynolds@unimelb.edu.au
pdv@unimelb.edu.au

[a]Oral Health Cooperative Research Centre, Melbourne Dental School, Bio21 Institute, The University of Melbourne, Victoria, Australia.

Running Title: Predicted OM proteome of *P. gingivalis*

Acknowledgements

This work was supported by the Australian Government Department of Industry, Innovation and Science Grant ID 20080108 and the Australian Research Council Grant ID DP200100941.

Summary

Porphyromonas gingivalis is a Gram-negative anaerobic pathogen found in subgingival plaque associated with progressive periodontitis. Proteins associated with the outer membrane (OM) of Gram-negative pathogens are particularly important for understanding virulence and for developing vaccines. The aim of this study was to establish a reliable list of outer membrane associated proteins (Omps) for this organism. Starting with a list of 99 experimentally determined Omps, several bioinformatics tools were used to predict a further 52 proteins, leading to a predicted OM proteome of 151 proteins. The tools used included databases of protein families, prediction of OM β-barrels and structural homology. The list includes 33 T9SS cargo proteins, 43 lipoproteins and 66 OM β-barrel proteins with some overlap between categories. The proteins are discussed both in these structural categories as well as their various functions in OM biogenesis, nutrient acquisition, protein secretion, adhesion and efflux. Proteins that were previously shown to be part of large complexes are highlighted and cross reference is provided to a previous major study of protein localisation in *P. gingivalis.* Finally, proteins were also scored according to their level of conservation within the *Bacteroidales* taxon. Low scores were shown to correlate with virulence factors and may be predictive of novel virulence factors.

Keywords: *Porphyromonas gingivalis*; outer membrane; proteome; bioinformatics

Introduction

Porphyromonas gingivalis is a Gram-negative anaerobic bacterium found in subgingival plaque. Numerous studies over more than three decades have demonstrated that *P. gingivalis* is a key pathogen associated with periodontitis in humans (Hajishengallis *et al.* 2011; Holt *et al.* 1988; Kirst *et al.* 2015; Socransky and *et al.* 1998; Ximenez-Fyvie *et al.* 2000). Periodontitis, and in some cases *P. gingivalis* in particular, has been linked to an increased risk of cardiovascular diseases, certain cancers, pre-term birth, rheumatoid arthritis and dementia (Dominy *et al.* 2019; Figuero *et al.* 2020; Haraszthy and *et al*. 2000; Linden *et al.* 2013; Orlandi *et al.* 2020).

P. gingivalis as with other Gram-negative pathogens have numerous virulence factor proteins associated with their outer membranes (OM). Proteins associated with the external surface of the OM are particularly important for virulence as they are needed for adhesion to host tissues, coaggregation with other bacteria to establish biofilms, competitive capture of nutrients and defense against both host and bacterial factors. For *P. gingivalis*, these virulence factors include cargo proteins secreted by the type IX secretion system (T9SS) such as the gingipain proteinases (Guo *et al.* 2010; O'Brien-Simpson *et al.* 2003), fimbrial adhesins (Amano 2010), and OM transporters such as the RagAB complex (Madej *et al.* 2020) and the HmuR/Y system (Smalley and Olczak 2017)

which transport peptides and heme respectively across the OM.

P. gingivalis is the most-studied species of the *Bacteroidetes* and therefore serves as a model organism for this unique phylum. With regard to OM proteins, *P. gingivalis* is the major model for the study of the T9SS (Lasica *et al.* 2017; Veith *et al.* 2017) and also for type V pili (Xu *et al.* 2016). An early report of the OM proteome of *P. gingivalis* identified the most abundant Omps in this species and demonstrated that most exhibited a pyroglutamate at their matured N-terminus (Veith *et al.* 2002), a finding that was later applied to the whole *Bacteroidetes* phylum and termed the "Q-rule" (Bochtler *et al.* 2018). The major structural proteins appear to be the abundant OmpA-like proteins Omp40 and Omp41 which have the unusual property of forming disulphide-bonded heterodimers (Veith *et al.* 2001). Unlike most Proteobacteria, *P. gingivalis* and related species naturally secrete pure outer membrane vesicles (OMVs). The analysis of these have proven to be a very effective means of localising proteins to the OM and periplasm (Friedrich *et al.* 2015; Veith *et al.* 2015; Veith *et al.* 2014; Veith *et al.* 2018).

The localisation of proteins to the OM is confounded by multiple issues. OM isolation generally involves cell lysis followed by membrane separation, sometimes after the sedimentation of both membranes. The OM can be separated from the inner membrane (IM) by density gradient centrifugation since the OM is denser (Schnaitman 1970), or by differential solubilisation with detergent (Filip *et al.* 1973), since the OM is less soluble in certain detergents. Both types of techniques tend to assume that membranes are homogeneous and that all OM-associated proteins (Omps) have the same properties. However, the OM is not homogeneous since many Omps are clustered together in "OM islands" (Rassam *et al.* 2015) which will form OM fragments of different density compared to the protein-poor regions of the OM. Furthermore, while some Omps are resistant to being extracted with detergents others are extracted more easily (Chen *et al.* 2011). Another method for Omp localisation is the use of OMVs in certain species as mentioned above. The major advantage of using natural OMVs is that they do not come in contact with the cytoplasm or IM, strongly limiting false positives due to artefactual interactions between proteins from these compartments and OM components (Veith *et al.* 2014). A disadvantage of using OMVs is the difficulty in detecting Omps involved in protein complexes that span the cell envelope since they are not incorporated into OMVs.

Likewise, the prediction of Omps using bioinformatics is also challenging. This is especially true for understudied phyla such as the *Bacteroidetes* where there are many proteins which do not have functionally or structurally characterised homologs. Recently, a genome-wide prediction of protein

localisation in *P. gingivalis* was conducted resulting in the prediction of 104 Omps plus 22 proteins secreted by the T9SS in strain W83 (Gabarrini *et al.* 2020).

In this study, we start with a list of experimentally-verified Omps from our previous study (Veith *et al.* 2014) and perform extensive bioinformatic searches to predict additional Omps. In this context we define Omps as those proteins expected to be localised to a hypothetically pure OM fraction and includes proteins that are attached to the extracellular or periplasmic sides of the OM but excludes proteins that are more closely associated with the IM. Using this strategy, we present a predicted OM proteome comprising 151 proteins and discuss their roles in OM biogenesis, nutrient acquisition, protein secretion, efflux and adhesion.

Methods

Conserved domains

Conserved domains were found by performing a batch search of the *P. gingivalis* W83 sequence database using the NCBI conserved domains database search tool in June 2020 (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) (Lu *et al.* 2020).

BLAST searches

Where putative OM proteins lacked an N-terminal signal peptide, protein BLAST was conducted with the taxonomy restricted to *P. gingivalis*. This often revealed that the signal peptide had been missed due to misidentification of the Met start site in the genome. The correct N-terminal signal was therefore obtained from further along the protein sequence or else translated from the upstream DNA sequence. PSI-BLAST searches (1 iteration) were also conducted for all proteins in **Table S1** with the taxonomy set to *Bacteroidales.* The number of protein hits obtained with the PSI threshold set to 0.005, sequence coverage >85% and identity >25% is provided in **Table S1** as an indication of how prevalent the protein is, and hence whether it is likely to perform a common function or a specialised function. BLAST searches were also performed with taxonomy restricted to *P. gingivalis* strain W83 to find paralogs.

Prediction of T9SS cargo

T9SS cargo proteins having a type A CTD were as previously predicted (Veith *et al.* 2013) with the following exceptions. PG0769, PG0787 and PG1326 were removed from the list due to the absence of N-terminal signal peptides. PG1035 was added to the list due to it having a type B CTD as

predicted by the conserved domain TIGR04131 (Kulkarni *et al.* 2017). PG0410 and PG1548 were added as predicted by the conserved domain describing type A CTDs, TIGR04183.

Prediction of OM β-barrel proteins using TMBB

OM β-barrel predictive scores were obtained using the Freeman-Wimley analysis tool obtained from the Wimley Laboratory (http://www.tulane.edu/~biochem/WW/Barrel.html). The program was used without considering the SignalP predictions (Freeman *et al.* 2011). Proteins with a TMBB score of >0.8 were considered to be putative OM β-barrel proteins.

Prediction of OM β-barrel proteins using structural homology

Putative OM-associated proteins, especially β-barrel proteins were analysed by Phyre-2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) (Kelley *et al.* 2015). Exclusive and high confidence matches (typically >95% confidence) to OM β-barrel proteins was considered a strong confirmation of OM β-barrel structure. Notes are provided in **Table S1** regarding whether the match was to a complete or partial β-barrel structure. In many cases, the matched protein provided in **Table S1** is just one of several equivalent hits of excellent confidence.

Lipoprotein prediction

The lipoproteins shown in **Table S1** were either predicted from conserved domains or else known from the literature, or else experimentally localised using our OMV proteome approach (Veith *et al.* 2014). From the OMV proteome, lipoproteins with a Mascot score <60 were removed due to the possibility of them being localised to the IM.

Criteria for inclusion

The inclusion criteria for T9SS cargo and lipoproteins were already given above. Proteins previously localised to the OM (vesicle membrane) with a Mascot score <100 (Veith *et al.* 2014) were only accepted if supported by additional evidence such as TMBB score, structural homology to OM β-barrels or having a conserved domain belonging to an Omp family. This resulted in the exclusion of PG2029 and PG2050. Proteins that had not been experimentally localised to the OM (Veith *et al.* 2014) needed to be experimentally verified in other studies or else predicted by multiple bioinformatic methods comprising TMBB, conserved domains, structural homology or manual sequence analysis. Manual analysis involved finding at least four putative β-strands each consisting of a string of eight or more residues with alternating hydrophobic side-chains.

Native MW profiles

The native MW profiles (**Fig S1**) were constructed using data from the comprehensive blue native electrophoresis study of whole cell lysates. A manuscript describing this work has been submitted for publication, and a preprint that includes the data has been made publicly available (Gorasia *et al.* 2020).

Results and Discussion

As a starting point, we filtered our previous list of 109 experimentally identified Omps (Veith *et al.* 2014) to remove ten low scoring Omps (see methods) resulting in a list of 99 Omps divided into 30 T9SS cargo proteins, 29 lipoproteins and 40 other OM-associated proteins This list was extended by evidence from the literature, looking for paralogs of these proteins and by screening proteins for belonging to Omp-related families or having a high propensity to form an OM β-barrel structure. The assignment of many Omps was also confirmed by structure prediction tools, especially where the tool strongly predicted an OM β-barrel structure. Using this approach, the Omp list was extended to contain 151 proteins comprising 33 T9SS cargo proteins including one lipoprotein, 42 additional lipoproteins including 10 type V pilins and two β-barrel proteins, 64 additional β-barrel proteins, and 12 other OM-associated proteins (**Table S1**). Where appropriate, reference is made to the comprehensive blue native electrophoresis interactome of *P. gingivalis* strain ATCC 33277 which provides experimental evidence for protein complex formation (**Table S1, Fig S1**)(Gorasia *et al.* 2020). The list of conserved domains and the list of TMBB scores for each protein in the *P. gingivalis* W83 database is provided in **Table S2** and **Table S3** respectively.

OM biosynthesis

The OM is composed of lipopolysaccharide (LPS), phospholipids, lipoproteins and integral membrane proteins which mostly have a β-barrel architecture. Systems involved in the synthesis of the OM include the related β-barrel assembly machines BAM and TAM for incorporating β-barrel proteins into the OM (Konovalova *et al.* 2017; Stubenrauch and Lithgow 2019), the Lol pathway for incorporating lipoproteins into the OM (Konovalova and Silhavy 2015), the Lpt system for incorporating LPS (Okuda *et al.* 2016) while the mechanisms for transporting and incorporating phospholipids into the OM are poorly understood (Shrivastava and Chng 2019). The *P. gingivalis* proteins predicted to perform these roles are presented in **Fig 1**. *P. gingivalis* was found to have three BamA-related proteins (**Table S1**). PG0191 matches very well to the BamA family (COG4775, Expect value = 4E-93) over its entire length which includes the β-barrel domain and the five periplasmic POTRA domains and is therefore designated BamA. *E. coli* BamA is known to associate with four lipoproteins (BamB-D), with BamD being strongly conserved throughout Bacteria (Konovalova *et al.* 2017). The likely *P. gingivalis* BamD is PG1215 (Gabarrini *et al.* 2020)

but its association with BamA has not yet been confirmed. BamA was previously identified to comigrate with the predicted lipoprotein PG0188 during Blue-Native PAGE (Glew *et al.* 2014). PG0188 does not match to any of the Bam protein families and appears to be a novel BamA-associated lipoprotein. The other BamA-related proteins PG0980 and PG2095 are predicted lipoproteins and are paralogs of each other sharing 28% sequence identity. The gene encoding PG2095 is sandwiched between genes encoding proteins with TamB domains suggesting that PG0980 and PG2095 may be equivalent to TamA (**Fig 1**). Only PG2095 appeared to complex with the TamB proteins (**Fig S1**). Structural homology searches using Phyre-2 confirm full-length matches to TamA inclusive of the three POTRA domains. The presence of the lipoprotein signal suggests that the string of POTRA domains curl back around to the OM (**Fig 1**). The same *P. gingivalis* Bam and Tam proteins were also recently predicted by others (Shoji *et al.* 2020).

P. gingivalis appears to have a complete LPS transport system including the OM components LptD (PG1405) and LptE (PG0147) which appear to form the expected complex (**Fig S1**). Close to the full length of PG1405 (residues 77-908) was matched to LptD using Phyre-2. LptD comprises a C-terminal 26-stranded β-barrel domain and an N-terminal domain which forms a periplasmic bridge with LptA and LptC (Okuda *et al.* 2016). LptE is a lipoprotein which plugs part of the LptD barrel and was also matched by Phyre-2 (**Fig 1**). In addition to LPS, some *P. gingivalis* strains produce an exopolysaccharide capsule. In *E. coli*, the group 1 capsular polysaccharides are exported by Wza, an octameric Omp that is rare in its ability to form an alpha-helical trans-OM domain (Dong *et al.* 2006; Sande *et al.* 2019). In *P. gingivalis*, PG0437 matches strongly to the Wza family (**Table S1**) and the adjacent genes, PG0435-6 encode for the other two members of the export apparatus, Wzb and Wzc (**Fig 1**). PG0437 was found at a high native MW consistent with its oligomerisation to form a pore (**Fig S1**).

Nutrient Acquisition

Omps such as TonB-dependent receptors (TDRs) and porins are usually involved in the uptake of nutrients. Members of the TDR family were identified by PSI-BLAST searches of known members, and the presence of TDR domains such as “Plug”, “ligand_gated_channel” and “TonB_Dep_Rec”. Furthermore, PSI-BLAST searches of proteins matching the “OMP_b-brl_3” family resulted in identifying these as also belonging to the TDR family. Including the PG0534 TDR associated with the T9SS (see below), 15 TDRs were predicted based on the presence of a large C-terminal β-barrel, and an N-terminal plug domain including a putative TonB box for interaction with the TonB protein (**Table S1, Table 1, Fig 2B**). Analysis of conserved domains and multiple sequence alignments demonstrated that twelve of these contained a “carbopepD_reg_2” domain (Cpdr)

followed by a plug domain N-terminal to the β-barrel domain (**Table 1**, **Fig 2A**). Three were missing the Cpdr domain (**Table 1**). Multiple alignments showed that ten sequences exhibited a conserved region of ~32 aa in the plug domain (**Fig 2C**) while the remaining five did not. The function of the Cpdr domain is not known but is located on the periplasmic side of the barrel in the RagA structure (Madej *et al.* 2020). Structural homology searches with Phyre-2 for most *P. gingivalis* TDRs matched best to the ferripyoverdine receptor (FpvA) and the ferrioxamine b transporter (FoxA) of *Pseudomonas aeruginosa*. Phyre-2 matched the Cpdr domain to the N-terminal "signalling domain" of both of these TDRs, however conserved domain searches of FpvA and FoxA match to the STN family (Smart00965) and not to Cpdr. This signalling domain of FpvA and FoxA is very flexible and located in the periplasm (Brillet *et al.* 2007; Greenwald *et al.* 2009; Josts *et al.* 2019) as can be seen in the model of PorF (**Fig 3**) and appears to be structurally homologous to the Cpdr domain.

While a few TDRs exhibited a native MW between 130-170 kDa which may correspond to monomer plus detergent, most of the TDRs exhibited a native MW in the 300's, probably reflecting a state of oligomerisation. They are likely to be dimers since RagA is also known to dimerise in a complex containing two subunits of RagA and two subunits of RagB (Madej *et al.* 2020). Furthermore, the three putative porins (see below) which are likely to be trimers based on homology, exhibited a similar native MW to the TDRs, therefore if the TDRs were also trimers they would have migrated at a larger size. PG1414 exhibited a native MW of 430-610 kDa suggesting that it may complex with another protein (**Fig S1**).

Of the 15 predicted TDRs, only two have a definitive function. RagA (PG0185) was shown to complex with the RagB lipoprotein and was determined to be a peptide transporter based on structural studies, binding studies and growth experiments (Glew *et al.* 2014; Madej *et al.* 2020; Nagano *et al.* 2007). RagB serves as the initial peptide binder and also functions as a hinged lid on the extracellular side of RagA (**Fig 2D**). RagA/B are very abundant proteins consistent with their important role in acquiring peptides for energy and growth, since *P. gingivalis* cannot grow on carbohydrates. The other TDR system of well characterised function is HmuR (PG1552) and its cognate lipoprotein HmuY (PG1551) which together capture heme on the cell surface and transport it into the periplasm (Smalley and Olczak 2017). While the RagB lipoprotein is firmly bound to RagA, the HmuY lipoprotein is more abundant than HmuR and appears to be mostly free on the OM and enriched in OMVs (Veith *et al.* 2018). A similar situation exists for the IhtA TDR (PG0668) and its more abundant cognate lipoprotein IhtB (PG0669) (Veith *et al.* 2014; Veith *et al.* 2018). The mechanism of secretion for these lipoproteins across the OM has not been determined

and cannot be easily elucidated by bioinformatic approaches since there is no known general pathway for this to occur.

Nutrient acquisition on the cell surface begins with T9SS cargo proteins. Seven of these are proteases including the abundant gingipains RgpA, RgpB and Kgp whose proteolytic activity and contribution to virulence has been extensively characterised (Guo *et al.* 2010; O'Brien-Simpson *et al.* 2003). The proteases cleave host proteins into peptides which can be taken up by the RagAB transporter.

Secretion Systems

The major secretion system of *P. gingivalis* is the T9SS, and now includes about 20 components (Gorasia *et al.* 2020; Naito *et al.* 2019). In contrast to other complex secretion systems in Gram-negative bacteria which generally have a single integral Omp to serve as the translocon, the T9SS has seven OM β-barrel proteins including the Sov translocon which is the largest known OM β-barrel produced by a single polypeptide chain (Lauber *et al.* 2018). PorG and PorT (Nguyen *et al.* 2009) have eight predicted strands; PorP, PorQ and PorV each have 14 strands similar to *E. coli* FadL (van den Berg *et al.* 2004); and PorF is a predicted TDR with 22 strands **(Fig 3)**. The structure of PorV was determined in complex with Sov (Lauber *et al.* 2018). The T9SS also has three OM-associated lipoproteins, PorE, PorK and PorW (Gorasia *et al.* 2020; Gorasia *et al.* 2016; Heath *et al.* 2016; Sato *et al.* 2013)(**Table S1**).

P. gingivalis also has a predicted Type IVb system, expected to be involved in conjugation. The genes encoding this system are all located on a mobile genetic element (conjugative transposon) spanning from PG1473 to PG1490 and annotated with the prefix *tra* for transfer, originally by their homology to the genes first sequenced in a Bacteroides conjugative transposon named CTnDOT (Bonheyo *et al.* 2001). Most of the encoded proteins are now understood to be part of the Type IVb secretion system which is responsible for the conjugative transfer of DNA (Guglielmini *et al.* 2014). The OM complex is predicted to include domains from TraN and the TraQ lipoprotein (both related to VirB9) and TraM (related to VirB10). The structure of core complexes demonstrate that VirB10 homologs contribute two trans-OM helices per subunit to produce the OM translocation pore (Chandran *et al.* 2009; Sgro *et al.* 2018). *P. gingivalis* appears to have three different pilin proteins, TraE (related to VirB2) as well as TraI and TraO (both related to VirB5) (Guglielmini *et al.* 2014) and are included in our table since they extend from the OM channel into the extracellular environment. The OM components TraN, TraM and TraO produced a high MW complex,

consistent with a secretion complex, that could be seen in band 2 (**Fig S1**). The expected arrangement of the proteins across the cell envelope is shown in **Figure 4.**

Adhesion

A new type of pilus was recently characterised in *P. gingivalis* and related species. The type V pilins are produced as lipoproteins and polymerised on the cell surface through proteolytic cleavage of the lipidated N-terminal region and donor-strand exchange (Shibata *et al.* 2020; Shoji *et al.* 2020; Xu *et al.* 2016). Three pili have been identified, the major (longer) pili (or fimbriae) composed of Fim subunits, the minor pili composed of Mfa subunits and PG1881 (Nagano *et al.* 2017), which is yet to be fully characterised (**Table S1**). The Fim and Mfa pili are composed of an anchor pilin which retains its lipid anchor, a stalk pilin (FimA and Mfa1) which multimerises to form the stalk and multiple tip pilins. Mfa1 (PG0178) was taken from strain ATCC33277 since the gene is disrupted in strain W83. PG1881 appears not to be associated with anchor or tip pilins. Two Omps, PG2130 (FimX) and PG2131 (PgmA) are encoded by conserved genes located directly upstream of FimA in multiple species (Nagano *et al.* 2012). While expected to play a role in fimbrial biogenesis, a role for these proteins has not yet been demonstrated. In addition, the secretion system for the type V pili has not yet been identified.

Several T9SS cargo proteins are also involved in adhesion. The most abundant are the sequence-related adhesins of the gingipains RgpA and Kgp and the haemagglutinin HagA (O'Brien-Simpson *et al.* 2003).

Efflux

P. gingivalis has six proteins belonging to the TolC family of OM efflux proteins (Gabarrini *et al.* 2020; Inoue *et al.* 2015) (**Table S1**). Members of this family are trimeric with each monomer contributing four β-strands to a 12-stranded β-barrel. The remainder of the protein forms a long α-helical barrel that connects the OM barrel to a "membrane fusion protein" anchored in the IM and associated with one of several types of IM transporters (Du *et al.* 2018). The cargo are very diverse and include toxic substances such as heavy metals and a wide assortment of drugs/antimicrobials that need to be expelled, as well as self-made proteins, antimicrobial peptides and siderophores that are secreted (Du *et al.* 2018). *P. gingivalis* mutants lacking each of the six membrane fusion proteins were shown to be more sensitive to various antibiotics (Inoue *et al.* 2015). The proteins produced by the PG0538 – PG0541 efflux system were highly upregulated in response to excess heme conditions suggesting the system is able to excrete excess amounts of iron or heme (Veith *et al.* 2018).

Small β-barrels

Barrels composed of eight strands are the smallest known OM β-barrels and are sometimes involved in virulence (McClean 2012). Generally, these barrels are too small to transport nutrients, however if they oligomerise they may form large enough channels for this purpose (Sanganna Gari *et al.* 2018). In *P. gingivalis,* 13 proteins were predicted to form 8-stranded β-barrels (**Table S1**). Seven of these, PG0083, PG0189, PG0751, PG1823, PG2041, PG2106 and PG2149 could be matched to each other by PSI-BLAST searches and showed full-length structural homology to 8-stranded β-barrels such as OmpW and the N-terminal domain of OmpA (**Table S1**). The remaining six proteins also modelled well to at least 7 of the 8 β-strands of 8-stranded barrels. In each case, the missing strand could be predicted from inspection of the protein sequence by the presence of at least four alternating hydrophobic residues. The most interesting structural homology match was of PG0409 to CarO, an 8-stranded β-barrel with a glove-like extracellular domain. CarO was shown to transport small molecules such as glycine and ornithine in *Acinetobacter baumannii* (Zahn *et al.* 2015). One of the 8-stranded barrels is PG2130 (FimX) which has a paralog, PG2168 (**Table S1**). PG2168 (22 kDa) was found to form a ~290 kDa complex with the PgmA paralog, PG2167 (53 kDa) suggesting that one or both subunits oligomerise (Glew *et al.* 2014). Similarly, FimX and PgmA also form a putative complex (**Fig S1**). PG2167 and PgmA have the peptidoglycan-binding OmpA-C domain and are therefore likely to reside in the periplasm.

Omp40 and Omp41 form a heterodimer and are predicted to form 8-stranded β-barrels in their N-terminal domains similar to OmpA from *E. coli* and are likely to be important for the structural integrity of the OM as their C-terminal domains bind to peptidoglycan (Pautsch and Schulz 1998; Veith *et al.* 2001).

Two proteins were concluded to be 12-stranded barrels. PG0937 produced a good Phyre-2 match to the nucleoside transporter Tsx which has an elongated cross-section like a keyhole (Ye and van den Berg 2004) and PG1786 matched well to the NanC porin responsible for the uptake of N-acetylneuraminic acid in *E. coli.*

The number of β-strands in the other small predicted barrels could not be assigned. Of these, PG0373 has a predicted mature MW of 22 kDa but was found to be part of a large complex >700 kDa suggesting that it may multimerise and associate with other proteins (**Fig S1**). Many of the small barrels have closely related paralogs such as Omp40 with Omp41, PG2168 with PG2130, PG0217 with PG0218, PG1823 with PG2106 and PG0479 with PG2112 **(Table S1)**.

FadL family

FadL is a family of 14-stranded β-barrels with an N-terminal plug domain. Where characterised these barrels are involved in the import of hydrophobic compounds such as fatty acids (van den Berg 2005). In *P. gingivalis*, five members of the FadL family were identified by matching to conserved domains or by structural homology modelling. In addition to the T9SS components PorV, PorQ and PorP, PG1626 and PG0448 also match to this family. It is interesting that three of the members are components of the T9SS where PorV and PorQ are thought to bind to the T9SS signal (the CTD) (Glew *et al.* 2017). PorP was alone in not having the complete plug domain. All exhibited C-terminal extensions of 6-18 residues beyond the most C-terminal β-strand except for PG1626. PG0448 was interesting as it was found in a large complex >450 kDa (**Fig S1**).

Porins

Porins can be used to describe any OM pore, but the term is also used more specifically to describe protein families usually containing 16-18 β-stands and involved in the passive import of nutrients. No porins have yet been functionally or structurally characterised for *P. gingivalis*. Two *P. gingivalis* Omps (PG0326 and PG1382) have similarity to the Porin_O_P family of anion-selective 16-stranded porins, a prediction supported by Phyre-2 matches to porins such as OmpF and OprP, and a third putative porin, PG1178 was also predicted to form a 16-stranded barrel by Phyre-2 (**Table S1**). All appeared to produce complexes of a similar native size which may correspond to trimers, since most other porins are trimeric (**Fig S1**).

Lipoproteins

OM lipoproteins are functionally and structurally diverse. BamA-associated lipoproteins and the two TamA lipoproteins are involved in the assembly of β-barrel proteins while LptE and Wza are involved in the secretion/assembly of LPS and capsule respectively, as described above (**Fig 1**). This example alone has lipoproteins forming both α-barrels and β-barrels in the OM as well as being located inside β-barrels or in the periplasm (**Fig 1**). The three OM lipoproteins of the T9SS are thought to be located at the periplasmic face of the OM. PorK is part of the large (50 nm) periplasmic rings (Gorasia *et al.* 2020; Gorasia *et al.* 2016); PorE binds to the PorP β-barrel Omp and potentially tethers the T9SS to the cell wall (Gorasia *et al.* 2020); and PorW is associated with the Sov translocon (Gorasia *et al.* 2020). *P. gingivalis* also has several lipoproteins known to be located on the cell surface, including RagB, HmuY and IhtB which are all associated with TDRs and nutrient acquisition. The type V pili are also surface proteins composed of pilin subunits which are made as lipoprotein precursors (see above). An important gap in our knowledge concerns the

Omps required for the secretion of these surface lipoproteins. Lipoproteins yet to be mentioned include PG2054 and PG1028 which are both abundant and contain the peptidoglycan-binding OmpA-C domain suggesting a potential structural role in tethering the OM to the cell wall. PG1341 has a different peptidoglycan-binding domain, "SPOR" and may be involved in cell division. Other lipoproteins are predicted enzymes including PG0082 with a "putative carbohydrate metabolism domain" that is potentially associated with the PG0083 OM β-barrel; PG1713, a putative "Rhodanese" domain-containing enzyme; PG2164, a peptidyl-prolyl cis-trans isomerase; PG1948, a putative hydrolase (e.g peptidase or esterase); and PG2197 that was strongly predicted to be a metalloendopeptidase (**Table S1**). Lastly, PG0183 which is a large protein of 240 kDa has both a strongly predicted lipoprotein signal as well as a C-terminal T9SS signal suggesting that it may be anchored to the OM at both N-terminal and C-terminal ends (Veith *et al.* 2014). Lipoproteins involved in large complexes included PG1093, PG1713, PG1835, PG2164 and PG2173 (**Fig S1**).

Unusual OM-associated proteins

Several proteins could not be classified either functionally or structurally by bioinformatic approaches, including PG0987 and PG1625. Their localisation to the OM was based on the experimental evidence (Veith *et al.* 2014). Interestingly, both of these were found to form high MW complexes suggesting that their association with the OM may be via protein-protein interactions (**Fig S1**). PG0987 was predicted to form an amphipathic helix from residue 179-236. The predicted helix displays a small hydrophobic face and a large hydrophilic face that can be divided into separate faces enriched in either Lys, Arg or Glu (**Fig 5**). PG1625 exhibited a high TMBB score, however inspection of the sequence did not reveal any likely β-barrel strands. Instead, the sequence could be divided into three regions based on sequence bias. The N-terminal region was acidic and enriched with charged residues (Asp, Glu, Arg and Lys). The middle region in particular was notable in comprising 38% aromatic residues, particularly Tyr and also15% Gly (**Fig 5**). The C-terminal region was extremely basic with an estimated isoelectric point of 12.6 and was also enriched with Gly, Asn and Ser residues, with Gly and Ser particularly concentrated within the last 35 amino acids. The native MW profile of PG1625 closely followed that of the FadL-like β-barrel, PG1626 suggesting they may form a complex together (**Fig S1**).

Using BLAST to predict virulence factors

Bacteria, like all organisms conserve those proteins required for common functions such as central metabolism, gene expression and multiplication to name just a few. For processes involving Omps, common functions include OM biogenesis and maintenance, the acquisition of nutrients, the efflux of toxic substances and secretion systems. However, processes that involve the interaction of the

bacterium with its immediate environment can involve more specialised (less conserved) proteins. To use this property to predict Omps that may be involved in such functions, and therefore likely to be involved in specific modes of virulence, each protein was searched with BLAST against the *Bacteroidales* taxon, and the number of hits were recorded (**Table S1**). As proof of principle, it was observed that OM proteins involved in the above-listed conserved processes had high numbers of homologs (**Fig 6A**) while the T9SS cargo proteins which are known to be largely virulence factors located on the cell surface had on average, 10-fold lower numbers of homologs (**Fig 6B**). Excluding the T9SS cargo, the 20 Omps with the lowest number of homologs are shown (**Fig 6B**). These included seven small β-barrels and seven lipoproteins of which three are associated with pili. Some of these are likely to be important virulence factors enabling *P. gingivalis* to thrive in its environment and interact with its host in a species-specific manner.

The OM proteome in perspective

According to the Ecocyc database (Keseler *et al.* 2013), 131 *Escherichia coli* K12 proteins are localised to the "cell OM". This does not include cell surface appendages such as fimbria and flagella but does include proteins with multiple localisations such as TamB and LptA, which in this study were considered IM and periplasm respectively. In round terms, the OM proteomes of *E. coli* and *P. gingivalis* are a similar size. *E. coli* has flagellar, curli and a larger number of fimbrillin/pilin proteins and up to 11 OM usher proteins for the secretion of pilins but does not have any T9SS cargo proteins. *P. gingivalis* has six more TDRs, two more OM efflux proteins but possibly lacks autotransporters of which *E. coli* has four.

The *P. gingivalis* OM proteome presented here is expected to be complete for T9SS cargo proteins due to their ease of prediction. For the β-barrel proteins, approximately half (35/66) were experimentally localised to the OM. Of these 35, only 3 would not have been found by our bioinformatics screen. This suggests that >90% of the β-barrel proteins have been identified by our approach. Of the 65 *P. gingivalis* lipoproteins predicted, 43 could be assigned to the OM so far. The largest unknown is the number of proteins that associate with the OM through protein-protein interactions. In all, there are likely to be < 200 OM-associated proteins which is <10% of the total proteome of ~2,200 proteins. This is a much larger proportion than *E. coli* K12 which has twice the number of proteins (~4,400) encoded in its genome.

Of the 151 proteins presented here, a previous bioinformatics study (Gabarrini *et al.* 2020) predicted that 81 were OM proteins, 24 were periplasmic, 20 were T9SS cargo, 10 were IM proteins, 9 were cytoplasmic, 2 were extracellular with no signal peptide, with the remainder missing due to

database differences (**Table S1**). The differences can largely be accounted for as follows. Predicted cytoplasmic proteins and several IM proteins were misassigned in the aforementioned study due to errors in the W83 sequence database that prevented automatic detection of signal peptides. IM lipoproteins were predicted based on the presence of Gly at the +2 or +3 positions (Gabarrini *et al.* 2020), however this prediction does not appear to be universal as the PorK protein and the BamD protein (PG1215) while both having Gly in the +2 position are very likely to be OM-associated (Gorasia *et al.* 2020; Gorasia *et al.* 2016). Most of the predicted periplasmic proteins had been localised to the OM in our previous study or are known T9SS cargo (Veith *et al.* 2014; Veith *et al.* 2013). The remaining 8 predicted periplasmic proteins included LptE, 3 components of the type IVb secretion system and 4 strongly predicted β-barrel proteins. Looking at it from the other direction, the bioinformatics study (Gabarrini *et al.* 2020) predicted 6 OM proteins and 17 OM lipoproteins that were not identified in our study. Many of the lipoprotein predictions may be correct, depending on the accuracy of the "glycine rule". For the 6 OM proteins, PG0192 and PG0193 are the Skp chaperones involved in Omp assembly (**Fig 1**), PG1326 is a predicted T9SS cargo protein, but it has not been experimentally detected. BLAST searches of PG1326 against other *P. gingivalis* strains suggests that this protein has evolved a non-functional signal peptide in the W50/W83 strains. The remaining three predicted OM proteins, PG1773, PG1790 and PG1978 have potential to be OM proteins.

References

Amano, A. (2010). Bacterial adhesins to host components in periodontitis. *Periodontology 2000*, *52*, 12-37.

Bochtler, M., Mizgalska, D., Veillard, F., Nowak, M. L., Houston, J., Veith, P., Reynolds, E. C., & Potempa, J. (2018). The Bacteroidetes Q-Rule: Pyroglutamate in Signal Peptidase I Substrates. *Frontiers in Microbiology*, *9*, 230.

Bonheyo, G., Graham, D., Shoemaker, N. B., & Salyers, A. A. (2001). Transfer region of a bacteroides conjugative transposon, CTnDOT. *Plasmid*, *45*, 41-51.

Brillet, K., Journet, L., Celia, H., Paulus, L., Stahl, A., Pattus, F., & Cobessi, D. (2007). A beta strand lock exchange for signal transduction in TonB-dependent transducers on the basis of a common structural motif. *Structure*, *15*, 1383-1391.

Chandran, V., Fronzes, R., Duquerroy, S., Cronin, N., Navaza, J., & Waksman, G. (2009). Structure of the outer membrane complex of a type IV secretion system. *Nature*, *462*, 1011-1015.

Chen, Y. Y., Peng, B., Yang, Q., Glew, M. D., Veith, P. D., Cross, K. J., Goldie, K. N., Chen, D., O'Brien-Simpson, N., Dashper, S. G., & Reynolds, E. C. (2011). The outer membrane protein LptO is essential for the O-deacylation of LPS and the co-ordinated secretion and attachment of A-LPS and CTD proteins in *Porphyromonas gingivalis*. *Molecular Microbiology*, *79*, 1380-1401.

Dominy, S. S., Lynch, C., Ermini, F., Benedyk, M., Marczyk, A., Konradi, A., Nguyen, M., Haditsch, U., Raha, D., Griffin, C., Holsinger, L. J., Arastu-Kapur, S., Kaba, S., Lee, A., Ryder, M. I., Potempa, B., Mydel, P., Hellvard, A., Adamowicz, K., Hasturk, H., Walker, G. D., Reynolds, E. C., Faull, R. L. M., Curtis, M. A., Dragunow, M., & Potempa, J. (2019). *Porphyromonas gingivalis* in Alzheimer's disease brains: Evidence for disease causation and treatment with small-molecule inhibitors. *Sci Adv*, *5*, eaau3333.

Dong, C., Beis, K., Nesper, J., Brunkan-Lamontagne, A. L., Clarke, B. R., Whitfield, C., & Naismith, J. H. (2006). Wza the translocon for E. coli capsular polysaccharides defines a new class of membrane protein. *Nature*, *444*, 226-229.

Du, D., Wang-Kan, X., Neuberger, A., van Veen, H. W., Pos, K. M., Piddock, L. J. V., & Luisi, B. F. (2018). Multidrug efflux pumps: structure, function and regulation. *Nature Reviews: Microbiology*, *16*, 523-539.

Figuero, E., Han, Y. W., & Furuichi, Y. (2020). Periodontal diseases and adverse pregnancy outcomes: Mechanisms. *Periodontology 2000*, *83*, 175-188.

Filip, C., Fletcher, G., Wulff, J. L., & Earhart, C. F. (1973). Solubilization of the cytoplasmic membrane of *Escherichia coli* by the ionic detergent sodium-lauryl sarcosinate. *Journal of Bacteriology*, *115*, 717-722.

Freeman, T. C., Jr., Landry, S. J., & Wimley, W. C. (2011). The prediction and characterization of YshA, an unknown outer-membrane protein from *Salmonella typhimurium*. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1808*, 287-297.

Friedrich, V., Gruber, C., Nimeth, I., Pabinger, S., Sekot, G., Posch, G., Altmann, F., Messner, P., Andrukhov, O., & Schaffer, C. (2015). Outer membrane vesicles of *Tannerella forsythia*: biogenesis, composition, and virulence. *Molecular Oral Microbiology*, *30*, 451-473.

Gabarrini, G., Grasso, S., van Winkelhoff, A. J., & van Dijl, J. M. (2020). Gingimaps: Protein localization in the oral pathogen *Porphyromonas gingivalis*. *Microbiology and Molecular Biology Reviews*, *84*.

Glew, M. D., Veith, P. D., Chen, D., Gorasia, D. G., Peng, B., & Reynolds, E. C. (2017). PorV is an outer membrane shuttle protein for the type ix secretion system. *Scientific Reports*, *7*, 8790.

Glew, M. D., Veith, P. D., Chen, D., Seers, C. A., Chen, Y. Y., & Reynolds, E. C. (2014). Blue native-PAGE analysis of membrane protein complexes in *Porphyromonas gingivalis*. *Journal of Proteomics*, *110*, 72-92.

Gorasia, D., Chreifi, G., Seers, C., Butler, C., Heath, J., Glew, M., McBride, M., Subramanian, P., Kjær, A., Jensen, G., Veith, P., & Reynolds, E. (2020). *In situ* structure and organisation of the type IX secretion system. *bioRxiv*, *doi:10.1101/2020.05.13.094771*.

Gorasia, D. G., Veith, P. D., Hanssen, E. G., Glew, M. D., Sato, K., Yukitake, H., Nakayama, K., & Reynolds, E. C. (2016). Structural insights into the PorK and PorN components of the *Porphyromonas gingivalis* type ix secretion system. *PLoS Pathogens*, *12*, e1005820.

Greenwald, J., Nader, M., Celia, H., Gruffaz, C., Geoffroy, V., Meyer, J. M., Schalk, I. J., & Pattus, F. (2009). FpvA bound to non-cognate pyoverdines: molecular basis of siderophore recognition by an iron transporter. *Molecular Microbiology*, *72*, 1246-1259.

Guglielmini, J., Neron, B., Abby, S. S., Garcillan-Barcia, M. P., de la Cruz, F., & Rocha, E. P. (2014). Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Research*, *42*, 5715-5727.

Guo, Y., Nguyen, K. A., & Potempa, J. (2010). Dichotomy of gingipains action as virulence factors: from cleaving substrates with the precision of a surgeon's knife to a meat chopper-like brutal degradation of proteins. *Periodontology 2000*, *54*, 15-44.

Hajishengallis, G., Liang, S., Payne, M. A., Hashim, A., Jotwani, R., Eskan, M. A., McIntosh, M. L., Alsam, A., Kirkwood, K. L., Lambris, J. D., Darveau, R. P., & Curtis, M. A. (2011). Low-abundance biofilm species orchestrates inflammatory periodontal disease through the commensal microbiota and complement. *Cell Host & Microbe*, *10*, 497-506.

Haraszthy, V. I. *& et al*. (2000). Identification of periodontal pathogens in atheromatous plaques. *Journal of Periodontology*, *71*, 1554-1560.

Heath, J. E., Seers, C. A., Veith, P. D., Butler, C. A., Nor Muhammad, N. A., Chen, Y. Y., Slakeski, N., Peng, B., Zhang, L., Dashper, S. G., Cross, K. J., Cleal, S. M., Moore, C., & Reynolds, E. C. (2016). PG1058 is a novel multidomain protein component of the bacterial type ix secretion system. *PloS One*, *11*, e0164313.

Holt, S. C., Ebersole, J., Felton, M., Brunsvold, M., & Korman, K. S. (1988). Implantation of *Bacteroides gingivalis* in non-human primates initiates progression of periodontitis. *Science*, *239*, 55-57.

Inoue, T., Nakayama, M., Taguchi, Y., Kano, K., Ono, M., Shimizu, Y., Kuroda, T., & Ohara, N. (2015). Characterization of the tripartite drug efflux pumps of *Porphyromonas gingivalis* aTCC 33277. *New Microbiologica*, *38*, 101-108.

Josts, I., Veith, K., & Tidow, H. (2019). Ternary structure of the outer membrane transporter FoxA with resolved signalling domain provides insights into TonB-mediated siderophore uptake. *Elife*, *8*.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*, 845-858.

Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martinez, C., Fulcher, C., Huerta, A. M., Kothari, A., Krummenacker, M., Latendresse, M., Muniz-Rascado, L., Ong, Q., Paley, S., Schroder, I., Shearer, A. G., Subhraveti, P., Travers, M., Weerasinghe, D., Weiss, V., Collado-Vides, J., Gunsalus, R. P., Paulsen, I., & Karp, P. D. (2013). EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Research*, *41*, D605-612.

Kirst, M. E., Li, E. C., Alfant, B., Chi, Y. Y., Walker, C., Magnusson, I., & Wang, G. P. (2015). Dysbiosis and alterations in predicted functions of the subgingival microbiome in chronic periodontitis. *Applied and Environmental Microbiology*, *81*, 783-793.

Konovalova, A., Kahne, D. E., & Silhavy, T. J. (2017). Outer membrane biogenesis. *Annual Review of Microbiology*, *71*, 539-556.

Konovalova, A. & Silhavy, T. J. (2015). Outer membrane lipoprotein biogenesis: Lol is not the end. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *370*.

Kulkarni, S. S., Zhu, Y., Brendel, C. J., & McBride, M. J. (2017). Diverse c-terminal sequences involved in *Flavobacterium johnsoniae* protein secretion. *Journal of Bacteriology*, *199*.

Lasica, A. M., 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.

Lauber, F., Deme, J. C., Lea, S. M., & Berks, B. C. (2018). Type 9 secretion system structures reveal a new protein transport mechanism. *Nature*, *564*, 77-82.

Linden, G. J., Lyons, A., & Scannapieco, F. A. (2013). Periodontal systemic associations: review of the evidence. *Journal of Periodontology*, *84*, S8-S19.

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research*, *48*, D265-D268.

Madej, M., White, J. B. R., Nowakowska, Z., Rawson, S., Scavenius, C., Enghild, J. J., Bereta, G. P., Pothula, K., Kleinekathoefer, U., Basle, A., Ranson, N. A., Potempa, J., & van den Berg,

B. (2020). Structural and functional insights into oligopeptide acquisition by the RagAB transporter from Porphyromonas gingivalis. *Nature Microbiology*.

McClean, S. (2012). Eight stranded beta -barrel and related outer membrane proteins: role in bacterial pathogenesis. *Protein & Peptide Letters*, *19*, 1013-1025.

Nagano, K., Hasegawa, Y., Abiko, Y., Yoshida, Y., Murakami, Y., & Yoshimura, F. (2012). *Porphyromonas gingivalis* FimA fimbriae: fimbrial assembly by fimA alone in the fim gene cluster and differential antigenicity among fimA genotypes. *PloS One*, *7*, e43722.

Nagano, K., Hasegawa, Y., Yoshida, Y., & Yoshimura, F. (2017). Novel fimbrilin PGN_1808 in *Porphyromonas gingivalis*. *PloS One*, *12*, e0173541.

Nagano, K., Murakami, Y., Nishikawa, K., Sakakibara, J., Shimozato, K., & Yoshimura, F. (2007). Characterization of RagA and RagB in Porphyromonas gingivalis: study using gene-deletion mutants. *Journal of Medical Microbiology*, *56*, 1536-1548.

Naito, M., Tominaga, T., Shoji, M., & Nakayama, K. (2019). PGN_0297 is an essential component of the type IX secretion system (T9SS) in *Porphyromonas gingivalis*: Tn-seq analysis for exhaustive identification of T9SS-related genes. *Microbiology and Immunology*, *63*, 11-20.

Nguyen, K. A., 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*, *155*, 328-337.

O'Brien-Simpson, N. M., Veith, P. D., Dashper, S. G., & Reynolds, E. C. (2003). *Porphyromonas gingivalis* gingipains: the molecular teeth of a microbial vampire. *Current Protein & Peptide Science*, *4*, 409-426.

Okuda, S., Sherman, D. J., Silhavy, T. J., Ruiz, N., & Kahne, D. (2016). Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model. *Nature Reviews: Microbiology*, *14*, 337-345.

Orlandi, M., Graziani, F., & D'Aiuto, F. (2020). Periodontal therapy and cardiovascular risk. *Periodontology 2000*, *83*, 107-124.

Pautsch, A. & Schulz, G. E. (1998). Structure of the outer membrane protein A transmembrane domain. *Nature Structural Biology*, *5*, 1013-1017.

Rassam, P., Copeland, N. A., Birkholz, O., Toth, C., Chavent, M., Duncan, A. L., Cross, S. J., Housden, N. G., Kaminska, R., Seger, U., Quinn, D. M., Garrod, T. J., Sansom, M. S., Piehler, J., Baumann, C. G., & Kleanthous, C. (2015). Supramolecular assemblies underpin turnover of outer membrane proteins in bacteria. *Nature*, *523*, 333-336.

Sande, C., Bouwman, C., Kell, E., Nickerson, N. N., Kapadia, S. B., & Whitfield, C. (2019). Structural and functional variation in outer membrane polysaccharide export (OPX) proteins

from the two major capsule assembly pathways present in *Escherichia coli*. *Journal of Bacteriology*, *201*.

Sanganna Gari, R. R., Seelheim, P., Marsh, B., Kiessling, V., Creutz, C. E., & Tamm, L. K. (2018). Quaternary structure of the small amino acid transporter OprG from *Pseudomonas aeruginosa*. *Journal of Biological Chemistry*, *293*, 17267-17277.

Sato, K., Yukitake, H., Narita, Y., Shoji, M., Naito, M., & Nakayama, K. (2013). Identification of *Porphyromonas gingivalis* proteins secreted by the Por secretion system. *FEMS Microbiology Letters*, *338*, 68-76.

Schnaitman, C. A. (1970). Protein composition of the cell wall and cytoplasmic membrane of *Escherichia coli*. *Journal of Bacteriology*, *104*, 890-901.

Sgro, G. G., Costa, T. R. D., Cenens, W., Souza, D. P., Cassago, A., Coutinho de Oliveira, L., Salinas, R. K., Portugal, R. V., Farah, C. S., & Waksman, G. (2018). Cryo-EM structure of the bacteria-killing type IV secretion system core complex from *Xanthomonas citri*. *Nature Microbiology*, *3*, 1429-1440.

Shibata, S., Shoji, M., Okada, K., Matsunami, H., Matthews, M. M., Imada, K., Nakayama, K., & Wolf, M. (2020). Structure of polymerized type V pilin reveals assembly mechanism involving protease-mediated strand exchange. *Nature Microbiology*, *5*, 830-837.

Shoji, M., Shibata, S., Sueyoshi, T., Naito, M., & Nakayama, K. (2020). Biogenesis of type V pili. *Microbiology and Immunology*.

Shrivastava, R. & Chng, S. S. (2019). Lipid trafficking across the Gram-negative cell envelope. *Journal of Biological Chemistry*, *294*, 14175-14184.

Smalley, J. W. & Olczak, T. (2017). Heme acquisition mechanisms of *Porphyromonas gingivalis* - strategies used in a polymicrobial community in a heme-limited host environment. *Molecular Oral Microbiology*, *32*, 1-23.

Socransky, S. S. & *et al.* (1998). Microbial complexes in subgingival plaque. *Journal of Clinical Periodontology*, *25*, 134-144.

Stubenrauch, C. J. & Lithgow, T. (2019). The TAM: A translocation and assembly module of the beta-barrel assembly machinery in bacterial outer membranes. *EcoSal Plus*, *8*.

van den Berg, B. (2005). The FadL family: unusual transporters for unusual substrates. *Current Opinion in Structural Biology*, *15*, 401-407.

van den Berg, B., Black, P. N., Clemons, W. M., Jr., & Rapoport, T. A. (2004). Crystal structure of the long-chain fatty acid transporter FadL. *Science*, *304*, 1506-1509.

Veith, P. D., Chen, Y. Y., Chen, D., O'Brien-Simpson, N. M., Cecil, J. D., Holden, J. A., Lenzo, J. C., & Reynolds, E. C. (2015). *Tannerella forsythia* outer membrane vesicles are enriched

with substrates of the type ix secretion system and tonb-dependent receptors. *Journal of Proteome Research*, *14*, 5355-5366.

Veith, P. D., Chen, Y. Y., Gorasia, D. G., Chen, D., Glew, M. D., O'Brien-Simpson, N. M., Cecil, J. D., Holden, J. A., & Reynolds, E. C. (2014). *Porphyromonas gingivalis* outer membrane vesicles exclusively contain outer membrane and periplasmic proteins and carry a cargo enriched with virulence factors. *Journal of Proteome Research*, *13*, 2420-2432.

Veith, P. D., Glew, M. D., Gorasia, D. G., & Reynolds, E. C. (2017). Type IX secretion: the generation of bacterial cell surface coatings involved in virulence, gliding motility and the degradation of complex biopolymers. *Molecular Microbiology*, *106*, 35-53.

Veith, P. D., Luong, C., Tan, K. H., Dashper, S. G., & Reynolds, E. C. (2018). Outer membrane vesicle proteome of *Porphyromonas gingivalis* is differentially modulated relative to the outer membrane in response to heme availability. *Journal of Proteome Research*, *17*, 2377-2389.

Veith, P. D., Nor Muhammad, N. A., Dashper, S. G., Likic, V. A., Gorasia, D. G., Chen, D., Byrne, S. J., Catmull, D. V., & Reynolds, E. C. (2013). Protein substrates of a novel secretion system are numerous in the Bacteroidetes phylum and have in common a cleavable C-terminal secretion signal, extensive post-translational modification, and cell-surface attachment. *Journal of Proteome Research*, *12*, 4449-4461.

Veith, P. D., Talbo, G. H., Slakeski, N., Dashper, S. G., Moore, C., Paolini, R. A., & Reynolds, E. C. (2002). Major outer membrane proteins and proteolytic processing of RgpA and Kgp of *Porphyromonas gingivalis* W50. *Biochemical Journal*, *363*, 105-115.

Veith, P. D., Talbo, G. H., Slakeski, N., & Reynolds, E. C. (2001). Identification of a novel heterodimeric outer membrane protein of Porphyromonas gingivalis by two-dimensional gel electrophoresis and peptide mass fingerprinting. *European Journal of Biochemistry*, *268*, 4748-4757.

Ximenez-Fyvie, L. A., Haffajee, A. D., & Socransky, S. S. (2000). Comparison of the microbiota of supra- and subgingival plaque in health and periodontitis. *Journal of Clinical Periodontology*, *27*, 648-657.

Xu, Q., Shoji, M., Shibata, S., Naito, M., Sato, K., Elsliger, M. A., Grant, J. C., Axelrod, H. L., Chiu, H. J., Farr, C. L., Jaroszewski, L., Knuth, M. W., Deacon, A. M., Godzik, A., Lesley, S. A., Curtis, M. A., Nakayama, K., & Wilson, I. A. (2016). A distinct type of pilus from the human microbiome. *Cell*, *165*, 690-703.

Ye, J. & van den Berg, B. (2004). Crystal structure of the bacterial nucleoside transporter Tsx. *EMBO Journal*, *23*, 3187-3195.

Zahn, M., D'Agostino, T., Eren, E., Basle, A., Ceccarelli, M., & van den Berg, B. (2015). Small-molecule transport by caro, an abundant eight-stranded beta-barrel outer membrane protein from *Acinetobacter baumannii*. *Journal of Molecular Biology*, *427*, 2329-2339.

Figure Legends

Figure 1. Proteins predicted to be involved in *P. gingivalis* OM biogenesis. β-barrel Omps are transported across the periplasm with the help of chaperones such as Skp or TamB and inserted into the OM with the aid of BamA and its associated lipoproteins or TamA. LPS is transported to the outer leaflet of the OM by the Lpt system, which was partly predicted previously (Chen *et al.* 2011). Lipoproteins are transported by the Lol system. *P. gingivalis* has multiple copies of proteins related to LolD and LolC/E, however only PG0922, PG1252 and PG1692 were not part of efflux systems. Proteins forming barrels in the OM are shown in blue, OM lipoproteins besides TamA and Wza are shown in pink, periplasmic proteins in orange, IM proteins in green and cytoplasmic proteins in yellow.

Figure 2. Prediction and analysis of TDRs. A. The two different domain organisations observed for TDRs in *P. gingivalis* (see Table 1). PG2008 and PG1552 (HmuR) are shown as examples for the two different architectures. The domains shown are CarbopepD_reg_2 (Cpdr, pfam13715), Plug (pfam07715) and Ligand_gated_channel (cd01347). The location of the TonB box is indicated by the green triangle. B. Multiple alignment of the TonB box sequences. C. Multiple alignment of a subset of TDRs containing a conserved region in the plug domain. D. Structure of RagAB complex in open state. The PDB accession is 6SML (Madej 2020). RagB is the extracellular lid shown in green, the RagA barrel is shown in purple and the plug is shown in yellow. The TonB box is the most N-terminal part that is visible and is shown in red. The Cpdr domain is not visible in the solved structure.

Figure 3. Structural models of the OM β-barrel components of the T9SS. For each model, the name of the protein, PDB accession code and modelled residues are given. Residue numbering is from the start Met residue as provided in Table S1, signal peptide sequence. PorG and PorT were modelled against the OmpA β-barrel domain (*E. coli*), PorF was modelled against a ferripyoverdine receptor (*P. aeruginosa*), and the remaining proteins were all modelled against the SprA-PorV complex structure (*F. johnsoniae*). All models were created using Phyre-2, and graphics were

produced using Chimera. Residues internal to the barrels and the periplasmic carboxypeptidase regulatory domain of PorF were coloured red.

Figure 4. Predicted location of T4SS components in *P. gingivalis*. TraE, TraI and TraO are predicted to form the conjugative pilus. TraM spans both membranes. TraN and the TraQ lipoprotein are predicted to be part of the OM-associated complex and the remaining proteins are predicted to be associated with the IM.

Figure 5. Sequence analyses of PG0987 and PG1625. A. Predicted amphipathic helix of PG0987(179-236) using a helical wheel. B. Sequence of PG1625 showing the sequence bias present in the three regions. Acidic residues (D, E) are shown in blue, basic residues (R, K) in pink, aromatic residues (F, W, Y) in green and other residues of interest (G, S, N) in orange.

Figure 6. The prediction of putative virulence factors using BLAST. The number of BLAST matches with greater than 25% identity and 85% coverage within the *Bacteroidales* taxon is plotted for each group of proteins. A. Protein groups associated with common functions as shown. B. T9SS cargo and after excluding the cargo, the 20 proteins with the fewest number of BLAST hits. Note the very different scales in A and B.

Table 1. Presence of conserved domains in the predicted TDRs

Locus	Cpdr	Plug	LGC/TDR
PG0185	y	y	y
PG0441	y	Mod	y
PG0534	y	y	y
PG0668	y	y	y
PG0707	y	y	y
PG1006	y	y	y
PG1020	y	y	Y
PG1414	y	y	y
PG1493	y	y	Mod
PG1715	y	Mod	y
PG2008	y	y	y
PG2226	y	y	y
PG0TLR	absent	y	y
PG1552	absent	y	y

PG1899	absent	y	y

Absent: no domain present

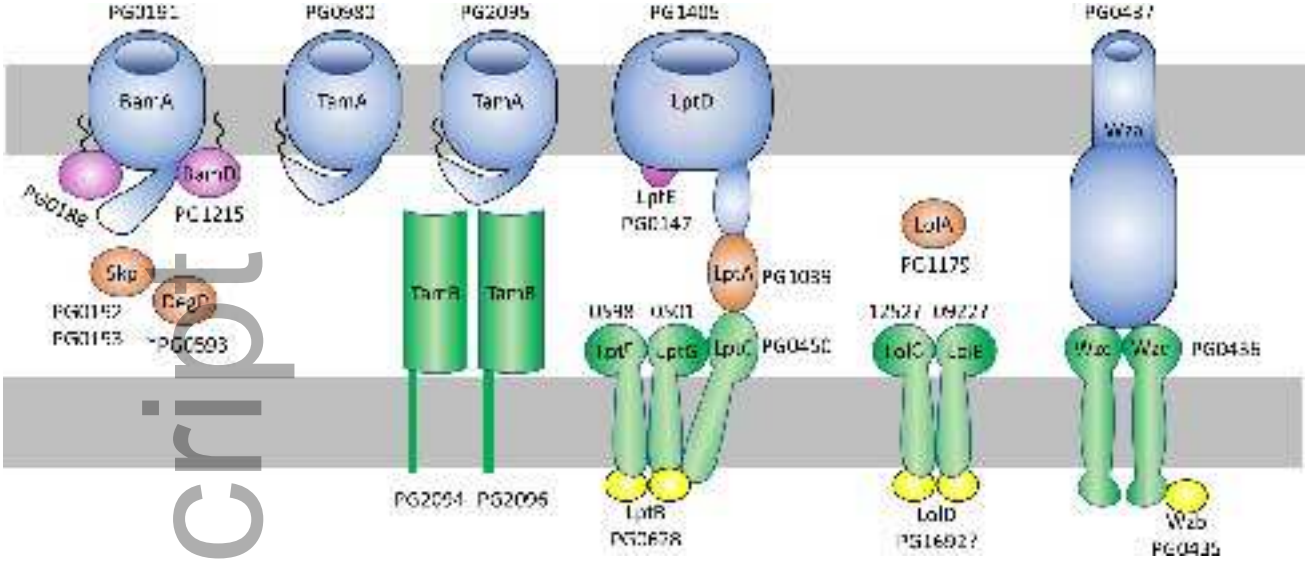
Mod: not detected by conserved domains but modelled by Phyre-2

Cpdr. “CarbopepD_reg_2” (pfam13715)

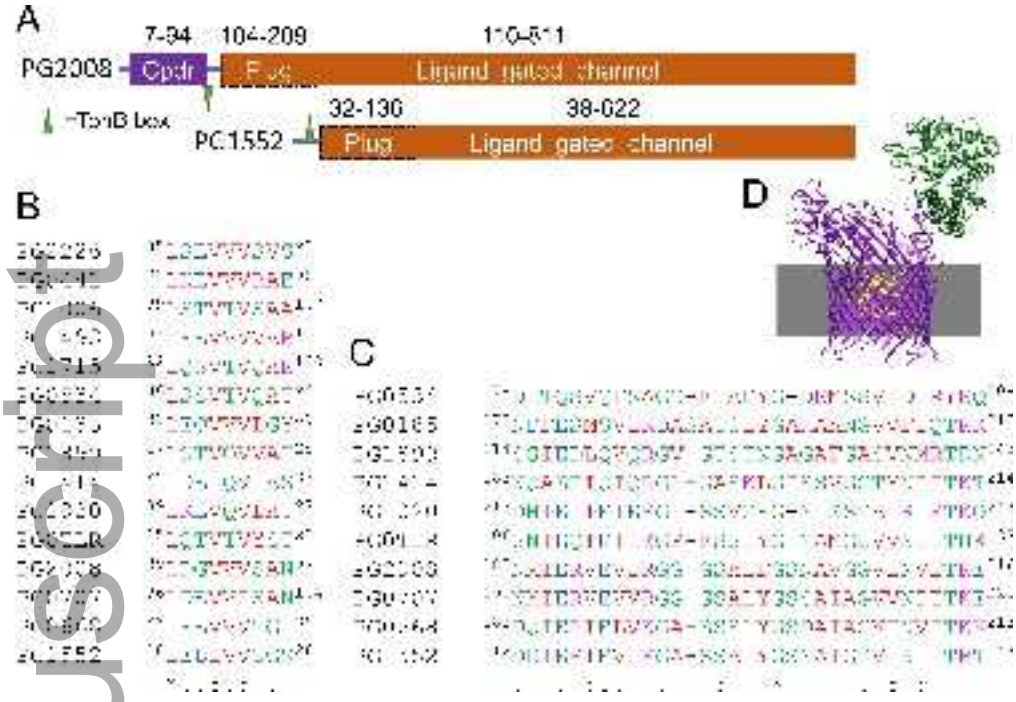
Plug. Pfam07715

LGC. “ligand_gated_channel” (cd01347)

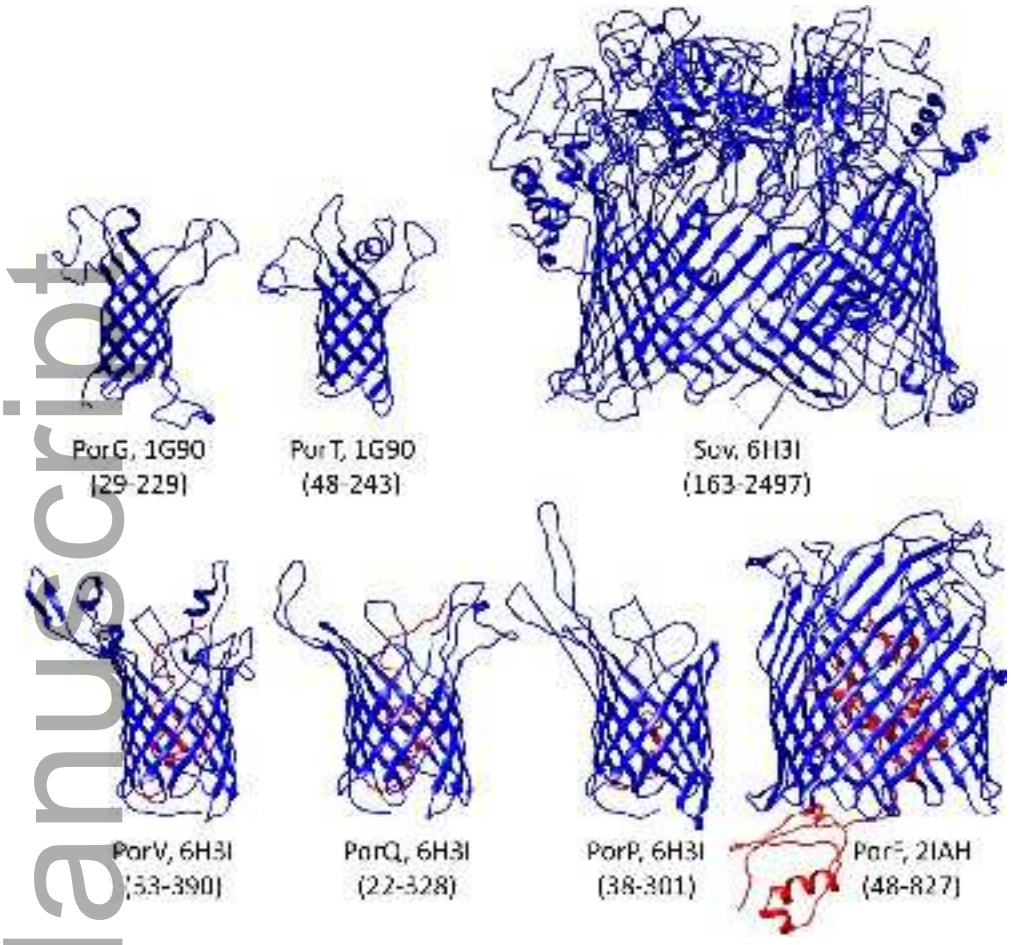
TDR. “TonB_dep_Rec” (pfam00593)



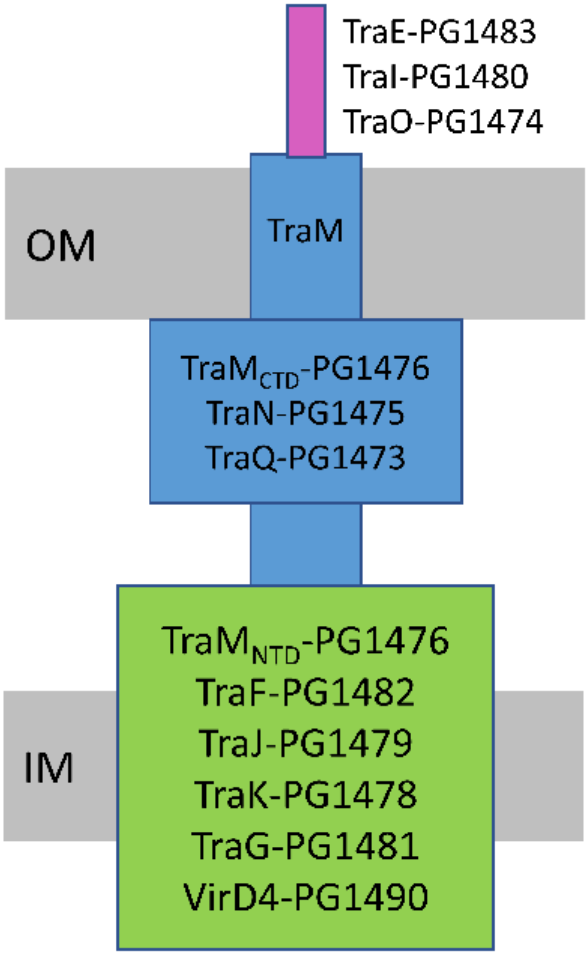
omi_12320_f1.tif



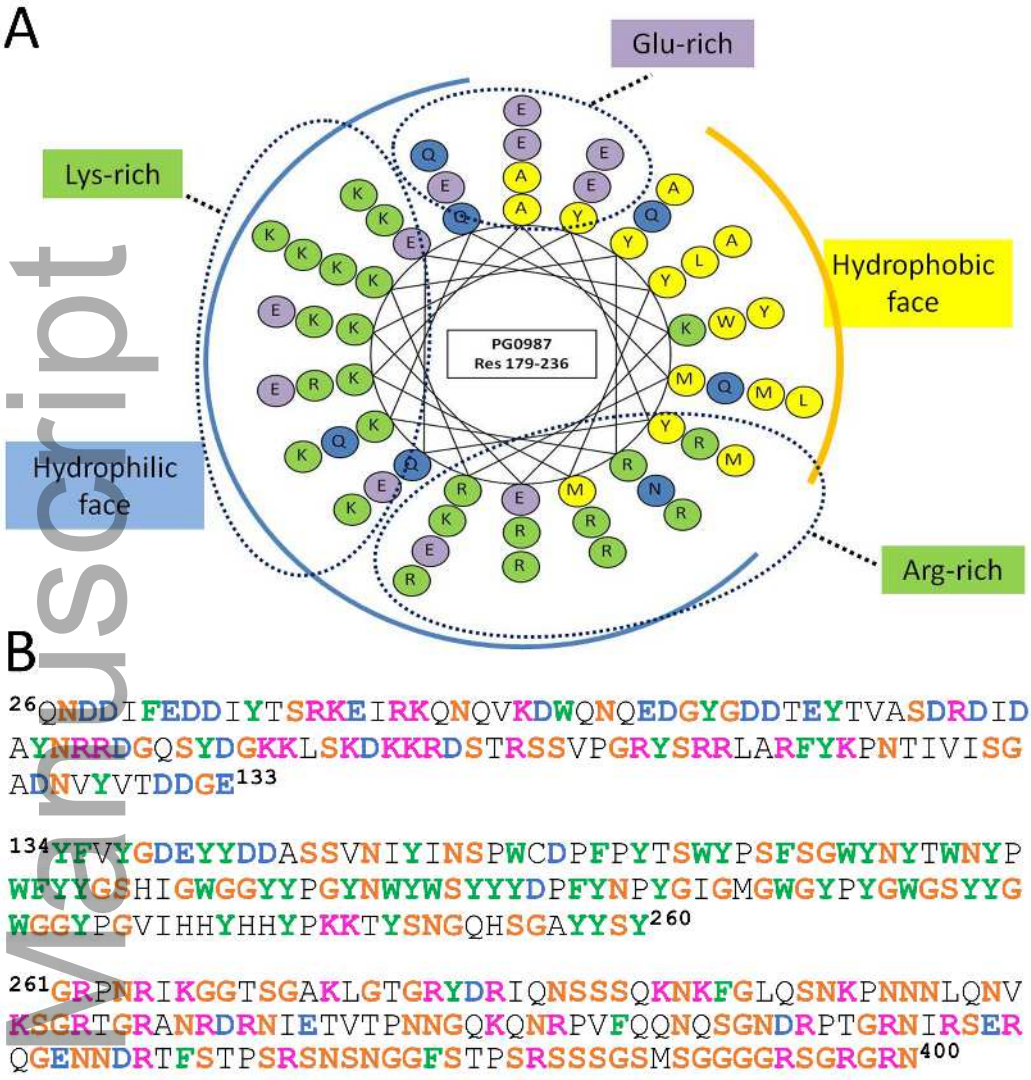
omi_12320_f2.tif

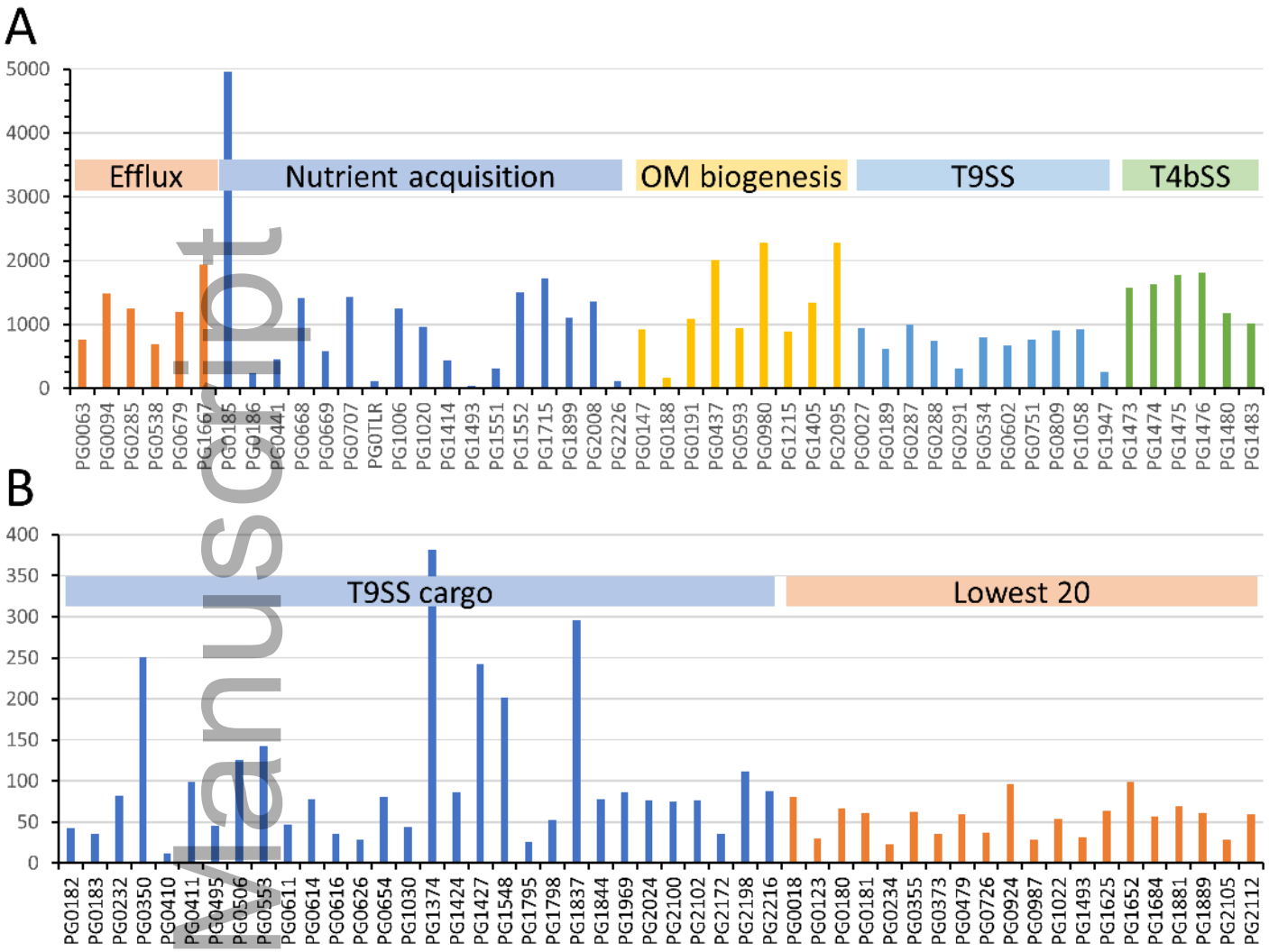


omi_12320_f3.tif



omi_12320_f4.tif





omi_12320_f6.tif