

SUPPLEMENTARY MATERIAL

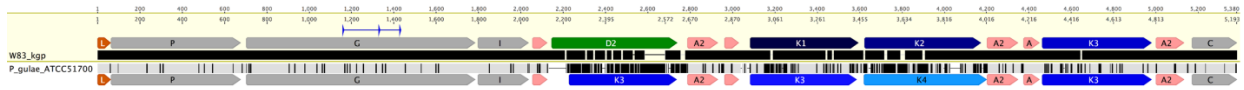


Figure S1. Structure of the *P. gulae* ATCC 51700 kgp gene compared with the *P. gingivalis* W83 kgp gene. Kgp is comprised of a leader peptide (L) that is cleaved during translocation across the inner membrane, a long propeptide (P) of ~200 amino acids and a catalytic domain (G) of around 360 amino acids that is followed by an immunoglobulin fold (I). There are three conserved catalytic residues in the catalytic domain, denoted by the blue lines above the W83 catalytic domain. There are four modular cleaved adhesin domains (CADs) comprised of CADs (K1-K4) or a DUF2436 domain (D2), these are separated by small adhesin binding domains (ABMs). There is a conserved C-terminal domain (C) of approximately 80 amino acids. The black lines denote differences between the predicted amino acid sequence of the *P. gulae* ATCC 51700 kgp gene and the *P. gingivalis* W83 kgp gene.

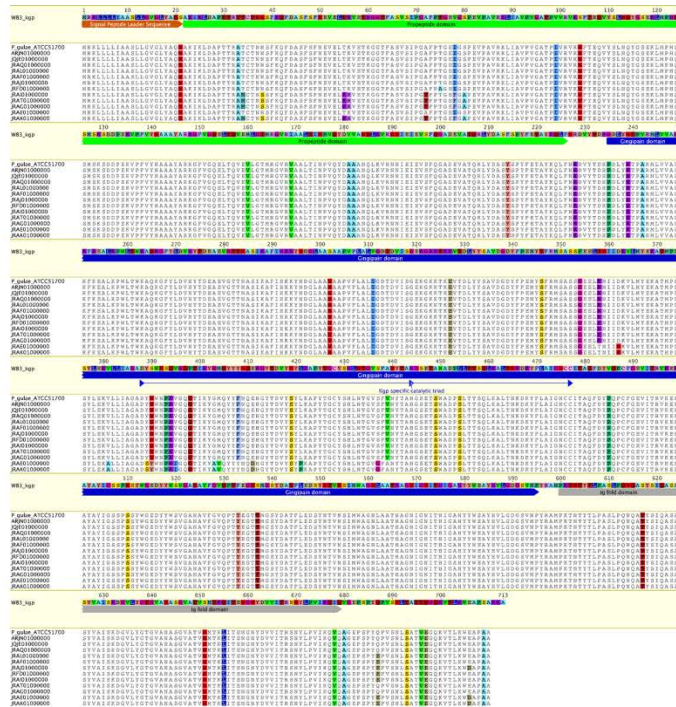


Figure S2. Amino acid sequence comparison of the portion of the *kgp* genes encoding the gingipain leader peptide and catalytic domain derived from the recently published draft genomes of 12 *P. gulae* strains (1), *P. gulae* ATCC 51700 and *P. gingivalis* W83. Differences between the *P. gingivalis* W83 gingipain catalytic domain and those of the *P. gulae* strains are highlighted. The three conserved catalytic residues are shown by the blue line below the *P. gingivalis* W83 sequence. Sequence alignments were performed using the commercial software package Geneious v 8.1.7 (Biomatters Ltd.). Individual genes were codon aligned using the MAFFT plugin v 7.017 from within Geneious (2). The E-INS-I algorithm was used with a BLOSUM62 scoring matrix with a gap opening penalty of 1.03. Protein alignments were extracted from the DNA alignments.

References

1. **Coil DA, Alexiev A, Wallis C, O'Flynn C, Deusch O, Davis I, Horsfall A, Kirkwood N, Jospin G, Eisen JA, Harris S, Darling AE.** 2015. Draft genome sequences of 26 *Porphyromonas* strains isolated from the canine oral microbiome. *Genome Announc* **3**:e00187-00115.
2. **Katoh K, Standley DM.** 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**:772-780.

TABLE S1 Primers used for Sanger sequencing of *P. gulae* ATCC51700 *kgp*.

| Primer Name | Sequence |
|--------------------|-----------------------|
| kgp_F1 | ATTATTATTGCTGATCGCGGC |
| kgp_F2 | TYATGCCRCATCAACCCTCT |
| kgp_F3 | GGAACRACMAACGCCTCT |
| kgp_F4 | CRGCGCATGGATCTGAGAC |
| PgulK5f | CGCGCCGAATTGCTTAATGA |
| PgulK5f2 | GTGTACTCACAGGGTGGAGC |
| PgulK5r1 | GCTCCACCCTGTGAGTACAC |
| PgulK5r2 | TCAAAGTCAGATGCTGCCGT |
| PgulK5r3 | TCATTAAGCAATTCGGCGCG |