Bioinformatic investigation of the cost management strategies of five oral microbes

Keith J. Cross, N. Laila Huq, Seung Hyun Park, Jun Sung Park, Pauline Pham, Minn Quah,
Mathuri Ranjan, Eric C. Reynolds*

Oral Health CRC, Melbourne Dental School, Bio21 Institute, The University of Melbourne,
720 Swanston Street, Melbourne, Victoria, 3010, Australia.

Running title: Energy-saving strategies of oral microbes

Keywords: Protein synthetic cost, management strategy, extracellular unrecyclable proteins, oral
microbes

*Corresponding author: Professor Eric C Reynolds, Melbourne Dental School, The
University of Melbourne, 720 Swanston Street, Victoria 3010, Australia. Tel +61 3 9341
1547, Email: e.reynolds@unimelb.edu.au

SUMMARY

Some amino acids are more energetically costly to synthesise *de novo*, therefore many microbes have evolved to regulate the metabolic expenditure of the cell and reduce the energy burden of extracellular unrecyclable proteins. Several oral bacterial species take up amino acids and peptides obtained from proteolysis of host proteins and hence do not rely only on *de novo* synthesis. The aim of this study was to investigate if five oral bacterial species implement cost management strategies to reduce the energy burden of extracellular unrecyclable proteins. Since the relative *de novo* amino acid synthetic costs are proportional to the masses of the amino acids, the energy costs of producing proteins were assessed by calculating the mean amino acid mass for each protein. For *Porphyromonas gingivalis*, *Treponema denticola*, *Tannerella forsythia*, *Prevotella intermedia*, and *Streptococcus sanguinis*, the outer membrane/extracellular proteins are made up of a much larger percentage of lower average mass amino acids whereas cytoplasmic proteins are made up of a larger proportion of higher average mass amino acid residues. These results are consistent with the five oral bacterial species employing energy-saving mechanisms in the production of extracellular unrecyclable proteins. Interestingly, the *P. gingivalis* and *S. sanguinis* genomes exhibited significantly lower predicted mean amino acid masses compared with that of the genomes of the other three species suggesting this may provide them with an energy advantage with respect to protein biosynthetic cost.

INTRODUCTION

Periodontitis is a chronic inflammatory disease of the supporting tissues of the teeth associated with subgingival plaque (Covani *et al.* 2008; Lamont and Jenkinson 1998; Socranksy *et al.* 1998). The chronic inflammation destroys the tooth's supporting tissues such that eventually the tooth becomes mobile and is lost. Chronic periodontitis is associated with the emergence of three bacterial species in subgingival plaque: *Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia* (Table 1). These bacteria are proteolytic, anaerobic, and together form a closely related aetiological cluster described as the 'Red complex' (Rocas *et al.* 2001; Socranksy *et al.* 1998). These organisms have also been identified as having the highest association with the severity of periodontal disease as measured by probing depth and bleeding on probing (Socranksy *et al.* 1998). Of the 'Red complex' group, *Porphyromonas gingivalis* is considered to be a keystone pathobiont associated with dysbiosis and the initiation and progression of chronic periodontitis (Hajishengallis *et al.* 2011; Lamont and Jenkinson 1998; Socranksy *et al.* 1998). *Prevotella intermedia*, part of the "Orange" complex (which is thought to have close associations with the "Red" complex), is related to various oral infections including periodontitis (Yu *et al.* 2007). *Streptococcus sanguinis*, although considered a commensal species plays an integral role in the colonization of the oral cavity by facilitating the attachment of other oral microorganisms; hence, contributing to the formation of dental plaque (Xu *et al.* 2007). The bacterium has also been associated with infective endocarditis as an opportunistic pathogen (Fan *et al.* 2012) (Table 1). It has been predicted that *S. sanguinis* possesses *de novo* synthesis pathways for all amino acids, however it has been demonstrated *in vitro* that the bacterium will obtain all required amino acids from the growth medium if they are available (Xu *et al.* 2011).

P. gingivalis requires the iron and protoporphyrin IX source, hemin, and proteins as

a source of amino acids, in order to proliferate. Proteins on the outer membrane, called adhesins, attach to host cells and the proteases on the outer membrane cleave host proteins. In addition, the outer membrane proteases called gingipains are packaged into outer membrane vesicles that are released into the host environment. *P. gingivalis* does not utilize glucose to generate energy; rather, it depends on the uptake of peptides and amino acids obtained through proteolytic cleavage of host proteins for energy (Lamont and Jenkinson 1998).

Organisms thrive by way of effective and economical management of their proteins (Smith and Chapman 2010). Proteins can be classified into various groups based on location: cellular, extracellular, periplasmic, and outer membrane proteins. Intracellular proteins are generally recycled; and extracellular proteins that are destined to be exported, are usually lost to the environment. Thus microbes have employed cost management strategies to reduce the economic burden of these lost unrecyclable proteins (Smith and Chapman 2010).

It was hypothesized that *P. gingivalis* and other oral bacteria implement cost-effective protein management strategies in response to economic pressures. Several studies have investigated cost management strategies in bacteria, namely the intensively studied *E.coli* (Akashi and Gojobori 2002; Smith and Chapman 2010). However, there is currently no research available on the cost-saving strategies of these clinically relevant oral bacteria.

In this study, the five oral bacteria described in Table 1 were selected for analysis. The genomes have been sequenced although not fully annotated, Since the relative *de novo* amino acid synthetic costs are proportional to the masses of the amino acids (Smith and Chapman 2010), the costs of producing the proteins was assessed as the mean amino acid mass of each protein calculated as the mass divided by the number of amino acids for each protein. The mean amino acid mass distribution and amino acid compositional biases of

1
2
3 proteins grouped according to predicted and known cellular locations of the proteins were
4 compared. Protein co-factors, such as NADH, FAD, Fe-S and iron centres were not
5 examined in this study since they have a fixed biosynthetic costs and hence are not
6
7
8
9
10 amenable to cost savings strategies in the manner of amino acid preference.

11 12 13 14 15 16 METHODS

17
18 Five oral bacterial species were selected for the study (Table 1): *Porphyromonas gingivalis*,
19
20 *Tannerella forsythia*, and *Treponema denticola* (members of the Red complex), *Prevotella*
21
22 *intermedia*, and *Streptococcus sanguinis*. The protein data of the bacteria were extracted from
23
24 publicly available online databases at Oralgen (www.oralgen.lanl.gov). The protein data were
25
26 cross-referenced with UniProt and GI accession numbers. The molecular weight (MW) of the
27
28 protein sequences was calculated using the Greengene website (greengene.uml.edu). The
29
30 amino acid compositions of the proteins were extracted from an online database (COPid)
31
32 (Kumar *et al.* 2008). The average amino acid compositions for the proteins were sorted
33
34 according to their physicochemical properties (i.e. charged, aliphatic, aromatic, polar, neutral,
35
36 hydrophobic, positively charged, negatively charged, tiny, small and large). The cellular
37
38 locations of the proteins (extracellular, outer-membrane, periplasmic, cytoplasmic,
39
40 cytoplasmic membrane, and unknown location) were based on the published literature or
41
42 were predicted using PSORTb (Yu *et al.* 2010). Hydrophobicity distribution of amino acids
43
44 was evaluated using the hydrophobicity scale (Gaur 2009). Experimentally derived data for
45
46 outer-membrane and cell wall proteins of all 5 genomes were then obtained through relevant
47
48 literature (Ang *et al.* 2008; Veith *et al.* 2009a; Veith *et al.* 2009b; Veith *et al.* 2002; Yang *et al.*
49
50 2012 ; Yu *et al.* 2007) and compared to the predicted data. The Wilcoxon-Mann-Whitney U-
51
52
53
54
55
56
57
58
59
60 test was used for the statistical analysis of the mean amino acid mass data for the five total

genomes and for the analysis of the mean amino acid mass data for proteins distributed across the different cellular locations.

RESULTS

Comparison of the physicochemical properties of oral bacterial proteins

Figure 1 shows the distribution of mean amino acid mass for all five genomes. The mean mass ranged from 113-114 (s.d.=3.4-3.8) for the five genomes. A statistical analysis of the mean amino acid mass for the genomes of *P. gingivalis* (112.776) and *S. sanguinis* (112.557) indicated they were significantly different ($p<0.0001$) to those for *T. denticola* (113.385), *T. forsythia* (113.829) and *P. intermedia* (113.624) (Figure 1). The relative amino acid abundances were compared as a function of hydrophobicity (Figure 2). All 5 genomes have Leu as the most abundant amino acid followed by Ala, Ile, Gly, Lys, and Val. The three least abundant amino acids are Trp, His, and Cys.

All five genomes displayed similar physicochemical properties (Figure 3). There were considerably fewer aromatic than aliphatic amino acids (Figure 3). Additionally, the majority of amino acids were “small” (Ala, Cys, Asp, Gly, Ser, Thr) followed by “tiny” (His, Ile, Leu, Lys, Met, Asn, Pro, Gln, Val) and then “large” (Phe, Tyr, Arg, Trp) as classified by Kumar *et al.*, (2008).

Distribution of the predicted and experimentally demonstrated cellular locations of the proteins from the 5 genomes

The predicted locations for all proteins in the 5 genomes were obtained from PSORTb (Yu *et al.* 2010). The majority of proteins were predicted to reside in the cytoplasm (56% to 68%) compared to the outer-membrane and extracellular environment. Out of 1911 proteins for *P.*

gingivalis, only 0.2% were predicted to be extracellular proteins. The extracellular proteins were predicted to be 0.7%, 0.7%, 0.6%, and 1% of the total annotated proteins for *T. denticola*, *P. intermedia*, *T. forsythia*, and *S. sanguinis*, respectively (Figure 4).

Table 2 shows the calculated mean amino acid mass of proteins and grouped according to their predicted location: extracellular, outer-membrane, periplasmic, cytoplasmic, cytoplasmic membrane, and unknown. Experimentally determined outer-membrane and or cell surface located proteins for all five genomes were obtained from the relevant literature (Ang *et al.* 2008; Veith *et al.* 2009a; Veith *et al.* 2002; Yang *et al.* 2012 ; Yu *et al.* 2007). A statistical analysis of the mean amino acid mass of the proteins distributed across the experimentally determined and/or predicted cellular locations demonstrated that the mean amino acid mass of the extracellular proteins was significantly smaller than the total mean amino acid mass for these oral bacterial species (Table 2). Specifically the mean amino acid mass within the subset of experimentally determined outer-membrane/cell surface proteins was significantly lower than the mean amino acid mass for the entire proteome of these organisms: *P. gingivalis* ($p=0.00017$), *T. forsythia* ($p<0.0001$), *T. denticola* ($p=0.024$) and *P. intermedia* ($p=0.002$), and was close to significance for *S. sanguinis* ($p=0.07$). The known outer-membrane located proteins are predicted to be both acidic and basic (Table 3).

DISCUSSION

Organisms have evolved and adapted to environmental pressures by tightly regulating their cellular metabolic expenditures. One mechanism by which organisms can do this is by reducing the average synthetic costs of extracellular proteins to reduce the economic cost of proteins irretrievably lost and excreted out of the cell. This has been shown to occur in the Gram-negative anaerobe *E.coli* (Smith and Chapman 2010). Smith and Chapman (2010)

suggested that the average synthetic cost (ASC) of amino acids is strongly influenced by the cellular location of the proteins (Smith and Chapman 2010).

This study investigated the relationship between the biosynthetic cost of amino acids, which are known to scale as the amino acid molecular weight (Smith and Chapman 2010) and predicted and known protein localizations in five oral bacterial species. Despite living in an environment which can provide host peptides as a supply of amino acids these five oral bacterial species still practice “good management” by using energetically “cheaper” amino acids in outer-membrane and extracellular proteins. The results comparing the predicted protein locations of *P.gingivalis* and the other selected oral bacterial species revealed that the majority of proteins reside in the cytoplasm rather than the outer membrane and extracellular environment. The average mean mass of amino acids within the subset of experimentally determined, outer-membrane proteins was shown to be significantly less than the total proteome mean amino acid mass. These results suggest that these oral bacterial species have evolved to employ cost-saving measures irrespective of their reliance on *de novo* synthesis. Interestingly the mean amino acid mass for the genomes of *P. gingivalis* and *S. sanguinis* was significantly smaller than that for the genomes of *T. denticola*, *T. forsythia* and *P. intermedia* suggesting that the two former species may be more economic with respect to the overall cost of protein biosynthesis. This may give the two species an energy advantage with respect to protein biosynthetic cost helping them to compete with other species with similar metabolic/energy characteristics sharing their environment in a polymicrobial biofilm. This speculation is consistent with *S. sanguinis* being a frequently found commensal in health and *P. gingivalis* being a pathogen frequently associated with periodontal disease (Caufield *et al.* 2000; Griffen *et al.* 1998; Stingu *et al.* 2008).

Figure 2 indicated that Trp, His, and Cys were the least abundant amino acids and this trend was reflected in the five genomes that were studied. This can be explained by the

relatively higher average synthetic costs of these amino acids, especially aromatic amino acids like Trp and His (Smith and Chapman 2010). However, Trp has been found to be an integral part in the stabilization of extracellular proteins. Cys (in its reduced form) is highly hydrophobic and shares physicochemical properties similar to that of aliphatic and aromatic residues, like Trp and Tyr (Nagano *et al.* 1999). Cys has the unique ability to form ‘disulphide bridges’ and the chemical reactions involving the sulphur atoms consume large quantities of energy. The occurrence of sulphur in the environment is rare; therefore it is unsurprising that Cys was found to be one of the least abundant amino acid residues (Marino and Gladyshev 2011). However, they are commonly found in extracellular proteins where they play an important role in maintaining protein structure and are involved in protein-protein interactions. Furthermore, the sulphur atom bonds readily to metal ions, which can be important in enzymatic functions, e.g. metal proteases and catalytic sites (Betts and Russell 2003). Hence, amino acid cost minimization possibly works within the limits allowed by functional constraints **that is** cost minimization should have a greater effect on amino acids found to be less crucial to protein structure **or function.**

Amino acids that were found to be most abundant in all **five** genomes include Leu, Ile, Ala, Gly, and Val (Figure 2). These amino acids were also found to be abundant in *E.coli* a Gram-negative, facultative anaerobic bacterium (Akashi and Gojobori 2002). Additionally, **the energetically cheaper** Leu and Ala, non-polar aliphatic amino acids were more abundant and play important roles in the formation of stable protein structures. **Furthermore** Lys, a hydrophilic amino acid was also found to be abundant **as expected** (Lodish *et al.* 2000). Lysine residues are quite frequent in binding sites as they contain a positively charge side chain that is involved in the formation of **ion pairs and** hydrogen bonds and are important for protein structure stability (Betts and Russell 2003).

The comparison of the physicochemical properties of overall amino acids of the five genomes demonstrated that small amino acids (>50%) are relatively more abundant and aromatic amino acids are rare (Figure 3). Other physicochemical properties did not appear to have much distinction. In some instances, the function of proteins overrides the metabolic cost and may help to explain the relatively higher abundance of a particular amino acid. This is exemplified by Gly, which is the smallest amino acid and cheapest to produce, however it is limited in its contribution to hydrophobic interactions and the formation of stable interactions. In contrast to Leu, which has a branched side chain and is energetically more expensive to produce than Gly, however has a functional purpose in the make-up of proteins, specifically in the formation of hydrophobic clusters in the center of proteins (Lehninger *et al.* 2008).

Additionally, extracellular factors, including availability of nutrients and the way in which organisms utilize these nutrients to produce proteins hugely influence the metabolic expenditure of the cell. *E. coli* is capable of producing its own amino acids without growth factors and has developed cost-minimization strategies to thrive in nutrient deficient environments (Smith and Chapman 2010; Todar 2012). Out of the five oral bacterial genomes studied, it was expected that *S. sanguinis*, an organism able to synthesize all the essential amino acids, would develop a similar economical mechanism (Xu *et al.* 2007). However, it was unexpected that *P. gingivalis*, *T. denticola*, *T. forsythia*, and *P. intermedia* would also adopt such cost-saving strategies.

CONCLUSIONS

1
2
3 The results of this study showed that the five oral bacterial species limit the cost of their
4
5 extracellular proteins by using “cheaper” amino acids despite growing in a nutrient-rich
6
7 environment. Additionally, the deduced proteins from the five bacterial genomes displayed
8
9 similar physicochemical properties with fewer costly aromatic amino acids and more
10
11 inexpensive, small-sized amino acids present. Interestingly the genomes of *P. gingivalis* and
12
13 *S. sanguinis* exhibited significantly lower predicted mean amino acid masses compared with
14
15 that of the genomes of *T. denticola*, *P. intermedia* and *T. forsythia* suggesting this may
16
17 provide *P. gingivalis* and *S. sanguinis* with an energy advantage with respect to protein
18
19 biosynthetic cost.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

Akashi, H. and Gojobori, T. (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **99**: 3695–3700.

Ang, C.S., Veith, P.D., Dashper, S.G., and Reynolds, E.C. (2008) Application of 16O/18O reverse proteolytic labeling to determine the effect of biofilm culture on the cell envelope proteome of *Porphyromonas gingivalis* W50. *Proteomics* **8**: 1645-60.

Betts, M.J. and Russell, R.B. (2003) *Bioinformatics for Geneticists* Wiley.

Caufield, P.W., Dasanayake, A.P., Li, Y., Pan, Y., Hsu, J., and Hardin, J.M. (2000) Natural history of *Streptococcus sanguinis* in the oral cavity of infants: evidence for a discrete window of infectivity. *Infect Immun* **68**: 4018-23.

Covani, U., Marconcini, S., Giacomelli, L., Sivozhelevov, V., Barone, A., and Nicolini, C. (2008) Bioinformatic prediction of leader genes in human periodontitis. *J Periodontal Res* **79**: 1974-1983.

Fan, J., Zhang, Y., Chuang-Smith, O.N., Frank, K.L., Guenther, B.D., Kern, M., Schlievert, P.M., and Herzberg, M.C. (2012) Ecto-5'-nucleotidase: a candidate virulence factor in *Streptococcus sanguinis* experimental endocarditis. *PLoS One* **7**: e38059.

Gaur, R. (2009) Prokaryotic and eukaryotic non-membrane proteins have biased amino acid distribution. *J Comput Sci Syst Biol* **2**: 298-299.

Ge, X., Kitten, T., Munro, C.L., Conrad, D.H., and Xu, P. (2010) Pooled protein immunization for identification of cell surface antigens in *Streptococcus sanguinis*. *PLoS One* **5**: e11666.

Griffen, A.L., Becker, M.R., Lyons, S.R., Moeschberger, M.L., and Leys, E.J. (1998) Prevalence of *Porphyromonas gingivalis* and periodontal health status. *J Clin Microbiol* **36**: 3239-42.

- 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., and Curtis, M.A. (2011) Low-abundance biofilm species orchestrates inflammatory periodontal disease through the commensal microbiota and complement. *Cell Host Microbe* **10**: 497-506.
- Kumar, M., Thakur, V., and Raghava, G. (2008) COPid: Composition based protein identification. *In Silico Biol* **8**: 121-8.
- Lamont, R. and Jenkinson, H. (1998) Life below the gum line: Pathogenic mechanisms of *Porphyromonas gingivalis*. *Am Soc Microbiol* **62**: 1244-1263.
- Lehninger, A., Cox, M., and Nelson, D. (2008) *Lehninger Principles of Biochemistry*. 5th ed. edn. New York: W.H. Freeman.
- Lodish, H., Berk, A., and Zipursky, S. (2000) *Mol Cell Biol*. 4th edn. New York.
- Marino, S. and Gladyshev, V. (2011) Redox biology: computational approaches to the investigation of functional cysteine residues *Antioxid Redox Signal* **15** 135-146.
- Nagano, N., Otaa, M., and Nishikawaa, K. (1999) Strong hydrophobic nature of cysteine residues in proteins *FEBS Lett* **458**: 69-71.
- Rocas, I., Siqueira, J., Santos, K., and Coelho, A. (2001) "Red complex" (*Bacteroides forsythus*, *Porphyromonas gingivalis*, and *Treponema denticola*) in endodontic infections: a molecular approach. *Oral Surg Oral Med Oral Path Oral Radiol Endo* **91**: 468-471.
- Smith, D.R. and Chapman, M.R. (2010) Economical evolution: Microbes reduce the synthetic cost of extracellular proteins. *mBio* **1**: e00131-10.
- Socranksy, S.S., Haffajee, A.D., Cugini, M.A., Smith, C., and Jr, R.L.K. (1998) Microbial complexes in subgingival plaque *J Clin Periodontol* **25**: 134-144.
- Stingu, C.S., Eschrich, K., Rodloff, A.C., Schaumann, R., and Jentsch, H. (2008) Periodontitis is associated with a loss of colonization by *Streptococcus sanguinis*. *J*

- 1
2
3 *Med Microbiol* **57**: 495-9.
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
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Todar, K. (2012) *Online Textbook of Bacteriology*. Wisconsin.
- Veith, P.D., Dashper, S.G., O'Brien-Simpson, N.M., Paolini, R.A., Orth, R., Walsh, K.A., and Reynolds, E.C. (2009a) Major proteins and antigens of *Treponema denticola*. *Biochimica et Biophysica Acta* **1794**: 1421-32.
- Veith, P.D., O'Brien-Simpson, N.M., Tan, Y., Djatmiko, D.C., Dashper, S.G., and Reynolds, E.C. (2009b) Outer membrane proteome and antigens of *Tannerella forsythia*. *J Prot Res* **8**: 4279-92.
- Veith, P.D., Talbo, G.H., Slakeski, N., Daspher, S.G., Moore, C., Paolini, R.A., and Reynolds, E.C. (2002) Major outer membrane proteins and proteolytic processing of RgpA and Kgp of *Porphyromonas gingivalis* W50. *J Biochem* **363**: 105-115.
- Xu, P., Alves, J.M., Kitten, T., Brown, A., Chen, Z., Ozaki, L.S., Manque, P., Ge, X., Serrano, M.G., Puiu, D., Hendricks, S., Wang, Y., Chaplin, M.D., Akan, D., Paik, S., Peterson, D.L., Macrina, F.L., and Buck, G.A. (2007) Genome of the opportunistic pathogen *Streptococcus sanguinis*. *J Bacteriol* **189**: 3166-75.
- Xu, P., Ge, X., Chen, L., Wang, X., Dou, Y., Xu, J.Z., Patel, J.R., Stone, V., Trinh, M., Evans, K., Kitten, T., Bonchev, D., and Buck, G.A. (2011) Genome-wide essential gene identification in *Streptococcus sanguinis*. *Sci Rep* **1**: 1-9.
- Yang, J., Yuanyuan, S., Junyi, S., Jianzhong, G., and Ang, L. (2012) Identification of common outer membrane proteins of *Porphyromonas gingivalis* by shotgun proteomics. *Afr J Microb Res* **6** 5980-6000.
- Yu, F., Anaya, C., and Lewis, J.P. (2007) Outer membrane proteome of *Prevotella intermedia* 17: identification of thioredoxin and iron-repressible hemin uptake loci. *Proteomics* **7**: 403-12.
- Yu, N., Wagner, J., Laird, M., Meili, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M.,

Foster, L., and Brinkman, F. (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26** 1608-1615.

For Review Only

Table 1. Description of selected oral bacterial species

Oral Bacterial Species Selected

Porphyromonas gingivalis W83

Gram-negative, anaerobic, key periodontal pathogen, member of Red complex

Treponema denticola ATCC 35405

Gram-negative, anaerobic, periodontal pathogen member of Red complex

Tannerella forsythia ATCC 43037

Gram-negative, anaerobic, periodontal pathogen member of Red complex

Prevotella intermedia 17

Gram-negative, anaerobic, member of Orange complex (close associations with Red complex)

Streptococcus sanguinis SK36

Gram-positive, facultative anaerobe, commensal important for plaque formation.
Associated with infective endocarditis.

Table 2. Statistical analysis of the mean mass of proteins from five genomes classified according to predicted locations extracted from PSORTdb 2.0 database of precomputed genomes (Yu *et al.* 2010). The data includes average, standard deviation, minimum, and maximum of the mean mass. The p-value was calculated using the Wilcoxon-Mann-Whitney (WMW) test. The p-value (WMW) significance refers to the mean mass of that predicted location group compared with the total average of mean mass. The known experimentally derived outer membrane proteins were extracted from the literature (Table 3) (Ang *et al.* 2008; Ge *et al.* 2010; Veith *et al.* 2009a; Veith *et al.* 2009b; Veith *et al.* 2002; Yang *et al.* 2012; Yu *et al.* 2007).

<i>Porphyromonas gingivalis</i> W83 (Ang <i>et al.</i> 2008; Veith <i>et al.</i> 2002; Yang <i>et al.</i> 2012)									
	Total	Experimental outer membrane	Annotated outer membrane	Extra cellular	Outer membrane	Peri-plasmic	Cyto plasmic membra ne	Cyto plasmic	Unkno wn
No. of proteins	1911	35	9	4	42	18	330	962	553
Avg. of mean mass	112.74	111.05	113.08	111.9	112.18	113.23	112.67	112.77	112.76
Std dev of mean mass	3.26	2.19	1.66	4.19	3.55	2.71	3.15	3.21	4.38
Min. of mean mass	95.13	106.52	111.00	106.30	101.48	106.29	100.19	101.49	95.12
Max. of mean mass	126.89	118.06	115.68	116.4	122.42	121.59	121.9	126.89	123.3
p-value (WMW)		0.00017	0.74	0.76	0.19	0.65	0.99	0.83	0.99
<i>Treponema denticola</i> ATCC 35405 (Veith <i>et al.</i> 2009a)									
No. of protein	2766	14	17	19	42	31	555	1205	914
Avg. of mean mass	113.44	111.28	112.21	109.3	112.61	110.992	111.78	113.47	114.62
Std dev of mean mass	3.81	2.82	2.91	2.83	2.49	2.4	3.36	3.28	4.31
Min. of mean mass	98.48	107.18	106.57	106.14	105.05	105.97	98.48	102.51	99
Max. of mean mass	129.91	115.64	116.82	115.31	116.79	114.66	123.74	124.38	129.91
p-value (WMW)		0.024	0.19	<0.0001	0.15	<0.0001	<0.0001	0.47	<0.0001
<i>Tannerella forsythia</i> ATCC 4307 (Veith <i>et al.</i> 2009b)									
No. of proteins	3106	231	124	18	101	19	434	1255	1066
Avg. of mean mass	113.76	112.33	112.81	112.96	113.63	114.56	113.45	113.78	113.77
Std dev of mean mass	3.37	2.76	2.06	4.28	3.86	2.74	3.56	3.29	3.42
Min. of mean mass	97.83	102.91	102.91	103.41	105.45	109.81	99.2	102.4	97.83
Max. of mean mass	131.17	118.84	116.72	119.72	131.17	119.5	125.2	125.61	125.78
p-value (WMW)		<0.0001	0.00025	0.6	0.55	0.33	0.38	0.88	0.72
<i>Prevotella intermedia</i> 17 (Yu <i>et al.</i> 2007)									
No. of proteins	2807	16	36	20	56	21	471	1091	1148
Avg. of mean mass	113.63	110.82	111.71	114.11	113.33	112.2	113.13	113.64	113.85
Std dev of mean mass	3.61	3.17	2.55	3.08	3.6	2.31	3.57	3.48	3.76
Min. of mean mass	99.17	102.96	102.84	110.23	106.2	106.26	99.17	99.77	99.98
Max. of mean mass	127.78	115.79	115.62	122.25	124.12	115.38	125.9	126.55	127.78
p-value (WMW)		0.002	0.0005	0.82	0.43	0.049	0.0055	0.77	0.092
<i>Streptococcus sanguinis</i> SK36 (Ge <i>et al.</i> 2010)									
No. of proteins	2268	10	8	23	57	N/A	658	1157	373
Avg. of mean mass	112.48	109.04	112.23	112.07	109.67	N/A	112.38	112.67	112.53
Std dev of mean mass	3.72	7.62	2.94	4.31	3.84	N/A	3.72	3.55	4.03
Min. of mean mass	93.6	96.22	107.34	101.39	99.97	N/A	96.39	93.6	97.4
Max. of mean mass	131.16	121.31	116.41	118.85	116.62	N/A	131.16	126.55	123.85
p-value (WMW)		0.07	0.84	0.92	<0.0001	N/A	0.4	0.21	0.54

Table 3. Experimentally determined, outer-membrane proteins for the five genomes extracted from the literature (Ang *et al.* 2008; Veith *et al.* 2009a; Veith *et al.* 2009b; Veith *et al.* 2002; Yang *et al.* 2012 ; Yu *et al.* 2007). The data include length of the protein, calculated mean mass and calculated pI.

<i>Porphyromonas gingivalis</i> W83												
Protein ID	Length	Mean Mass	pI	Protein ID	Length	Mean Mass	pI	Protein ID	Length	Mean Mass	pI	
PG_0026	1123	110.87	5.35	PG_0637	346	108.63	4.7	PG_1424	556	111.02	5.88	
PG_0027	391	110.66	5.23	PG_0654	390	115.04	5.75	PG_1442	142	118.06	8.53	
PG_0232	821	111.47	5.17	PG_0669	293	111.35	5.86	PG_1626	554	110.01	7.67	
PG_0290	516	109.58	8.87	PG_0694	380	111.72	7.69	PG_1798	405	112.84	5.41	
PG_0350	484	108.84	4.79	PG_0695	391	111.06	8.61	PG_1837	2105	107.94	4.58	
PG_0410	1294	111.31	5.46	PG_0707	848	111.39	5.58	PG_1844	1732	108.35	6.04	
PG_0495	469	114.94	4.81	PG_1006	894	111.43	9.16	PG_2008	833	111.34	5.74	
PG_0506	736	110.01	5.52	PG_1028	448	111.41	5.13	PG_2024	1706	108.84	4.99	
PG_0553	940	109.05	5.98	PG_1030	446	112.38	4.79	PG_2102	540	113.2	5.25	
PG_0611	316	112.83	5.6	PG_1242	528	110.56	5.48	PG_2173	290	111.78	5.74	
PG_0614	326	112.23	5.72	PG_1326	352	111.17	5.1	PG_2216	576	106.52	6.75	
PG_0616	344	109.26	6.61	PG_1414	871	109.78	8.01					
<i>Treponema denticola</i> ATCC 35408												
Protein ID	Length	Mean Mass	pI	Protein ID	Length	Mean Mass	pI	Protein ID	Length	Mean Mass	pI	
TDE0405	543	107.18	6.34	TDE1273	531	113.56	5.36	TDE2028	1338	109.57	5.76	
TDE0664	422	115.64	6.1	TDE1511	208	109.13	4.84	TDE2242	821	110.72	5.22	
TDE0761	639	109.08	5.59	TDE1658	358	112.85	8.65	TDE2601	818	114.67	6.12	
TDE0951	348	108.91	5.28	TDE1663	167	114.97	8.98	TDE2699	555	111.17	8.46	
TDE1071	591	112.86	5.287	TDE1950	357	107.56	5.47					
<i>Prevotella intermedia</i> 17												
Protein ID	Length	Mean Mass	pI	Protein ID	Length	Mean Mass	pI	Protein ID	Length	Mean Mass	pI	
PIN_0093	190	107.67	5.41	PIN_A0611	802	111.19	6.5	PIN_A1341	389	112.45	6.17	
PIN_0414	226	102.96	9.35	PIN_A0975	509	109.75	9.12	PIN_A1373	1040	108.31	7.83	
PIN_A0102	381	110.28	6.15	PIN_A1050	184	109.3	7.89	PIN_A1455	390	110.52	9.39	
PIN_A0324	486	113.71	5.67	PIN_A1184	278	112.57	5.11	PIN_A1480	693	115.18	9.23	
PIN_A0326	1057	110.51	8.72	PIN_A1204	277	113.67	6.77	PIN_A1542	413	109.21	7	
PIN_A0437	341	115.79	6.52									
<i>Tannerella forsythia</i> ATCC43037												
Protein ID	Length	Mean Mass	pI	Protein ID	Length	Mean Mass	pI	Protein ID	Length	Mean Mass	pI	
TF0015	230	104.83	8.89	TF0977	546	116.28	6.11	TF2207	466	109.71	6.57	
TF0041	815	113.29	9.23	TF0980	818	112.89	8.99	TF2214	196	109.91	9.28	
TF0044	424	111.82	8.8	TF1015	508	113.97	9.47	TF2301	1111	112.34	7.62	
TF0045	795	113.8	9.13	TF1032	489	111.89	8.41	TF2302	452	113.93	6.69	
TF0063	767	116.13	9.17	TF1033	677	114.23	5.41	TF2320	1887	109.43	7.98	
TF0064	638	117.26	9.22	TF1038	421	116.25	9.44	TF2327	480	114.89	5.63	
TF0071	928	112.43	9.14	TF1039	205	115.08	9.46	TF2330	370	115.45	9.48	
TF0090	461	108.38	6.07	TF1052	600	112.46	6.33	TF2333	469	117.41	9.25	
TF0091	379	110.7	6.06	TF1053	1064	110.24	8.51	TF2339	1816	109.93	5.83	
TF0092	527	111.3	5.57	TF1055	284	114.93	8.59	TF2347	286	102.91	5.68	
TF0093	1014	110.72	8.54	TF1056	626	113.83	5.62	TF2349	550	112.03	6.42	
TF0111	1041	110.63	9.05	TF1057	1032	110.16	9.07	TF2402	521	112.98	5.07	
TF0112	568	114.84	7.01	TF1059	1019	112.35	7.11	TF2403	1166	112.81	9.03	
TF0183	178	106.86	8.66	TF1101	277	116.06	9.08	TF2411	514	114.92	5.47	
TF0216	115	115.73	11.58	TF1123	414	115.65	8.2	TF2412	1099	111.9	8.33	
TF0217	65	112.37	11.32	TF1150	903	112.97	5.5	TF2414	230	113.51	5.87	
TF0220	500	111.84	6.25	TF1151	87	114.21	8.37	TF2415	229	110.41	6.73	
TF0237	1062	112.68	7.2	TF1158	144	110.21	4.49	TF2416	517	112.46	5.27	
TF0238	669	113.42	6	TF1193	422	112.98	6.28	TF2417	1055	110.26	8.52	
TF0275	1009	111.53	8.24	TF1201	112	109.34	9.57	TF2421	536	112.84	8.7	
TF0277	544	113.11	5.36	TF1206	514	115.1	5.68	TF2447	454	111.89	8.99	

TF0299	155	110.98	8.83	TF1207	1137	111.82	8.72	TF2450	902	114.02	7.95
TF0301	880	112.04	7.13	TF1245	187	113.14	6.75	TF2531	686	114.41	5.27
TF0304	238	106.69	5.52	TF1259	2192	109.37	6.14	TF2551	101	112.8	9.9
TF0305	251	108.73	5.05	TF1300	227	107.03	8.77	TF2552	205	107.35	9.61
TF0312	562	114.85	5.83	TF1318	1054	111.31	6.88	TF2560	257	111.74	10.1
TF0313	998	113.37	8.85	TF1319	503	114.84	5.38	TF2566	186	113.99	9.75
TF0318	770	113.36	8.67	TF1325	597	110.58	6.66	TF2569	183	110.44	9.78
TF0322	514	117.29	8.94	TF1331	389	114.12	8.82	TF2574	446	110.49	9.91
TF0324	230	115.47	9.44	TF1342	472	115.5	9.06	TF2579	201	114.06	9.96
TF0334	112	111.01	9.6	TF1351	401	115.54	9.44	TF2592	1216	108.43	6.23
TF0348	146	110.24	8.42	TF1404	388	114.9	5.11	TF2595	518	114.01	9.12
TF0365	141	114.56	9.06	TF1409	507	115.52	9.25	TF2596	411	111.81	5.43
TF0368	118	112.23	8.78	TF1413	210	111.04	5.66	TF2597	952	112.18	9.31
TF0399	451	115.81	9.02	TF1415	1091	111.34	7.97	TF2605	780	111.94	7.64
TF0405	154	113.18	9.51	TF1416	586	111.35	5.71	TF2606	404	110.8	5.7
TF0421	446	113.92	6.9	TF1440	327	113.25	6.56	TF2613	414	113.72	5.94
TF0424	1005	111.35	5.97	TF1441	210	111.84	9.04	TF2646	540	108.31	9.12
TF0425	519	112.68	5.66	TF1443	455	109.79	9.26	TF2649	646	112.17	6.96
TF0436	514	118.84	9.58	TF1444	536	112.63	8.74	TF2650	251	108.65	6.88
TF0439	442	106.36	8.74	TF1476	467	114.02	9.14	TF2661	559	115.41	6.71
TF0447	104	102.92	9.06	TF1505	537	112.67	5.2	TF2663	1364	111.72	9.09
TF0454	429	105.76	8.25	TF1506	717	112.2	8.84	TF2725	1096	112.35	6.81
TF0477	519	111.48	9.16	TF1525	145	108.16	5.43	TF2726	176	116.38	6.58
TF0482	1097	111.75	8.58	TF1534	162	112.39	6.41	TF2728	1190	111.6	9.17
TF0483	581	115.8	5.57	TF1535	720	113.48	9.04	TF2729	495	114.74	6.7
TF0508	724	115.51	9.35	TF1565	277	112.58	8.89	TF2734	206	116.45	9.47
TF0546	282	115.28	4.75	TF1575	213	112.48	5.51	TF2778	783	114.47	9.1
TF0587	511	113.98	5.75	TF1595	333	113.94	5.37	TF2801	1179	112.74	6.88
TF0588	1154	112.08	7.03	TF1605	1125	111.51	8.17	TF2802	631	115.68	6.44
TF0640	1049	112.42	8.67	TF1606	504	116.31	5.15	TF2803	493	111.07	8.22
TF0641	486	114.29	5.88	TF1733	156	113.54	9	TF2804	292	111.7	5.19
TF0652	717	110.18	8.98	TF1741	1210	107.98	6.19	TF2806	249	114.63	9.33
TF0654	1038	113.3	6.38	TF1755	338	115.7	7.09	TF2838	228	113.88	8.91
TF0655	627	113.97	6.46	TF1775	468	110.89	7.3	TF2843	289	113.81	5.64
TF0661	263	112.22	8.51	TF1793	382	112.61	9.25	TF2852	392	111.34	4.87
TF0682	1001	112.17	8.76	TF1822	505	115.98	6.55	TF2901	2482	113.69	7.06
TF0683	443	110.19	8.27	TF1843	1092	106.58	5.58	TF2920	396	111.05	6.47
TF0706	412	113.99	9.42	TF1897	923	114.5	9.33	TF2924	985	114.04	5.67
TF0743	108	115.83	10.42	TF1940	594	110.45	9.25	TF2925	777	111.58	5.46
TF0749	730	113.74	5.31	TF1959	350	110.67	8.65	TF2998	1087	105.42	5.42
TF0750	281	113.41	6.93	TF1964	247	106.8	5.54	TF3006	241	115.3	5.12
TF0761	863	115.04	9.06	TF1970	387	106.24	8.44	TF3007	228	105.41	9.57
TF0765	211	117.61	9.55	TF1989	1164	111.44	7.21	TF3011	1144	112.27	8.64
TF0773	463	112.56	6.39	TF1990	595	112.56	6.56	TF3012	594	115.06	6.24
TF0778	1013	112.64	8.6	TF2016	155	113.96	6.09	TF3013	434	115.63	5.55
TF0779	483	112.1	6.37	TF2031	554	113.31	7.49	TF3024	477	113.51	8.98
TF0789	766	108.82	5.28	TF2032	1023	109.57	6.84	TF3036	435	105.66	8.45
TF0797	487	109.41	5.89	TF2035	369	114.66	5.92	TF3080	1288	109.31	6.26
TF0810	393	114.52	7.79	TF2096	1085	111.65	8.99	TF3087	875	111.87	9.64
TF0813	475	115.61	4.8	TF2116	1252	104.74	7.12	TF3099	189	113.67	9.12
TF0841	268	110.86	7.56	TF2123	999	114.46	5.37	TF3103	396	114.49	5.64
TF0875	781	115.75	9.3	TF2124	609	113.96	9.31	TF3104	1062	111.15	8.92
TF0945	366	110.43	6.25	TF2190	185	112.18	5.67	TF3114	375	115.63	9.78
TF0955	1143	111.73	6.21	TF2192	672	113.24	5.51	TF3137	205	108.66	7.59
TF0959	1081	112.83	7.17	TF2193	1075	112.85	7.55	TF3163	1933	110.05	8.71
TF0976	1067	110.39	5.88	TF2206	313	109.58	5.34	TF3165	374	110.34	7.55

***Streptococcus sanguinis* SK36**

Protein ID	Length	Mean Mass	pI	Protein ID	Length	Mean Mass	pI	Protein ID	Length	Mean Mass	pI
SSA_67	448	109.82	5.35	SSA_699	341	103.81	5.28	SSA_1573	1510	105.1	4.9
SSA_128	500	116.2	5.04	SSA_712	397	96.22	4.73	SSA_2228	332	117.77	5.33
SSA_136	797	103.76	4.53	SSA_844	405	110.12	8.24				
SSA_205	427	121.31	5.06	SSA_979	364	106.32	9.82				

FIGURE LEGENDS

Figure 1. Frequency vs mean amino acid mass in the five genomes. As a representation of the relative *de novo* amino acid synthetic cost, the mean amino acid mass of each protein calculated as the mass divided by the number of amino acids for each protein (Smith and Chapman 2010). The median values of the mean amino acid mass for the genomes of *P. gingivalis* (112.776) and *S. sanguinis* (112.557) are significantly different ($p < 0.0001$) to the median values of *T. denticola* (113.385), *T. forsythia* (113.829) and *P. intermedia* (113.624) based on the Wilcoxon-Mann-Whitney U-test.

Figure 2. Relative amino acid abundancies in the five genomes. The amino acids are ordered on a scale of decreasing hydrophobicity. The relative amino acid frequencies are typical of prokaryotic genomes (Smole *et al.*, 2011).

Figure 3. Comparison of physico-chemical properties of overall amino acids of the five genomes. The amino acid compositions of the proteins were extracted from an online database (COPid) (Kumar *et al.* 2008). These commonly used physico-chemical properties complement the relative amino acid abundancies (Figure2) and provide a more comprehensive description of the proteomes that enable the interpretation of biophysical adaptations to specific environments (Smole *et al.*, 2011). These profiles are typical of non-halophilic and non-thermophilic prokaryotes.

Figure 4. Comparison of the number of proteins in predicted cellular location for the five genomes. The predicted cellular locations were extracted from PSORTb (Yu *et al.* 2010).

131x95mm (300 x 300 DPI)

107x66mm (300 x 300 DPI)

118x72mm (300 x 300 DPI)

108x66mm (300 x 300 DPI)

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Cross, KJ;Huq, NL;Park, SH;Park, JS;Pham, P;Quah, M;Ranjan, M;Reynolds, EC

Title:

Bioinformatic investigation of the cost management strategies of five oral microbes

Date:

2015-04-01

Citation:

Cross, K. J., Huq, N. L., Park, S. H., Park, J. S., Pham, P., Quah, M., Ranjan, M. & Reynolds, E. C. (2015). Bioinformatic investigation of the cost management strategies of five oral microbes. MOLECULAR ORAL MICROBIOLOGY, 30 (2), pp.87-96. <https://doi.org/10.1111/omi.12071>.

Persistent Link:

<http://hdl.handle.net/11343/108633>