

Pg London 2015

Comparative genomics and proteomics of 13 *Porphyromonas gingivalis* strains

Tsute Chen¹, Huma Siddiqui² and Ingar Olsen²

¹Department of Microbiology, The Forsyth Institute, Cambridge, MA, USA; ²Department of Oral Biology, University of Oslo, Oslo, Norway

At the current time, genome sequences of a total of 13 *Porphyromonas gingivalis* strains are available, including five completed genomes (strains ATCC 33277, HG66, TDC60, JCVISC001, and W83) and eight high-coverage draft sequences (F0185, F0566, F0568, F0569, F0570, SJD2, W4087, and W50) that are assembled into fewer than 300 contigs. This study compared these genomes at both nucleotide and protein sequence levels in order to understand their phylogenetic and functional relatedness. There are four copies of 16S rRNA gene sequences in each of the strains of ATCC 33277, HG66, TDC60, and W83 and one copy in the other nine genomes. These 25 16S rRNA sequences represent only 13 unique sequences. The five copies in W83 and W50 are identical and the three copies in HG66 are identical to the four copies in ATCC 33277, suggesting close evolutionary lineage between W83 and W50, as well as HG66 and ATCC 33277. Genome-wide comparison based on “Rapid Annotation using Subsystem Technology” (RAST) also showed that for the overall biological functions of the genomes, W83 is closer to W50, and HG66 to ATCC33277, than to other genomes. The comparison of the RAST subsystems identified biological functions that are unique to individual, shared by some, or by all genomes. Functions unique to individual genomes include: a tetracycline resistance protein TetQ, DNA metabolism gene YcfH, and DNA repair gene exonuclease SbcC (only in SJD2); very-short-patch mismatch repair endonuclease and a phage packaging terminase similar to *Bacteroides* phage B124-14 (in W4087); an internalin similar to a *Listeria* surface virulence protein (W83); a Type I restriction-modification system (F0569); an iron acquisition/heme transport protein (F0566); colicin I receptor and carbamoylputrescine amidase (W50); L-serine dehydratase (TDC60); and spermidine synthase and ribokinase (JCVISC001). The results also identified biological functions that are missing in individual or several genomes. For example, JCVISC001 does not contain the CRISPR (clustered regularly interspaced short palindromic repeats) system – a prokaryotic immune system that confers resistance to foreign genetic elements such as plasmids and phages. Some genomes are enriched with multiple copies of certain genes [e.g., TDC60, W50, and W83 encode 2–4 copies of 4-alpha-glucanotransferase (amylomaltase in glycan metabolism)], while others only have a single copy in the genome. Complete results of this study will be presented and available online for download.