Genetic background of Porphyromonas gingivalis capsule biosynthesis

Brunner, J.

Publication date
2011

Citation for published version (APA):
Brunner, J. (2011). *Genetic background of Porphyromonas gingivalis capsule biosynthesis*.

Chapter 5

The core genome of the anaerobic oral pathogenic bacterium *Porphyromonas gingivalis*

Jorg Brunner[1], Floyd R. A. Wittink[2], Martijs J Jonker[2], Mark de Jong[2], Timo M. Breit[2], Marja L. Laine[3], Johannes J. de Soet[1] and Wim Crielaard[§1]

[1] Department of Oral Microbiology, Academic Centre for Dentistry Amsterdam, University of Amsterdam and VU university Amsterdam Amsterdam, Amsterdam, the Netherlands
[2] Microarray Department and Integrative Bioinformatics Unit, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, the Netherlands
[3] Department of Periodontology, Academic Center for Dentistry Amsterdam, University of Amsterdam and VU university Amsterdam, Amsterdam, the Netherlands

Abstract

The Gram negative anaerobic bacterium *Porphyromonas gingivalis* has long been recognized as a causative agent of periodontitis. Periodontitis is a chronic infectious disease of the tooth supporting tissues eventually leading to tooth-loss. Capsular polysaccharide (CPS) of *P. gingivalis* has been shown to be an important virulence determinant. Seven capsular serotypes have been described. Here, we used micro-array based comparative genomic hybridization analysis (CGH) to analyze a representative of each of the capsular serotypes and a non-encapsulated strain against the highly virulent and sequenced W83 strain. We defined absent calls using *Arabidopsis thaliana* negative control probes, with the aim to distinguish between aberrations due to mutations and gene gain/loss.

Our analyses allowed us to call aberrant genes, absent genes and divergent regions in each of the test strains. A conserved core *P. gingivalis* genome was described, which consists of 80% of the analyzed genes from the sequenced W83 strain. The percentage of aberrant genes between the test strains and control strain W83 was 8.2% to 13.7%. Among the aberrant genes many CPS biosynthesis genes were found. Most other virulence related genes could be found in the conserved core genome. Comparing highly virulent strains with less virulent strains indicates that *hmuS,* a putative CobN/Mg chelatase involved in heme uptake*,* may be a more relevant virulence determinant than previously expected. Furthermore, the description of the 39 W83-specific genes could give more insight in why this strain is more virulent than others.

Analyses of the genetic content of the *P. gingivalis* capsular serotypes allowed the description of a *P. gingivalis* core genome. The high resolution data from three types of analysis of triplicate hybridization experiments may explain the higher divergence between *P. gingivalis* strains than previously recognized.

Background

Periodontitis is a chronic destructive infectious disease of the tooth-supporting tissues. It is one of the most prevalent infectious diseases in the world. With percentages of moderate disease ranging from just below 20% in an age group of 30 to 40 year-olds in Swedish and Norwegian studies to even up to 38% of severe cases in the United States in an on average 75 year-old male population [1-3]. Besides high prevalence also links to systemic diseases have been described. Periodontitis has been associated with, amongst others, cardiovascular diseases, diabetes mellitus and rheumatoid arthritis [4-7].

Periodontitis leads to loss of sound teeth as supporting bone and connective tissue are slowly degraded as a result of an exaggerated host immune response triggered against a polymicrobial biofilm [8].

In the oral cavity around 7000 species can be detected, in subgingival and supragingival biofilm/plaque over 400 bacterial species are present [9-11]. Many disease-related bacterial species in the subgingival plaque have been shown to be Gram negative anaerobes. Among them, *Porphyromonas gingivalis* a black-pigmented bacterium from the phylum *Bacteroidetes* is a major causative agent in periodontal disease [12].

Interaction with other bacteria residing in the periodontal pocket is important to sustain the infectious biofilm. One of the structures involved in the inter-species adherence is the capsular polysaccharide (CPS) of *P. gingivalis* [13]. CPS has been described as a virulence factor of various pathogenic bacteria, mainly as being involved in evasion of the host immune system [14-16]. In *P. gingivalis* encapsulated strains have been shown to be more resistant to serum killing and phagocytosis. The explanation for this increased resistance compared to the non-encapsulated strains may be the increased hydrophilicity and the lower induction of the alternative complement pathway [17].

Encapsulated *P. gingivalis* strains have also been shown to be more virulent than non-encapsulated strains in the mouse infection model [18]. To date, six capsular serotypes (K1–K6) have been described [19, 20] and a seventh serotype (K7) has been suggested by R. E. Schifferle (personal communication). In a mouse subcutaneous infection model several strains of each of the serotypes have been shown to be highly virulent [18]. The variation of virulence within serotypes shows that besides CPS there have to be more virulence factors of importance in *P. gingivalis*. Many of its virulence factors have been studied in the last decades including fimbriae,

93

hemagglutinins, lipopolysaccharide (LPS), outer membrane proteins (OMPs) and an extremely wide variety of proteinases. High quality reviews have been published on the wide variety of *P. gingivalis* virulence factors [21-23].

Using comparative whole-genome hybridization analysis of the encapsulated W83 strain and the non-encapsulated ATCC33277 a CPS biosynthesis locus had been found, after which a knock-out study has proven that the CPS locus was functional [24, 25]. K1 CPS from W83 has been shown to induce a stronger chemokine response than CPS from the other serotypes in murine macrophages [26]. Recent work in our group, however, has shown that an isogenic W83 mutant lacking CPS triggers a higher pro-inflammatory immune response in human gingival fibroblasts than strain W83 carrying K1 CPS [27]. The exact roles of CPS in *P. gingivalis* are still unclear, but reducing the host immune response is certainly one of them.

In the latest years an increasing number of genomes have been sequenced paving the path for genomics-based approaches. For *P. gingivalis* genome sequences of the virulent strain W83 and the less-virulent strain ATCC33277 have become available [28, 29]. Comparative genomic hybridization (CGH) analysis using microarrays of these well-described bacterial strains could yield new insights in the virulence mechanisms of *P. gingivalis*. A recent study reported on the CGH analysis of several *P. gingivalis* strains to describe the genetic variety among them [30].

In this study we analyzed the genetic contents of representative strains of each of the seven capsular serotypes (table 1): W83 (K1), HG184 (K2),

Table 1. *P. gingivalis* strains used in this study

Strain	Capsular serotype	Origin	Virulence[c]
W83[a]	K1	Clinical specimen	High
HG184	K2	Periodontitis patient	Medium
HG1025	K3	Periodontitis patient with diabetes mellitus	High
ATCC49417	K4	Advanced adult periodontitis patient	High
HG1690	K5	37-year-old male periodontitis patient	High
HG1691	K6	28-year-old female periodontitis patient	Medium
34-4	K7	Severe periodontitis patient	Low
FDC381[b]	K-	Adult periodontitis patient	Low

[a] A kind gift of H. N. Shah (NCTC, London, UK)
[b] A kind gift of S. S. Socransky (The Forsyth Institute, Boston, MA, USA)
[c] As determined in a subcutaneous mouse infection model [18, 32]

ATCC53977 (K3), ATCC49417 (K4), HG1690 (K5), HG1691 (K6), 34-4 (K7). We also included the non-encapsulated strain FDC381 (K-) in the CGH analysis to compare with each of the encapsulated strains. Strain FDC381 does however express a non-CPS anionic extracellular polysaccharide as do the other strains [31]. The strains were classified into three virulence levels as determined by using a subcutaneous mouse infection model [18, 32]. Although not an optimal measure for the ability to cause periodontitis, this classification has long been used [33] and proven useful in studying virulence determinants [34-37].

Triplicate hybridization experiments and three types of analysis, 1) aberrant gene calling, 2) breakpoint analysis and 3) absent gene calling, have been performed for optimal use of the new genetic information. The careful design of the experiment and the thorough analysis of the data lead to a high resolution data set, yielding more detailed information on the genetic differences between strains than has been shown before. In this study we initiate the description of a core-gene set of *P. gingivalis* allowing a more focused search for potential important virulence factors.

Results and discussion

Microarray performance and data interpretation

The *P. gingivalis* version 1 microarray from the PFGRC used in this study has been used in several studies before [30, 38] and consists of 1907 probes and 500 negative control probes (*Arabidopsis thaliana*) printed in four replicates. The microarray was designed to cover all non-redundant coding sequences (CDSs) of the *P. gingivalis* W83 genome. Before our study all probes were analyzed for their unique- and perfect matching with the genome, as downloaded from the NCBI, using BLAST. Twenty-nine of the 1907 probes of the microarray gave non-specific hits, mostly related to transposases (table 2). These probes were excluded from further analyses together with four probes that were not in use anymore annotated "obsolete" by the manufacturer, so that 1874 probes remained. The comparison of each test strain to W83 using this array gives insights into described virulence associated genes. A limitation of the method, however, is that genes from the variable gene pool from other strains will not be detected.

Data were normalized and technical and biological replicates were collapsed as described in the Materials and Methods. Detailed analysis of the probe intensities indicated that 22 probes gave systematically low intensity values for strain W83 as well as for all the other strains. The
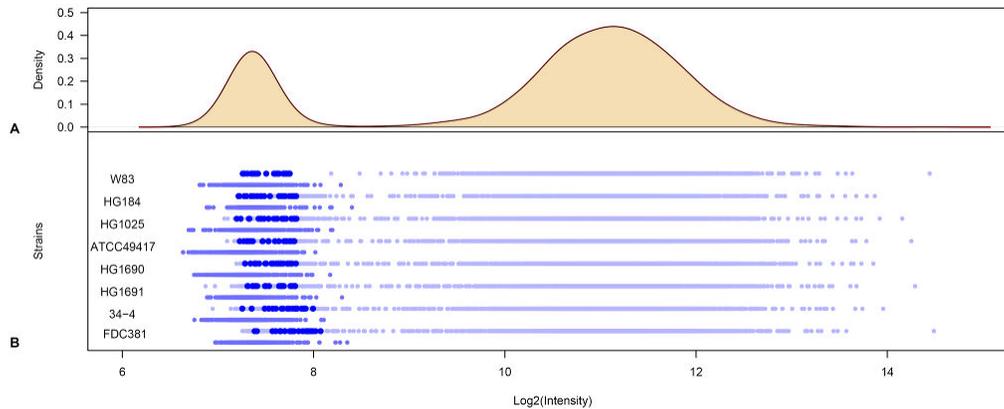
Figure 1 - Hybridization signals of *P. gingivalis* strains – dead probes.
A. The total intensity distribution of probe signals of W83 DNA hybridized to the W83 array. The density peak around 7.5 contains the negative controls (empty spots and *A. thaliana* probes). The peak around 12 should contain all present genes in strain W83. B Probe signal intensities of each *P. gingivalis* test strain are represented in light blue dots; medium blue dots, slightly below that, symbolize *A. thaliana* negative control genes. Dark blue dots represent *P gingivalis* probes, which show the same low intensity as the negative control probes. These 22 probes are called dead probes as they do not give any significant hybridization signal.

intensity levels were at the same low levels as the intensity levels of the negative control probes (figure 1). These probes were labeled as "dead probes" and excluded from the results (table 3). Our data do not explain why dead probes have occurred in our experiments, but the consistent low signal for these probes suggests that the sequencing information used for designing these probes was imperfect.

In order to maximize the mining of the genomic information, we subjected the data to three complementary analyses: 1) analysis for aberrations as detected by individual probes, 2) analysis for breakpoints, and 3) analysis for genomic loss. The rationale behind the three analyses is as follows. The probed genomic sites are on average 1250 bp apart from each other (median was 1018), which was not considered to be a high interrogation density. We therefore decided to analyze each probe individually for indication that the genomic site interrogated is aberrant from W83. Deviations from W83 that were detected with a false discovery rate corrected p-value (FDR) < 0.05 were considered significant. This aberrance could have occurred due to mutations or loss (or due to W83 gain), and this was regarded as point-variability between the strains. Nevertheless, if several neighboring probes indicate aberrations, then this may indicate
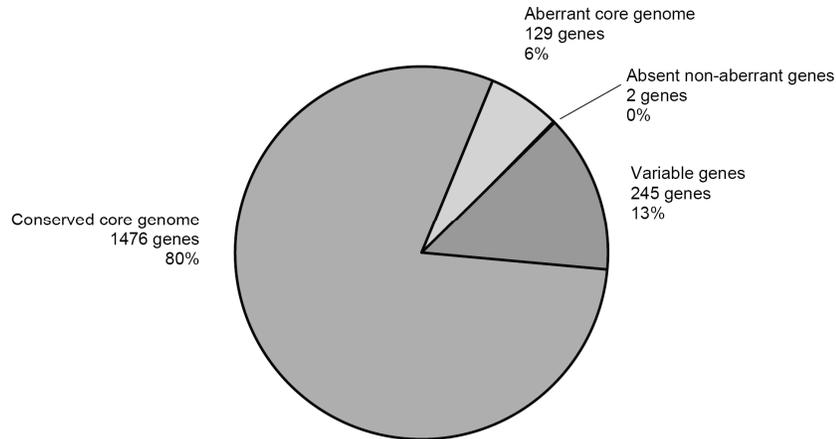
96

Figure 2. *P. gingivalis* core genome. Pie diagram representing all probes included in the results divided into pieces representing the conserved core genome, aberrant core genome and the variable genes. The percentages show the proportions of the total of functional probes. 80% of the strain W83 genes is present and conserved among the test strains. 6% of the W83 genes is present but aberrant and 13% of the genes is absent in at least one of the test strains. Two probes with very low signals were found as non-aberrant but absent.

highly variable regions due to mutations or loss. Hence, a breakpoint analysis was executed to quantitatively specify such regions. Finally, we used the negative controls to define absent calls with the aim to distinguish whether an aberration was found more likely due to mutation or loss. If the probes that indicated aberrations in the first analysis also showed the same intensities as the negative controls with FDR corrected p-value < 0.01 (see M&M), the genomic site was considered as mutated, and otherwise it was considered as lost. This last analysis enhanced our interpretation of the data and the definition of the core genome.

P. gingivalis core genome

Research on microbial pathogens is mostly performed to unravel mechanisms of virulence in order to design effective treatments. Virulence mechanisms present in all strains of a species are especially attractive. The description of a core set of genes present in a species is thus a key step for better understanding. From an analysis of eight *P. gingivalis* strains we found that 1476 genes were non-aberrantly present in each of the strains and 2 hypothetical genes were called absent but non-aberrant (Additional file 1). The conserved core genes make up 80 % of all genes included in this study. Hence, 20 % (374) of all genes of W83 were aberrant in at least one of the strains. Core genomes from several bacterial species have been

described [39-45]. The fraction of a bacterial genome that consists of core genes depends highly on the amount of strains included to describe the core genome [43]. The more strains are used, the smaller the core genome will be. As such, the very well studied *Escherichia coli* core genome makes up only 46% of the average *E. coli* genome. Other bacterial species, including Gram positives and Gram negatives, have been found to have a core genome which covers 52% to 85% of a genome [39-45]. The 80% of W83 genes which are part of the conserved core genome can therefore be understood. It must be clear though that the core genome of *P. gingivalis* as described here must be seen as a first step. The core genome will be found to be smaller as more genetic information on different *P. gingivalis* strains will become available.

We could distinguish two gene sets in the aberrant set, namely the present and absent genes (figure 2). Using aberrance and absent call analysis we were thus able to describe the *P. gingivalis* core genome in two ways. Aberrance represents mutations within the probe sequence, whereas absent calls represents the total absence of the probe sequence interpreted as gene absence. The fully conserved *P. gingivalis* core genome is comprised of 1476 genes. The variable core genome is comprised of a total of 1605 genes, which are aberrant, but called present (figure 2). In the further analyses the conserved core genome was taken as the core genome.

Combining our findings on the core genome with a study describing 1490 conserved CDSs when comparing the genome sequences of W83 and ATCC33277 [28], makes it tempting to speculate that the core genome as described here may already be close to its final size. An analysis combining the conserved CDSs from that study with our 1476 conserved core genes showed that when strain ATCC33277 is included the core genome size decreased to 1384 genes.

The conserved core gene set was analyzed for the presence of virulence genes. When it was analyzed for the presence of the 153 potentially virulence associated gene set originating from when the genome sequence of W83 became available (selected by presence of a signal peptide, or transmembrane domains or by homology to previously described virulence agents) [29], it was found that 128 genes were present in all strains (figure 3A). Only 25 genes were aberrant in at least one strain, among which 9 usual suspects from the CPS locus, but also four hemagglutinins.

Another virulence gene set was also tested for presence in the conserved core gene set of *P. gingivalis*. The set was composed of genes shown to be up-regulated in infection experiments [46, 47]. Genes up-regulated in an *in*
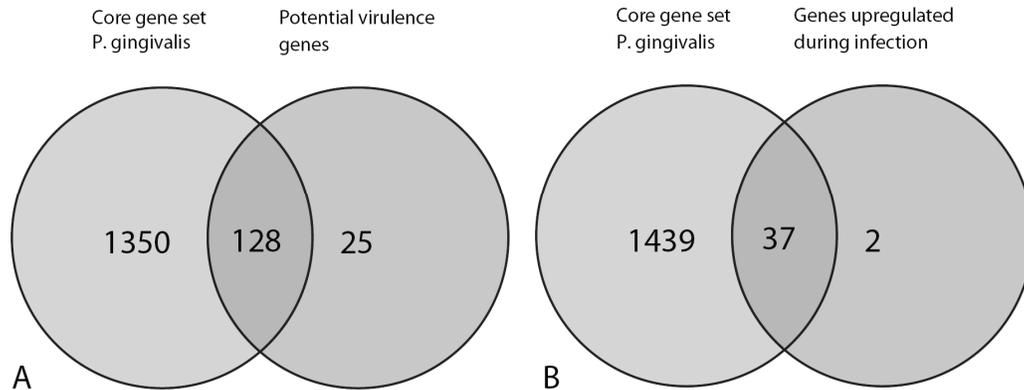
Figure 3. Virulence associated genes in the conserved core genome of *P. gingivalis.* A. 153 potential virulence genes from the genome annotation of W83 combined with the conserved core genome of *P. gingivalis* [29]. B 39 genes known to be up-regulated during infection combined with the conserved core genome of *P. gingivalis* [46, 47]. The number in the overlapping part of the circles is the number of potential virulence associated genes that was found in the conserved core genome of *P. gingivalis*

vitro human epithelial cell infection experiment were combined with a gene set *in vivo* up-regulated on protein level in a mouse subcutanuous chamber experiment to make a set of 39 virulence genes. The former experiment was chosen as an early response gene set, whereas the latter includes genes involved in sustaining an infection *in vivo*. 37 of the 39 virulence genes were present among the core gene set (figure 3B). The two genes that were not in the core gene set were a thiol protease (*PG1055*) [48] and *tetR* a transcription regulator (*PG1240*). The thiol protease is aberrant in each strain except for strain ATCC49417, from the 16S-23S ISR heteroduplex type that together with the type of strain W83 has the highest association with disease [49]. This is another indication that this thiol protease may be an important determinant in virulence of *P. gingivalis*. Transcription regulator *tetR* was only found to be aberrant in strain FDC381, which is the least virulent and the only non-encapsulated strain [18, 32].

The analysis of the core gene set shows the presence of almost all virulence related genes. The genes that are not present in the core genome may be determinants of the differences in virulence found between the strains.

Table 2. Aberrant and absent CDSs of *P. gingivalis* strains.

Strain	Aberrant CDSs	% aberrant	Absent CDSs	% absent
HG184	213	11,4	133	7,8
HG1025	214	11,4	135	7,8
ATCC49417	153	8,2	88	4,7
HG1690	187	10,0	107	5,7
HG1691	227	12,1	158	8,5
34-4	207	11,0	126	6,8
FDC381	256	13,7	195	10,5

Strain divergence

The divergences of the test strains were determined by the percentage of aberrant CDSs from the total number of 1874 CDSs included in this study. We found 8.2 % to 13.7 % of aberrant genes per strain, with ATCC49417 having the lowest and FDC381 having the highest percentage of aberrant genes (table 4). These percentages of aberrant genes are higher than the 7 % of aberrant genes from a previous genomic hybridization study on strain ATCC33277, a close relative of strain FDC381 [25]. From the 64 highly aberrant genes in ATCC33277 41 genes were included in our study from which 33 were in the aberrant gene list of strain FDC381. A recent study reported even lower percentages 0-5.1% of divergence between *P. gingivalis* strains [30]. Although they used the same arrays and also used some identical strains the differences between our data sets were substantial. We detect a much higher number of aberrant genes probably because of higher resolution due to the use of three arrays per strain. We also excluded a set of 55 genes before the analyses (see above) which further elevated the percentages found in this study.

Proteases

P. gingivalis is known to have a vast arsenal of proteases. The main function of these enzymes is to provide peptides for growth. These peptides can be derived from host-proteins, involved in defence against pathogens, thereby potentially disrupting the host immune response. Other proteases degrade collagen, thereby weakening the tooth-supporting tissues. Proteases have therefore been regarded as important virulence factors. A selection of 64 proteases/peptidases was made by text searches in the *P. gingivalis* W83 genome annotation combined with peptidases found in the MEROPS *P. gingivalis* peptidase database [50] (http://merops.sanger.ac.uk/index.shtml). This selection was analyzed for presence in the test strains. From the analysis it was clear that most proteases, 58 in total, belong to the core gene set of *P. gingivalis*. From the 6 non-core protease genes (table 5) *tpr* was already mentioned earlier. The gene *prtC*, a collagenase, was found to be aberrant only in three strains with medium/low virulence in a subcutaneous mouse model. Interestingly, in early studies on *P. gingivalis* virulence one of the discriminatory factors between virulent and avirulent strains was described to be collagenase activity, which was found to be low in avirulent strains [51]. Another non-core protease gene is the well-described *rgpA*, an arg-gingipain which has regularly been described as one of the most important virulence factors of *P. gingivalis* [52, 53]. *RgpA* is aberrant in the highly virulent strain ATCC53977. This finding is however in line with a murine periodontitis model study in which *rgpA* was found to be not important in virulence using *P. gingivalis* knockouts [34]. From the present study, however, no hard conclusion should be drawn as no functional changes have been explored.

Table 3. Non-core protease genes of *P. gingivalis*.

| GeneID	Annotation	Remark
PG0317	peptidase, M49 family	Aberrant only in 34-4
PG1055	thiol protease	Non-aberrant only in W83 and ATCC49417 (absent in FDC381)
PG1542	collagenase	Aberrant in HG1691, 34-4 and FDC381
PG2024	hemagglutinin protein HagE	Aberrant and absent only in HG1025
PG2115	protease PrtT, degenerate	Non-aberrant only in W83
PG2185	transporter, putative	Aberrant in HG184, HG1025 and FDC381
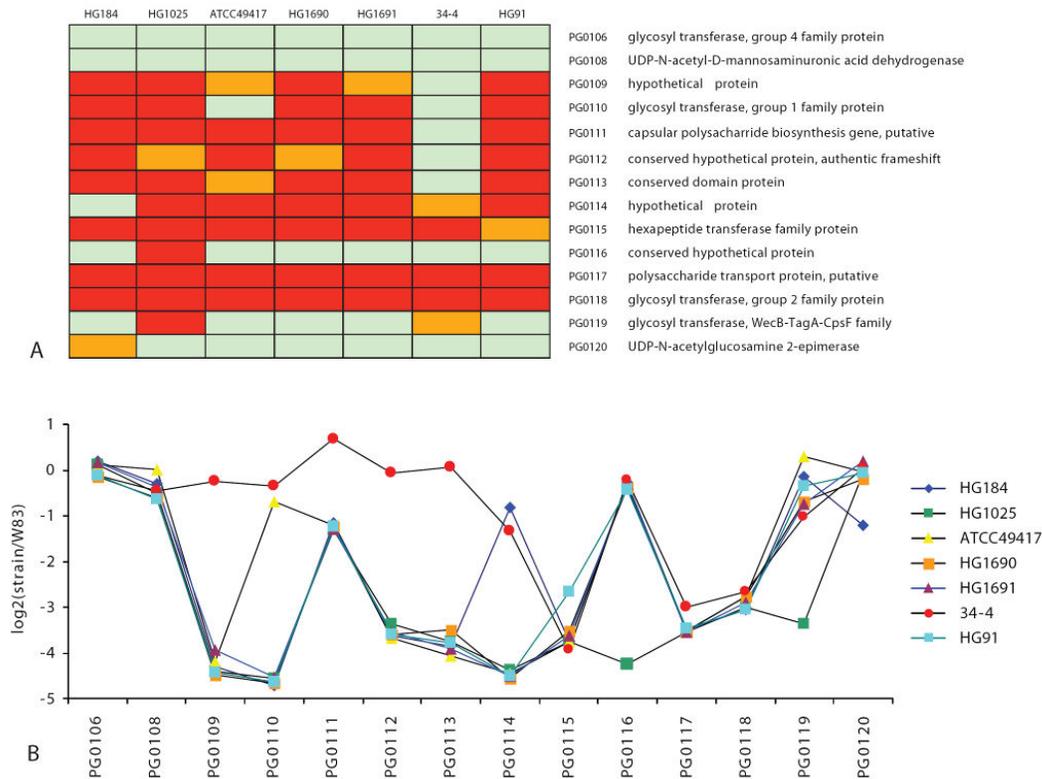
Figure 4. CPS biosynthesis locus diversity. A. Heat map showing presence (green), aberrance (orange) and absence (red) of each gene in each test strain, showing the variation within the CPS biosynthesis locus. The CPS locus of the serotype K7 strain 34-4 shows the highest similarity with the K1 serotype strain W83. B. For each probe in the CPS biosynthesis locus and for each test strain a log-ratio value compared to strain W83 is depicted by a data point, supporting the heat map results as shown in figure 4A.

The capsular polysaccharide biosynthesis locus

Analysis of the CPS biosynthesis locus [24] of the test strains revealed a high variation as seen in figure 4A showing the hybridization log-ratios against W83. Our interpretation of the log-ratios depicted as a heat map showing presence, aberrance and absence of each of the CPS-locus genes is shown in figure 4B. Only *PG0106* and *PG0108* show no divergence in any strain and are thus among the core gene set as described earlier. The other genes in the locus show at least some aberrance. *PG0117* and *PG0118* are called absent in each test strain as concluded from our hybridization experiments. This supports the choice of these genes to design a K1-
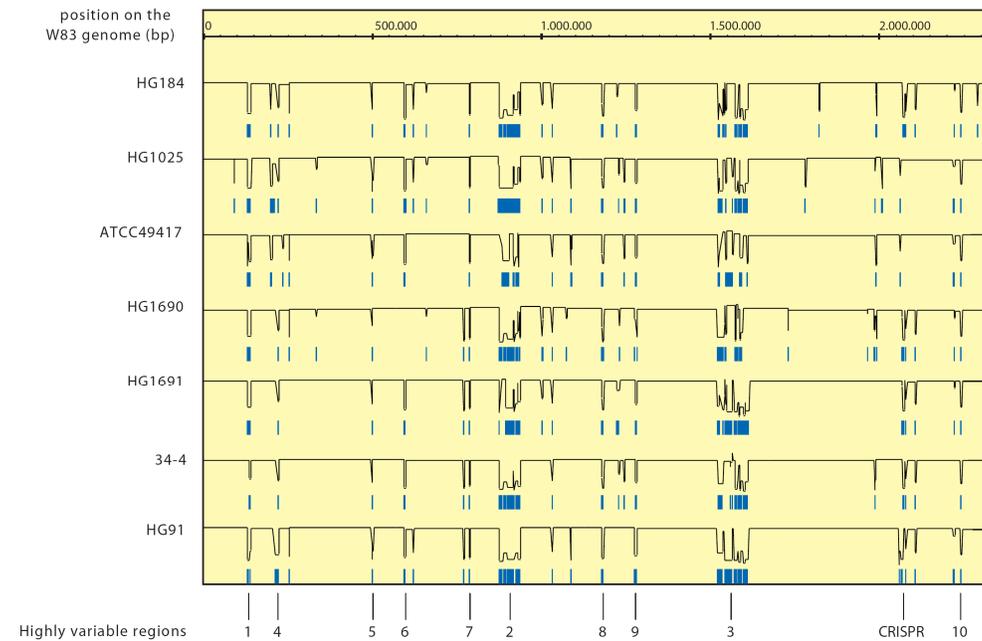
Figure 5. Highly variable regions of *P. gingivalis.* Breakpoint analysis of test strains describing potential lacking genomic regions as positioned on the W83 genome sequence. Black lines depict breakpoint data. As long as the line is flat there is low variability of the test strain compared to W83. Dips in the line indicate variability. Blue lines/rectangles below depict potential absent regions. At the top the probe positions are given as described in the W83 genome [29]. The numbers at the bottom label the 10 highly variable regions in each strain which are explained in the text. CRISPR represents a region of interest with CRISPR associated genes as described in the text.

specific PCR for serotyping in our group [54]. All test strains are found to be aberrant for at least 8 genes, except strain 34-4 (K7) which only shows aberrance in 5 genes. These findings may suggest that the different capsular serotypes can be highly variable in structure and that K7 CPS may share more common elements with the K1 type of CPS than the other test strains.

Highly variable regions

An analysis was performed to calculate the chance that certain genetic regions of the W83 genome are missing in the test strains included in the hybridization experiments. This was done using breakpoint analysis, which takes the divergence of neighbouring genes into account. In this analysis 10 highly variable regions were found (figure 5). Three regions, regions 1, 2 and 3, have already been reported earlier based on aberrance in strain

ATCC33277 [25] (table 6), but only a function for the CPS biosynthesis locus has been described. The function of the other two may be pathogenicity islands, although no prove has been reported yet. Region 4 which includes *ragA* and *ragB* is in addition to W83 only present in strain ATCC49417. Both strains are representatives of the 16S-23S ISR heteroduplex types that have the strongest association with disease. The other strains lack this region. This region has also been described as disease related directly by PCR of subgingival samples [55]. Region 5 includes *pgaA*, which also has been described as a virulence determinant [56]. The other highly variable regions may be involved in virulence, but too little is known to speculate on the functions.

Another region that was found to be interesting in this analysis is region *PG1981-PG1986* which is comprised of clustered regularly interspaced short palindromic repeat (CRISPR) associated genes (CAS) [57]. Together with CRISPRs, located directly downstream of *PG1981*, these types of genes have been described as the immune system of bacteria against foreign DNA, *e.g.* plasmids and viruses. Recently they also have been described as a useful tool in epidemiology [58]. Variation is expected to be high in these regions as they encompass exogenous DNA sequence fragments from infection events that happened to the strain or its ancestors. Here variation within the CAS genes is evident, but not as high as the other regions mentioned in this section.

W83-specific genes

Strain W83 has been described as a highly virulent strain. What makes this strain special is however not specifically known. The purified CPS of W83 has been shown to induce a higher immune response than other types of CPS [26]. Removal of the capsular structure, by genetic interruption of CPS-biosynthesis, however resulted in a much higher immune response when infecting fibroblasts with viable *P. gingivalis* [27]. What this means for virulence in a mouse model has not yet been addressed. With the data presented here a more detailed study is possible to find specific traits that make W83 different. A list of all genes that are aberrant in each of the test strains and absent in each of the test strains is presented (see Additional file 2). Among the 65 genes that were aberrant in all test strains there were 10 DNA mobilization genes, 5 DNA modification genes, 3 CPS biosynthesis genes, 2 carbohydrate metabolism genes, 2 transcriptional regulator genes, 2 lipoprotein genes and 36 (conserved) hypothetical protein genes. From this gene set 39 genes were W83-specific as they were absent in each of

104

Table 4. Genes only aberrant in the non-encapsulated strain FDC381.

GeneID	Annotated function
PG0183	lipoprotein, putative
PG0204	hypothetical protein
PG0300	TPR domain protein
PG0492	hypothetical protein
PG1119	flavodoxin, putative
PG1199	hypothetical protein
PG1200	hypothetical protein
PG1373	hypothetical protein
PG1466	hypothetical protein
PG1467	methlytransferase, UbiE-COQ5 family
PG1473	conjugative transposon protein TraQ
PG1685	hypothetical protein
PG1711	alpha-1,2-mannosidase family protein
PG1777	conserved hypothetical protein
PG1786	hypothetical protein
PG1814	DNA primase
PG1969	hypothetical protein
PG1970	hypothetical protein
PG1972	hemagglutinin protein HagB
PG1977	hypothetical protein
PG1978	hypothetical protein

the test strains. In this way the *prtT* protease gene and a fimbrillin gene (*fimA*) were found to be aberrant in all test strains, but not W83-specific as they were present in one or more test strains. The results for *fimA* support the findings that the gene is widely distributed, but variable at the probe locus among *P. gingivalis* strains. Many of the genes found in this analysis are located within the highly variable regions described in earlier publications using whole-genome analysis. The existence of those regions was supported by data comparing the genome sequences of *P. gingivalis* strains W83 and ATCC33277 [28]. Also in this study we found these regions back in the analysis as described above

Genes only aberrant in FDC381

FDC381 is the only strain included in this study that does not produce CPS. It is also the least virulent strain in mouse studies. Here, an analysis was performed to find genes that are specifically aberrant in FDC381 and

not in all the other test strains (table 7). Alongside many genes encoding hypothetical proteins several genes of special interest were found. The genes *PG1711* encoding an alpha-1,2-mannosidase family protein, and *PG1972* encoding the hemagglutinin *hagB*, all thought to be involved in virulence either by a role in evasion of the immune system or by a role in adhesion to host cells [29, 59].

Although these data do not directly show any CPS biosynthesis specific genes aberrant only in the non-encapsulated FDC381 it does give hints towards other virulence associated traits that are missing in FDC381.

High versus lower virulence strains

When comparing the core gene set of only the highly virulent strains W83, HG1025, ATCC49417 and HG1690 with the genes aberrant in each of the less virulent strains HG184, HG1691, 34-4 and FDC381 an interesting result was seen. There is only a single gene, *hmuS,* that is present in all highly virulent strains but aberrant in each of the less virulent strains. *HmuS* is part of the *hmuYRSTUV* haemin uptake system [60]. Haemin acquisition is vital for *P. gingivalis* which makes the haemin uptake and storage system relevant study objects. Lacking part of an important uptake mechanism could have consequences for infection and survival. However, in these experiments no functional differences have been shown.

Conclusions

In this study we analyzed the genetic contents of representative strains of each of the seven capsular serotypes. Comparative genomic hybridization shows that gene aberrance among *P. gingivalis* strains can be up to 13.7 %, which is higher than previously reported. The *P. gingivalis* genome is variable with 20 % of the W83 gene content being aberrant in at least one of the seven test strains. Analysis of virulence-related genes conservation was performed; only a few virulence-related genes were shown to be aberrant among test strains. As could be expected due to the choice of strains it was found that among the most aberrant virulence genes were the CPS biosynthesis genes.

In this study we initiated the description of a core genome of the anaerobic bacterium *P. gingivalis*, one of the most important causative agents of periodontitis allowing a more focused search for potential important virulence factors of which several were identified.

Methods

Bacterial strains and maintenance

P. gingivalis strains used in this study are listed in table 1, including serotype, origin and virulence level. *P. gingivalis* strains were first grown on 5% horse blood agar plates (Oxoid no. 2, Basingstoke, UK) supplemented with hemin (5 µg/ml) and menadione (1 µg/ml) (BA+H/M plates) at 37ºC in an anaerobic atmosphere of 80% N_2, 10% H_2, and 10% CO_2. From these plates 10 ml of liquid brain heart infusion broth supplemented with hemin (5 µg/ml) and menadione (1 µg/ml) (BHI+H/M) was inoculated and grown overnight as a pre-culture at 37ºC in an anaerobic atmosphere. From the pre-culture a 300 ml 1:100 dilution in BHI+H/M was made, which was grown overnight at 37ºC in an anaerobic atmosphere. The bacteria were washed 3 times in phosphate-buffered saline (PBS) and then pelleted and stored at - 80 ºC until DNA isolation was performed.

Microarray design

Whole-genome microarrays made for *P. gingivalis* strain W83 kindly provided by the Pathogen Functional Genomics Resource Center (The Institute for Genomic Research (TIGR), Rockville, MD) were used in this study. The aminosilane-coated microarrays contain 1,907 70-mer oligonucleotide probes designed on the 1,990 annotated W83 ORFs as found by TIGR. Each probe was designed to be unique for an ORF, so ORFs that were not unique were excluded. The arrays also included 500 *Arabidopsis thaliana* control probes. Each probe was printed four times on an array. Specific information about the microarrays can be found at: http://pfgrc.jcvi.org/index.php/microarray/array_description/porphyromonas_ gingivalis/version1.html.

DNA isolation

P. gingivalis pellets were frozen at -80°C until DNA isolation. Frozen pellets were ground in liquid nitrogen with a mortar and pestle until a fine powder was obtained. 500 µl of this powder was transferred to a liquid nitrogen pre-chilled 15 ml tube. DNA was extracted by addition of 1500 µl 65°C CTAB extraction buffer made to 2% (v/v) 2-mercaptoethanol before use (100 mM Tris-Cl (pH 8.0), 2.0 M NaCl, 20 mM EDTA, 3% (w/v) CTAB (H6269, Sigma-Aldrich), 2% (w/v) PVP-40 (PVP40, Sigma-Aldrich); Filter sterilized and stored at room temperature). After incubation for 30 min at

65°C with occasional mixing, DNA was extracted with 1500 µl phenol/chloroform/isoamylalcohol (25:24:1) (pH 7.9) (AM9730, Ambion). After centrifugation at 6,000xg for 15 min, the aqueous phase was transferred to a clean 15 ml tube and DNA was precipitated with an equal volume of ice-cold isopropanol. DNA was pelleted at 6,000 x g for 15 min. The DNA pellet was washed twice with ice-cold 70% ethanol and centrifugation at 6,000 x g for 5 min. The remaining liquid was removed by decanting and the pellet was air dried. This pellet was resuspended in 600 µl TE and 1 µl RNAse A (10 mg/ml, R6513, Sigma-Aldrich) was added. Residual RNA was removed by overnight incubation at 37°C and DNA was re-extracted with an equal volume of phenol/chloroform/isoamylalcohol (25:24:1) pH 7.9. The aqueous phase was recovered by centrifugation at 6,000 x g for 15 min. The aqueous layer was treated with an equal volume of chloroform/IAA (96:4) and centrifuged at 6,000 x g for 10 min at room temperature. The final aqueous phase was treated with an equal volume of 100% ethanol and 1/10 volume of 3M sodium acetate (pH 5.2) and incubated for 30 min @ -20°C. DNA was pelleted for 15 min at 6,000xg. Residual liquid was removed and the pellet was washed once with ice-cold 70% ethanol. DNA was pelleted for 5 min at 6,000 x g and the pellet was air-dried. The DNA pellet was resuspended in an appropriate volume of TE. DNA quality was verified with gel electrophoresis (0.5% agarose in TAE).

Genomic DNA labelling, microarray hybridization, scanning and data extraction

1 µg of genomic DNA was labeled with Cy3 or Cy5 using the CGH labeling kit for oligo arrays (ENZO Life Sciences). Labeled genomic DNA was purified with the QiaQuick PCR purification kit (Qiagen). *P. gingivalis* (W83) version 1 arrays were obtained from the Pathogen Functional Genomics Resource Center (PFGRC). Individual arrays were hybridized with 5 µg Cy3- and 5 µg Cy5-labeled material (test strains versus FDC381, which served as common reference), without dye swap, according to the Oligonucleotide Array-Based CGH for Genomic DNA Analysis manual (Agilent Technologies version 5.0). Briefly, labeled DNA was combined with 52 µl 10x Blocking Agent and 260 µl 2x Gex Hybridization Buffer Hi-RPM (Gene Expression Hybridization Kit, Agilent Technologies) in a total volume of 520 µl. Hybridization samples were incubated at 95°C for 3 min, spun down and hybridized at 37°C for 30 min. Samples were spun down and 490 µl of each sample was loaded onto a 1x244k backing in a SureHyb hybridization chamber (Agilent Technologies) and a *P. gingivalis* version 1

array was placed on top. Hybridization was performed at 65°C for 24h and 10 RPM in a hybridization oven (G2545A, Agilent Technologies). After the hybridization the backings were removed in LSW (2x SSC, 0.1% Sarkosyl (L9150, Sigma-Aldrich) at room temperature, washed for 5 min at 42°C in LSW, washed for 10 min at room temperature in HSW (0.1x SSC, 0.1% Sarkosyl) and finally washed for 1 min at room temperature in FW (0.1x SSC). Each array was dipped 5 times in H_2O and quickly submerged in isopropanol. Microarrays were spun dry for 1 min at 232xg and scanned on an Agilent G2505B scanner at 5 μm resolution and data was extracted with Feature Extraction version 9.5.3.1. (Protocol GE2-NonAT_95_Feb07).

Experimental design and microarray data analysis

Each strain was cultured in triplicate, in three experimental batches. DNA isolations and hybridizations were therefore performed three times for each strain, each being a biological replicate analyzed in one experimental block. On each array four technical replicate spots were spotted.

After log2 transformation, the data was normalized by a global Lowess smoothing procedure, omitting the probes with highly divergent intensities because of the bias they induced. A mixed ANOVA model (as described in [61]) with group-means-parameterization was used to normalize the data and collapse the technical and biological replicates. The gene specific model was:

$$y_{iklmn} = \mu + \tau_i + \rho + S_k + A_l + B_m + \varepsilon_{iklmn}$$

y_{ijklmn} represents log2 expression intensities, μ is the gene specific mean, т represents fixed strain effects (i = 1, …, 8), ρ is an indicator variable indicating the common reference, S represents random spot effects (j = 1, …, 96), A represents random array effects (i = 1, …, 24), and B represents experimental batch effects (m = 1, …, 3). Normalized average (Cy5) intensities for each strain were calculated as y_i^* = μ + τ_i and normalized average log2-ratio's with respect to W83 were calculated as Y_i^* = τ_i - τ_1, for each $i \neq 1$ (which represents W83).

Hence, each strain was compared with W83, and deviations in log2-ratio's were interpreted as aberrations. Given j genes divergence from zero were modelled as posterior probabilities of change under a mixture model, where non-divergent $Y_{ij}^* \sim N(0, s_i^2)$ and divergent Y_{ij}^* follows a uniform distribution [62].

Highly variable regions due to mutations or loss were quantified according to [63], using their GLAD (Gain and Loss Analysis of DNA)

package with default parameter settings. Finally, we used the negative control probes from *Arabidopsis thaliana* to define absent calls with the aim to quantify whether an aberration was found more likely due to mutation or loss. The distributions of intensities suggested a distinguishable mixed distribution of intensities from probes interrogating present genes (high) and probes interrogating absent genes (low; figure 1). Given *j* probes, probe intensities were modelled using a standard Gaussian mixture model where probes interrogating present genes $y_{ij}{}^* \sim N_1(\mu_{1i}, s_{1i}{}^2)$ and probes interrogating absent genes $y_{ij}{}^* \sim N_2(\mu_{2i}, s_{2i}{}^2)$. The probe specific membership probabilities of $N_1(\mu_{1i}, s_{1i}{}^2)$ represents the null-hypothesis of "not absent", which is the hypothesis under test. False discovery rate correction as described by [64] was applied to both the test for quantifying aberrations as well as to the test for quantifying genomic losses. The data was visualized using the Integrated Genome Browser [65]. The final data set including dead probes and conserved, aberrant and absent genes is shown in additional file 3.

Additional files

Additional file 1. Conserved core gene set of *P. gingivalis*. The conserved core genes of *P. gingivalis* consisting of 1476 genes and two ambiguous genes, which are called non-aberrant but absent

Additional file 2. W83-specific genes 65 genes. aberrant in each test strain of which 39 W83-specific genes (marked in red)

Additional file 3. *P. gingivalis* CGH data set. Table listing each *P. gingivalis* probe included in the results of this study in the order of geneID, including annotation. Low adjP-values (<0.05) depicted in yellow indicate aberrance in a test strain. High adj Pvals. absent (>0.99) depicted in red indicate absence in the test strain. Black rows indicate the dead probes as found on the W83 array in this study. Zooming out gives an overview of the whole genomic diversity along the test strains.

Additional files can be found at the *BMC microbiology* website: http://www.biomedcentral.com/1471-2180/10/252/additional/

Requests for a CD or e-mail with the additional files can be directed to Jorg Brunner via e-mail: jorgbrunner@hotmail.com

Abbreviations

AFLP amplified fragment length polymorphism; BA+H/M Horse blood agar plates supplemented with hemin and menadione; CDSs Coding sequences; CGH Comparative genomic hybridization; CPS Capsular polysaccharide; LPS Lipopolysaccharide; OMP Outer membrane protein; ORF Open reading frame; PBS Phosphate-buffered saline; SSC Saline-sodium citrate buffer; TAE Tris-acetate-EDTA buffer; TE Tris-EDTA buffer

Acknowledgements

References

1. Hugoson A, Sjodin B, Norderyd O: **Trends over 30 years, 1973-2003, in the prevalence and severity of periodontal disease**. *J Clin Periodontol* 2008, **35**(5):405-414.
2. Phipps KR, Chan BK, Jennings-Holt M, Geurs NC, Reddy MS, Lewis CE, Orwoll ES: **Periodontal health of older men: the MrOS dental study**. *Gerodontology* 2009, **26**(2):122-129.
3. Skudutyte-Rysstad R, Eriksen HM, Hansen BF: **Trends in periodontal health among 35-year-olds in Oslo, 1973-2003**. *J Clin Periodontol* 2007, **34**(10):867-872.
4. Genco R, Offenbacher S, Beck J: **Periodontal disease and cardiovascular disease: epidemiology and possible mechanisms**. *J Am Dent Assoc* 2002, **133 Suppl**:14S-22S.
5. Grossi SG, Genco RJ: **Periodontal disease and diabetes mellitus: a two-way relationship**. *Ann Periodontol* 1998, **3**(1):51-61.
6. Loos BG, Craandijk J, Hoek FJ, Wertheim-van Dillen PM, van der Velden U: **Elevation of systemic markers related to cardiovascular diseases in the peripheral blood of periodontitis patients**. *J Periodontol* 2000, **71**(10):1528-1534.
7. Mercado FB, Marshall RI, Klestov AC, Bartold PM: **Relationship between rheumatoid arthritis and periodontitis**. *J Periodontol* 2001, **72**(6):779-787.

8.   Pathirana RD, O'Brien-Simpson NM, Reynolds EC: **Host immune responses to *Porphyromonas gingivalis* antigens**. *Periodontol 2000* 2010, **52**:218-237.

9.   Paster BJ, Boches SK, Galvin JL, Ericson RE, Lau CN, Levanos VA, Sahasrabudhe A, Dewhirst FE: **Bacterial diversity in human subgingival plaque**. *J Bacteriol* 2001, **183**(12):3770-3783.

10.   Keijser BJ, Zaura E, Huse SM, van der Vossen JM, Schuren FH, Montijn RC, ten Cate JM, Crielaard W: **Pyrosequencing analysis of the oral microflora of healthy adults**. *J Dent Res* 2008, **87**(11):1016-1020.

11.   Zaura E, Keijser BJ, Huse SM, Crielaard W: **Defining the healthy "core microbiome" of oral microbial communities**. *BMC Microbiol* 2009, **9**(1):259.

12.   Slots J, Bragd L, Wikstrom M, Dahlen G: **The occurrence of *Actinobacillus actinomycetemcomitans*, *Bacteroides gingivalis* and *Bacteroides intermedius* in destructive periodontal disease in adults**. *J Clin Periodontol* 1986, **13**(6):570-577.

13.   Rosen G, Sela MN: **Coaggregation of *Porphyromonas gingivalis* and *Fusobacterium nucleatum* PK 1594 is mediated by capsular polysaccharide and lipopolysaccharide**. *FEMS Microbiol Lett* 2006, **256**(2):304-310.

14.   Domenico P, Salo RJ, Cross AS, Cunha BA: **Polysaccharide capsule-mediated resistance to opsonophagocytosis in *Klebsiella pneumoniae***. *Infect Immun* 1994, **62**(10):4495-4499.

15.   Noel GJ, Hoiseth SK, Edelson PJ: **Type b capsule inhibits ingestion of *Haemophilus influenzae* by murine macrophages: studies with isogenic encapsulated and unencapsulated strains**. *J Infect Dis* 1992, **166**(1):178-182.

16.   Glynn AA, Howard CJ: **The sensitivity to complement of strains of *Escherichia coli* related to their K antigens**. *Immunology* 1970, **18**(3):331-346.

17.   Sundqvist G, Figdor D, Hanstrom L, Sorlin S, Sandstrom G: **Phagocytosis and virulence of different strains of *Porphyromonas gingivalis***. *Scand J Dent Res* 1991, **99**(2):117-129.

18.   Laine ML, van Winkelhoff AJ: **Virulence of six capsular serotypes of *Porphyromonas gingivalis* in a mouse model**. *Oral Microbiol Immunol* 1998, **13**(5):322-325.

19.   Laine ML, Appelmelk BJ, van Winkelhoff AJ: **Novel polysaccharide capsular serotypes in *Porphyromonas gingivalis***. *J Periodontal Res* 1996, **31**(4):278-284.

20.   van Winkelhoff AJ, Appelmelk BJ, Kippuw N, de Graaff J: **K-antigens in *Porphyromonas gingivalis* are associated with virulence**. *Oral Microbiol Immunol* 1993, **8**(5):259-265.

21.   Holt SC, Kesavalu L, Walker S, Genco CA: **Virulence factors of *Porphyromonas gingivalis***. *Periodontol 2000* 1999, **20**:168-238.

22. Lamont RJ, Jenkinson HF: **Life below the gum line: pathogenic mechanisms of *Porphyromonas gingivalis***. *Microbiol Mol Biol Rev* 1998, **62**(4):1244-1263.

23. Nakayama K: **Molecular genetics of *Porphyromonas gingivalis*: gingipains and other virulence factors**. *Curr Protein Pept Sci* 2003, **4**(6):389-395.

24. Aduse-Opoku J, Slaney JM, Hashim A, Gallagher A, Gallagher RP, Rangarajan M, Boutaga K, Laine ML, Van Winkelhoff AJ, Curtis MA: **Identification and characterization of the capsular polysaccharide (K-antigen) locus of *Porphyromonas gingivalis***. *Infect Immun* 2006, **74**(1):449-460.

25. Chen T, Hosogi Y, Nishikawa K, Abbey K, Fleischmann RD, Walling J, Duncan MJ: **Comparative whole-genome analysis of virulent and avirulent strains of *Porphyromonas gingivalis***. *J Bacteriol* 2004, **186**(16):5473-5479.

26. d'Empaire G, Baer MT, Gibson FC, 3rd: **K1 serotype capsular polysaccharide of *Porphyromonas gingivalis* elicits chemokine production from murine macrophages that facilitates cell migration**. *Infect Immun* 2006, **74**(11):6236-6243.

27. Brunner J, Scheres N, El Idrissi NB, Deng DM, Laine ML, van Winkelhoff AJ, Crielaard W: **The capsule of *Porphyromonas gingivalis* reduces the immune response of human gingival fibroblasts**. *BMC Microbiol* 2010, **10**(1):5.

28. Naito M, Hirakawa H, Yamashita A, Ohara N, Shoji M, Yukitake H, Nakayama K, Toh H, Yoshimura F, Kuhara S *et al*: **Determination of the Genome Sequence of *Porphyromonas gingivalis* Strain ATCC 33277 and Genomic Comparison with Strain W83 Revealed Extensive Genome Rearrangements in *P. gingivalis***. *DNA Res* 2008, **15**(4):215-225.

29. Nelson KE, Fleischmann RD, DeBoy RT, Paulsen IT, Fouts DE, Eisen JA, Daugherty SC, Dodson RJ, Durkin AS, Gwinn M *et al*: **Complete genome sequence of the oral pathogenic Bacterium *Porphyromonas gingivalis* strain W83**. *J Bacteriol* 2003, **185**(18):5591-5601.

30. Igboin CO, Griffen AL, Leys EJ: ***Porphyromonas gingivalis* strain diversity**. *J Clin Microbiol* 2009, **47**(10):3073-3081.

31. Paramonov N, Rangarajan M, Hashim A, Gallagher A, Aduse-Opoku J, Slaney JM, Hounsell E, Curtis MA: **Structural analysis of a novel anionic polysaccharide from *Porphyromonas gingivalis* strain W50 related to Arg-gingipain glycans**. *Mol Microbiol* 2005, **58**(3):847-863.

32. Chen PB, Davern LB, Aguirre A: **Experimental *Porphyromonas gingivalis* infection in nonimmune athymic BALB/c mice**. *Infect Immun* 1991, **59**(12):4706-4709.

33. van Steenbergen TJ, Kastelein P, Touw JJ, de Graaff J: **Virulence of black-pigmented Bacteroides strains from periodontal**

pockets and other sites in experimentally induced skin lesions in mice. *Journal of periodontal research* 1982, **17**(1):41-49.

34. Pathirana RD, O'Brien-Simpson NM, Brammar GC, Slakeski N, Reynolds EC: **Kgp and RgpB, but not RgpA, are important for Porphyromonas gingivalis virulence in the murine periodontitis model**. *Infect Immun* 2007, **75**(3):1436-1442.

35. Fletcher HM, Schenkein HA, Morgan RM, Bailey KA, Berry CR, Macrina FL: **Virulence of a *Porphyromonas gingivalis* W83 mutant defective in the *prtH* gene**. *Infect Immun* 1995, **63**(4):1521-1528.

36. Walters S, Rodrigues P, Belanger M, Whitlock J, Progulske-Fox A: **Analysis of a band 7/MEC-2 family gene of *Porphyromonas gingivalis***. *J Dent Res* 2009, **88**(1):34-38.

37. Shi X, Hanley SA, Faray-Kele MC, Fawell SC, Aduse-Opoku J, Whiley RA, Curtis MA, Hall LM: **The *rag* locus of *Porphyromonas gingivalis* contributes to virulence in a murine model of soft tissue destruction**. *Infect Immun* 2007, **75**(4):2071-2074.

38. Bagaitkar J, Williams LR, Renaud DE, Bemakanakere MR, Martin M, Scott DA, Demuth DR: **Tobacco-induced alterations to *Porphyromonas gingivalis*-host interactions**. *Environ Microbiol* 2009, **11**(5):1242-1253.

39. Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, Paulsen IT, de Marsac NT, Wincker P, Dossat C *et al*: **Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria**. *Genome Biol* 2008, **9**(5):R90.

40. Fischer W, Windhager L, Rohrer S, Zeiller M, Karnholz A, Hoffmann R, Zimmer R, Haas R: **Strain-specific genes of *Helicobacter pylori*: genome evolution driven by a novel type IV secretion system and genomic island transfer**. *Nucleic Acids Res* 2010 [epub ahead of print].

41. Foote SJ, Bosse JT, Bouevitch AB, Langford PR, Young NM, Nash JH: **The complete genome sequence of *Actinobacillus pleuropneumoniae* L20 (serotype 5b)**. *J Bacteriol* 2008, **190**(4):1495-1496.

42. Rasmussen TB, Danielsen M, Valina O, Garrigues C, Johansen E, Pedersen MB: **Streptococcus thermophilus core genome: comparative genome hybridization study of 47 strains**. *Appl Environ Microbiol* 2008, **74**(15):4703-4710.

43. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O *et al*: **Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths**. *PLoS Genet* 2009, **5**(1):e1000344.

44. Waterhouse JC, Swan DC, Russell RR: **Comparative genome hybridization of *Streptococcus mutans* strains**. *Oral Microbiol Immunol* 2007, **22**(2):103-110.

45.   Wu J, Yu T, Bao Q, Zhao F: **Evidence of extensive homologous recombination in the core genome of *rickettsia*.** *Comp Funct Genomics* 2009:510270.

46.   Hosogi Y, Duncan MJ: **Gene expression in *Porphyromonas gingivalis* after contact with human epithelial cells.** *Infect Immun* 2005, **73**(4):2327-2335.

47.   Yoshimura M, Ohara N, Kondo Y, Shoji M, Okano S, Nakano Y, Abiko Y, Nakayama K: **Proteome analysis of *Porphyromonas gingivalis* cells placed in a subcutaneous chamber of mice.** *Oral Microbiol Immunol* 2008, **23**(5):413-418.

48.   Bourgeau G, Lapointe H, Peloquin P, Mayrand D: **Cloning, expression, and sequencing of a protease gene (*tpr*) from *Porphyromonas gingivalis* W83 in *Escherichia coli*.** *Infect Immun* 1992, **60**(8):3186-3192.

49.   Rumpf RW, Griffen AL, Leys EJ: **Phylogeny of *Porphyromonas gingivalis* by ribosomal intergenic spacer region analysis.** *J Clin Microbiol* 2000, **38**(5):1807-1810.

50.   Rawlings ND, Barrett AJ, Bateman A: **MEROPS: the peptidase database.** *Nucleic Acids Res* 2010, **38**(Database issue):D227-233.

51.   Grenier D, Mayrand D: **Selected characteristics of pathogenic and nonpathogenic strains of *Bacteroides gingivalis*.** *J Clin Microbiol* 1987, **25**(4):738-740.

52.   Tokuda M, Karunakaran T, Duncan M, Hamada N, Kuramitsu H: **Role of Arg-gingipain A in virulence of Porphyromonas gingivalis.** *Infect Immun* 1998, **66**(3):1159-1166.

53.   Nakayama K, Kadowaki T, Okamoto K, Yamamoto K: **Construction and characterization of arginine-specific cysteine proteinase (Arg-gingipain)-deficient mutants of Porphyromonas gingivalis. Evidence for significant contribution of Arg-gingipain to virulence.** *J Biol Chem* 1995, **270**(40):23619-23626.

54.   Brunner J, Crielaard W, van Winkelhoff AJ: **Analysis of the capsular polysaccharide biosynthesis locus of *Porphyromonas gingivalis* and development of a K1-specific polymerase chain reaction-based serotyping assay.** *J Periodontal Res* 2008, **43**(6):698-705.

55.   Hanley SA, Aduse-Opoku J, Curtis MA: **A 55-kilodalton immunodominant antigen of *Porphyromonas gingivalis* W50 has arisen via horizontal gene transfer.** *Infect Immun* 1999, **67**(3):1157-1171.

56.   Rigg GP, Roberts IS: **The molecular cloning, nucleotide sequence and expression of an antigenic determinant from *Porphyromonas gingivalis*.** *Arch Oral Biol* 2000, **45**(1):41-52.

57.   Jansen R, Embden JD, Gaastra W, Schouls LM: **Identification of genes that are associated with DNA repeats in prokaryotes.** *Mol Microbiol* 2002, **43**(6):1565-1575.

58.   Horvath P, Barrangou R: **CRISPR/Cas, the immune system of bacteria and archaea**. *Science (New York, NY* 2010, **327**(5962):167-170.

59.   Progulske-Fox A, Tumwasorn S, Lepine G, Whitlock J, Savett D, Ferretti JJ, Banas JA: **The cloning, expression and sequence analysis of a second *Porphyromonas gingivalis* gene that codes for a protein involved in hemagglutination**. *Oral Microbiol Immunol* 1995, **10**(5):311-318.

60.   Lewis JP, Plata K, Yu F, Rosato A, Anaya C: **Transcriptional organization, regulation and role of the *Porphyromonas gingivalis* W83 *hmu* haemin-uptake locus**. *Microbiology* 2006, **152**(Pt 11):3367-3382.

61.   Kerr MK, Martin M, Churchill GA: **Analysis of variance for gene expression microarray data**. *J Comput Biol* 2000, **7**(6):819-837.

62.   Yang YH, Speed TP: **Design and analysis of comparative microarray experiments**. In: *Statistical Analysis of Gene Expression Microarray Data.* Edited by Speed T. Boca Raton: Chapman and Hall/CRC CRC Press LLC; 2003: 35-92.

63.   Hupé P, Stransky N, Thiery JP, Radvanyi F, Barillot E: **Analysis of array CGH data: from signal ratio to gain and loss of DNA regions**. *Bioinformatics* 2004, **20**(18):3413-3422.

64.   Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple esting**. *J R Statist Soc B* 1995, **57**(1):289-300.

65.   Nicol JW, Helt GA, Blanchard SG, Jr., Raja A, Loraine AE: **The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets**. *Bioinformatics* 2009, **25**(20):2730-2731.